# Refining Spatial Grocery Models of Consumer Behaviour: An Individual-Based Approach

Kim Maria Butterfield

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

The University of Leeds

Faculty of Earth and Environment

School of Geography

July 2023

# Acknowledgements

Completing this thesis was only possible with the interest, support and patience of many individuals I have met along the way.

First, I want to thank Andy Newing for his excellent support and guidance throughout the study. Andy has been there for me from my time at Leeds as an undergraduate and has always willingly provided time, effort and guidance. I also want to thank Alison Heppenstall for inspiring me as a young undergraduate to research agent-based modelling and for believing in me to undertake such a challenging project. I would also like to acknowledge the advice and input from my Research Support Group members, Nick Hood and Graham Clarke, both of whom are experts in retail geography and provided helpful and constructive feedback from the beginning.

I want to thank the Sainsbury's Property and Analytics team for providing me with a fantastic three months working with their team and this research opportunity - specifically, Tim Rains for his support and guidance acting as collaborating supervisor during the pandemic.

I want to express gratitude to Ric Colasanti for his expertise in agent-based modelling, for providing advice around model logic and software application, and for inspiring me to become a better programmer. I am grateful to Geolytix for their enthusiasm towards my research project and for hiring me as a Data Scientist at Geolytix, I look forward to becoming a full time employee whilst publishing the findings from this thesis.

I want to say an exceptionally massive thank you to my partner, Beth. She has been with me from start to finish throughout this process and has helped me through some of the most challenging times life has thrown my way. Words cannot express my gratitude for your support throughout this journey. Thank you, Beth.

Finally, thank you to the ESRC for funding this project and the ESRC WRDTP for supporting my internship at Sainsbury's under their Company Internship Scheme.

# Thesis abstract

Consumer grocery behaviours have evolved over the past few decades, with consumer transaction behaviours becoming increasingly individualised due to the changes in cultural lifestyles, the increased demand for convenience, and the expanded provision of e-commerce. The UK grocery sector is a competitive market, and optimising their store network is of most importance to retailers. Current tools used within grocery location analytics use a top-down methodology, such as spatial interaction models. In the past, these tools have been considered the most appropriate for location impact assessments and predicting new store revenue. Whilst robust for the time, these models do not necessarily account for one of the most integral parts of store performance, *the consumer's behaviours.* Due to their top-down nature, current spatial models cannot account for the heterogeneous behaviours of grocery consumers, which are becoming increasingly diverse. These models would benefit from refinement in their ability to capture the nuanced behaviours of customers.

Therefore, the research presented in this thesis provides a framework for developing an individual-based model using modelling elements from decision trees, microsimulation, and, largely, agent-based modelling. First, a rarely accessed loyalty card-linked transaction dataset provided by the study collaborator, Sainsbury's, was analysed and segmented, identifying seven unique consumer type groups. Using these groups, a fully reproducible individual-based model was developed. The study highlights the advantages these models could bring to retailers by modelling transaction temporality, and the notable challenges encountered when modelling spatiality. The research in this thesis provides a novel framework to create such data-driven bottom-up models that utilise transaction data by known customers. The results

of this thesis present how individual-based models could be developed to provide grocery retailers with an integral tool that captures the diverse behaviours of consumers for site location analysis via scenario testing.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

ABM Agent-based Model (*or* Modelling)

GIS Geographical Information System

IBM Individual-based Model (*or* Modelling)

KISS Keep it Simple, Stupid

LAD Local Authority District

ML Machine Learning

MSM Microsimulation Model (*or* Modelling)

OA Output Area

OAC Output Area Classification

ODD Overview, Design Concepts, Details protocol

ONS Office for National Statistics

PWC Population-weighted Centroid

RPM Resale Price Maintenance Act

SIM Spatial Interaction Model

TTW Travel-to-Work

# Chapter 1 Introduction

With vast amounts of consumer data being collected every day, retailers are gaining a wealth of knowledge about the purchasing behaviours of their most loyal customers. This is especially true in the UK grocery sector, where the major retailers have long-established loyalty card schemes accumulating customer insight across multiple channels (in-store, online, and click and collect) (Knox and Denison, 2000; Bombaij and Dekimpe, 2020; Hood et al., 2021; Rains and Longley, 2021). These data allow retailers to better understand how consumer behaviours adapt and respond to supply and demand-side changes. These data can support retailers in making location-based decisions, including site selection, store format development or the inclusion of online retailing opportunities (Thompson et al., 2012; Newing et al., 2015; Waddington et al., 2018). One of the critical opportunities these data could provide is the ability to build models that incorporate and simulate both the spatial and temporal components of consumer behaviours, identifying how those individualised behaviours can impact the retailer at the store level and vice versa (Sturley et al., 2018).

Both academics and retailers researching and working within location analytics have a longstanding history of developing and applying spatial models to solve and support location-based decisions (Nakaya et al., 2007; Birkin and Clarke, 2019). Historically within the grocery sector, the focus has been on estimating grocery demand or on undertaking spatial modelling to better understand where to locate, relocate, refurbish or close stores within their store network (Hernández and Bennison, 2000; Reynolds and Wood, 2010), along with an awareness of the need to provide sufficient capacity for online groceries (Urquhart et al., 2022). To support these decisions, location professionals have utilised modelling techniques, with the most common tools using a top-down methodology (discussed in section 3.2). These techniques include methods such as spatial interaction modelling (SIM), in which the flow of consumer expenditure to stores is calculated via the disaggregation of store revenue amongst geographical areas based on numerous types of grocery demand, such as residential-based demand, workplace-demand, tourist-based demand, retail-based demand, and event-based demand (Roy and Thill, 2004; Waddington et al., 2018; Newing et al., 2018; Clarke and Birkin, 2018). Whilst these methods have become industry-standard (Rowe et al., 2022) due to their robustness, reliability and continual extensive development (Birkin and Heppenstall, 2011; Birkin and Clarke, 2012; Beckers et al., 2022), there has been an increased interest in bottom-up methodologies.

Within grocery location analytics, there has been increasing interest shown in individual-based modelling (IBM) in which the behaviours of individual agents are simulated. The following studies used an agent-based modelling (ABM) approach, a type of IBM, in the context of retail. Schenk et al. (2007) focused on building an ABM of grocery behaviour using only population and survey data, whereas Sturley et al. (2018) designed a proof-of-concept ABM based on the possibility of using loyalty card data. Bell and Mgbemena (2018) designed an ABM for mobile device sales using a decision-tree structure embedded within the model's base.

The work within this thesis is the result of a collaboration between the University of Leeds and Sainsbury's Supermarkets Ltd., a major British grocery retailer. Sainsbury's seeks to understand how their loyalty card-linked transaction data can be used and implemented into an IBM of consumer behaviour, covering transaction temporality and spatiality. Since the production of this research study, the author's employers in the location analytics sector have sought to incorporate the model methodology developed here to enhance their spatial modelling tools. The interest surrounding incorporating individual-level data into IBMs is ever-present, and the research findings presented here are significant steps towards filling this gap.

## 1.1 Research context

IBMs are widely applied across a range of research areas, including space and population dynamics (Bae et al., 2016), economics and finance (Axtell and Farmer, 2022), and ecological processes (DeAngelis and Grimm, 2014). These models simulate the behaviour or actions of individuals or groups of agents, which can then be aggregated to identify macro-level impacts that emerge from those micro-behaviours (Boero et al., 2008). The application of IBMs in retail location analytics can provide two key insights: firstly, to replicate and scenario-test the observed spatiotemporal behaviours of grocery consumers using empirical data, and secondly, an opportunity to identify how changes in consumer behaviour can impact store performance, and how supply side changes can impact consumer store or channel choice. Through the simulation of individuals and groups, the modeller can explore consumers' complex and dynamic behaviours at the aggregate level from the individual, as opposed to disaggregating known data based on loose assumptions (Chen, 2012; Lorscheid et al., 2019).

Two routes can be taken to build an IBM. The first route is to build a theoretical model that is loosely designed around observed data, where modelled agent behaviours are rather unspecific, allowing for true stochasticity (Yang and Gilbert, 2008; Gomez, 2019). The alternative route is to build a model that entirely depends on empirical data, in which agent behaviours are directly derived from historical observations (Sajjad et al., 2016; Rosés et al., 2021). Ideally, such a model in this applied context of grocery retail would be data-driven to truly mimic real-world behaviour (Bell and Mgbemena, 2018). This would ensure that the model would capture the habitual behaviours typically observed in relation to grocery shopping, which is often seen as a mundane and routine activity. However, it would also benefit from some element of stochasticity as human behaviour does follow a pattern of habituality that fluctuates from time to time (Knox and Denison, 2000; De Kervenoael et al., 2006; Krumme et al., 2013; Waddington et al., 2018). Therefore, developing an IBM of consumer behaviour would require a substantial amount of individualised data. Fortunately, as mentioned above, grocery retailers have been collecting data regarding their customers' behaviour since the 1990s (Ing and Mitchell, 1994) during the introduction of digitally-linked loyalty card schemes (Lauer, 2020). These data have been used to refine SIMs in past studies of location analytics (Newing et al., 2015; Waddington et al., 2018; Hood et al., 2021), yet are still to be applied in more IBM techniques (Sturley et al., 2018).

One of the most notable reasons why more IBMs have not been created or used within location analytics is due to the lack of accessibility to individual-level data (Quach et al., 2022). Due to private ownership and preservation of consumer information, these proprietary data are notoriously difficult to obtain within academic research. They are also yet to be used within the industry by retailers to refine their spatial models (Hummel et al., 2021). This is driven partly by the need for more expertise by those in the industry and the limited time available to explore the development of such models alongside their usual work. Additionally, IBMs are incredibly data-hungry (van der Ploeg et al., 2014) and require substantial computational power and time to run (Heppenstall et al., 2016). Despite these challenges, IBMs can provide us with a better methodology to better understand the spatiotemporal behaviours of grocery consumers via the simulation of individuals and groups (Heppenstall et al., 2013). Over recent decades, consumer grocery store and channel choice behaviours have evolved and developed in several ways, ranging from when customers go shopping, where they shop, and what channel they use. All aspects have become increasingly complex when linked to shopping missions. These changes are largely due to a concoction of expanded personal mobility (Vanhaverbeke and Macharis, 2011), a rise in convenience culture (Wood and Browne, 2007; Hood et al., 2016; Waddington et al., 2018), the proliferation of online shopping (Kirby-Hawkins et al., 2018; Hood et al., 2020), and changes to routines and workplace location following the pandemic.

Owning the second largest grocery market share at 15.6% in the UK (at the time of writing) (Kantar, 2022), Sainsbury's is a multi-national supermarket chain that owns Nectar; the most extensive multi-retailer loyalty card scheme in the UK, with over 17 million cardholders (nectar360.co.uk, 2023). Since the publication of Sturley et al. (2018), there has been an increased interest by grocery retailers like Sainsbury's to explore the shopping behaviours of their consumers and identify how these may adapt and change as not only consumer behaviours evolve themselves but how they respond to supply-side adjustments. Since the start of this thesis, the world has been impacted by the Covid-19 pandemic, causing unprecedented changes affecting many facets of life. A notable impact was on the UK grocery sector, disturbing various areas such as the supply chain and movement of goods, product availability, a proliferated increase in e-commerce, and a drastic change in consumer behaviour as lockdowns allowed customers to shop more flexibly when working from home. It would benefit grocery retailers if they could simulate and scenario-test such events via modelling the potential changes in consumer behaviour at the individual level to better prepare for potential impacts before such events occur again.

Sainsbury's first proposed this thesis in collaboration with the University of Leeds' School of Geography, which has a longstanding relationship with collaborative projects across their location planning and property analytics teams (Wright, 2008; LIDA, 2023). As a result of ongoing collaborations, several incremental enhancements have been made to the spatial modelling tools used by Sainsbury's teams to support new store revenue estimation and impact assessments (Newing et al., 2015; Hood et al., 2016; Waddington et al., 2018). The study conducted by Sturley et al. (2018) was a first attempt at incorporating the complex behaviours of Sainsbury's customers within a novel individual-level modelling framework. After thoroughly analysing the prototype model's logic proposed by Sturley et al. (2018) (section 6.1.3), this thesis considerably expands our understanding of the potential for individual-level models in the grocery retail location planning sector. For this study, Sainsbury's provided access to extensive loyalty card-linked transaction data, an integral dataset in allowing this study to come to fruition.

The research presented is split in two segments: part one focuses on the exploration of a high-volume loyalty card-linked transaction dataset, including the segmentation of customers to derive general consumer behavioural rules and identify key behaviours; part two focuses on the design and implementation of an IBM that simulates the store and channel choices of customers, using a hybrid-methodology that incorporates aspects of decision trees, microsimulation (MSM) and ABM. Each part of the thesis adds to the field of consumer behaviour and location analytics and will be incorporated into separate publications, with one already ready for submission (see thesis contributions in section 1.4).

The research in this thesis was produced under a studentship awarded by the Economic and Social Research Council's (ESRC) White Rose Doctoral Training Partnership (WRDTP). As part of this PhD, 3-months were spent working with the Sainsbury's Property team as an Analytical Solutions Junior Developer. During this internship, training was provided on the methodologies used to collate, analyse, and investigate the transaction data captured by Sainsbury's and how these data are used to build their location planning tools. With the experience gained from this internship, domain knowledge of the grocery industry, and expertise from collaborators, this thesis has been guided by Sainsbury's requirements and interests while expanding the spatial modelling research in academic literature. Due to the confidentiality of the in-house methodologies used at Sainsbury's, these cannot be outlined fully in this text but have been discussed in more general terms to provide context. All methods, data, and analytical tools used in this study were confirmed as suitable by a supervisor at Sainsbury's, who had considerable experience in both industry and academia.

To support the future development of individual-based spatial planning in grocery analytics, this study mines a novel dataset capturing 2.7 million grocery transactions undertaken by over 216 thousand Sainsbury's consumers over a 12-week period in 2018, capturing habitual behaviours pre-Covid-19. These insights enable the identification of typical consumer behaviours with respect to store and channel choice in such a way that behavioural rules can be derived for subsequent use in individual-level modelling. Using aspects from both IBM methods of MSM and ABM, this thesis meets its stated aims and objectives (section 1.2) and provides a working model of consumer behaviour that is fully reproducible with similar data requirements. The output of this thesis provides a bespoke modelling framework for modelling consumer behaviour, allowing retailers, retail-consultancies, academics and regulators (e.g. the Competition and Markets Authority) to model known consumer behaviours to better understand their customers and how they may change under different scenario tests.

In the latter stages of thesis, the author has been employed as a Spatial Data Scientist for the location planning and data analytics firm Geolytix, working with dozens of clients in the UK and internationally across various retail sectors, from fast-food companies and restaurants to sports brands and grocery retailers. In that role and using Geolytix's vast array of data products and client data, the author has designed bespoke analytical models using a variety of spatial modelling techniques ranging from analogue models, scorecards, SIMs and machine learning algorithm models such as XGBoost, a decision-tree-based technique, for multiple clients. The research and methodologies developed in this thesis have been designed and influenced by a combination of the experiences of working in the spatial retail analytics industry, being an academic researcher, and collaborating with the study collaborator, Sainsbury's. The combination of academic and industry experience provides a unique opportunity to develop a novel model that adds to the academic understanding of modelling retail consumer behaviour in such a way that it can be applied to industry location analytics.

## 1.2 Aims and objectives

The research presented in this thesis seeks to enrich the tools and methodologies used within location analytics by developing a prototype IBM in the application of grocery retail location analytics. The model developed is a first attempt at creating an IBM that simulates individual-level consumer behaviours, including temporality, spatiality, channel, and basket type choices. The model is designed to replicate the observed transactions to allow scenario testing using the customer population provided in the loyalty card-linked transaction data. Though the model has been designed primarily to simulate consumer behaviours at Sainsbury's stores in West Yorkshire, the model can be adapted to other grocery retailers and other geographical contexts where similar data are available.

The overall aims of this research are:

1. To present a review of the historical and recent changes in consumer behaviour in the context of grocery retail planning and the tools used in location planning analytics.
2. To identify key segments of consumer type groups who exemplify particular consumer behaviours regarding when, where, how and what they purchase from grocery stores.
3. To design an individual-based model based on known customer behaviour observations, incorporating temporality and spatiality using a hybrid microsimulation and agent-based modelling methodology.

To meet these aims, this thesis addresses the following objectives:

1. To explore the literature around British grocery retail over the past decades, identifying key changes that impacted consumer behaviour and the key indicators of consumer behaviour (Chapter 2).
2. To articulate the benefits and limitations of current grocery location planning analytical tools, including the potential data sources available (Chapter 3).
3. To investigate the loyalty card-linked transaction dataset provided by Sainsbury's, identifying general behaviours of their customers in relation to their store choice behaviours in West Yorkshire (0).
4. To segment the observed customers into consumer type groups based on their key behaviours regarding transaction frequency, store and channel choice, and purchase purpose, i.e., basket type (Chapter 5).
5. To perform data mining on the customer segments, identifying the probability of making a transaction at any given time based on their linked loyalty card transactions (Chapter 5).

6. To design and build a prototype individual-based model of consumer behaviours suitable for simulating their transactions by day type, time of day, channel, and basket type (Chapter 6).
7. To discuss how the model's functionality and framework is beneficial in retail location analytics, and how it can be used and further developed by retailers and location professionals.(Chapter 6 and Chapter 7).

Additionally, the research aims to expand the knowledge of consumer behaviours by analysing the transactional dataset, specifically around transaction temporality (time of the day/day of the week) and store and channel choice behaviour. The research produced in the thesis helps better understand whether loyalty card-linked transaction data of this calibre, in terms of size and quality, is sufficient for such model developments and how it can be used to develop more sophisticated IBMs. These objectives are addressed systematically throughout this thesis using the structure presented in section 1.3.

## 1.3 Thesis structure and scope

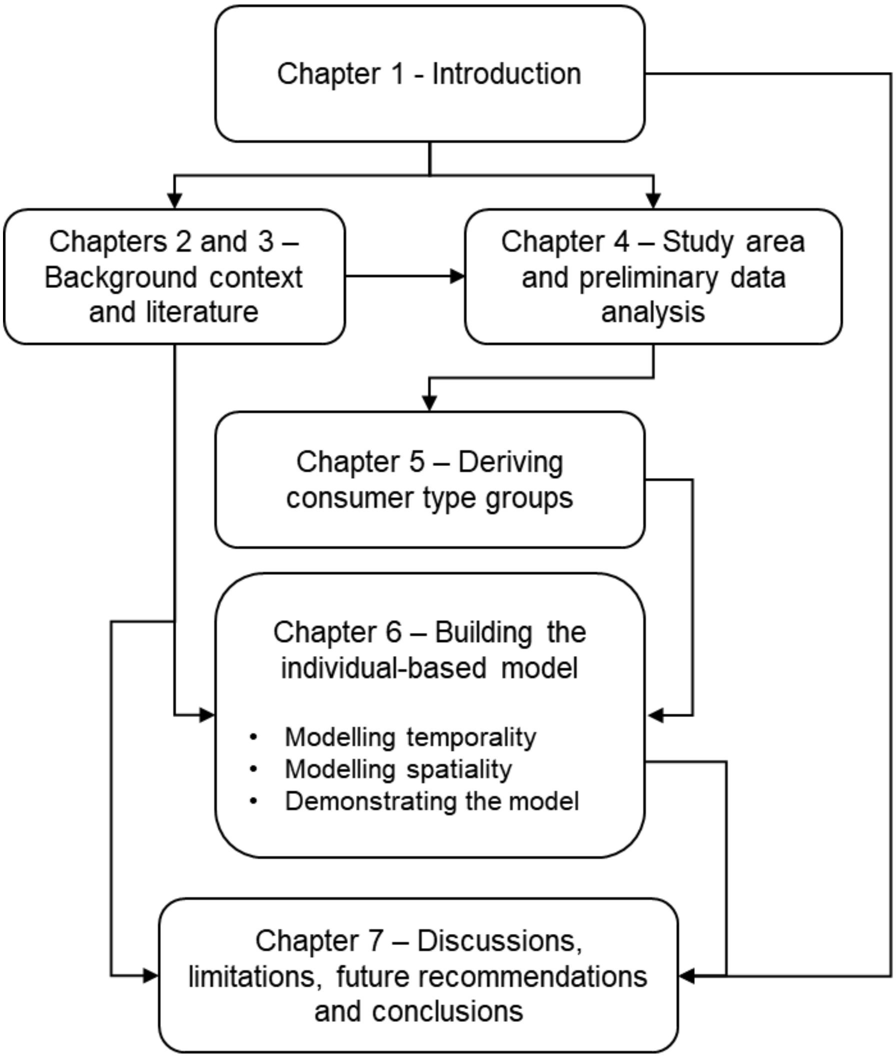



**Figure 1.1** Thesis structure

To meet the aims and objectives discussed in section 1.2, the thesis structure presented in Figure 1.1 was followed. Firstly, Chapter 2 and Chapter 3 establish the research in existing literature, emphasising the context in which this study is situated. As with all grocery retail studies in geography, Chapter 2 focuses on the changes in the grocery landscape over recent decades, reviewing the literature on the evolution of the British grocery sector. This chapter also relates these changes to the evolution of consumer behaviour, highlighting the impacts of convenience culture, the proliferation of e-commerce, and identifies the key indicators of consumer behaviour. The literature discussed highlights some of the most essential factors of grocery purchasing behaviour and is used to provide context during the analysis of the loyalty card-linked transaction dataset.

Chapter 3 focuses on the past and present methodologies used in location planning analytics. The chapter explores location planners and consultants' most used tools and practices, highlighting their key benefits and limitations. Alternative methods, such as IBMs and their application in grocery retail, are discussed. From these discussions, the prototype model developed in this thesis is situated in the literature, highlighting which features of different methods can be used and incorporated to help tackle the limitations of other models.

Chapter 4 presents the case study area of West Yorkshire, providing a summary of the area's profile, geodemographics, and other datasets, which are not incorporated into the model but provide context. This chapter introduces the loyalty card-linked transaction dataset provided by the study collaborator Sainsbury's. This novel dataset contains 2.7 million individual transactions recorded between May and July 2018 at all stores within West Yorkshire by those who reside within the county. The recorded transactions were performed by 216 thousand individual cardholders across 55 different stores, with each customer exhibiting unique and personal consumer behaviours. This chapter provides an overview of the dataset, including the types of transactions made, how far customers travelled to stores, and how often. The analysis in this chapter is used to support the segmentation of consumer type groups in Chapter 5.

Chapter 5 outlines the steps and approaches used to derive consumer type groups from the loyalty card dataset, including data cleansing, variable aggregation, and k-means clustering. This chapter discusses the challenges of working with such a large dataset and the rationale for the chosen methodology. In this chapter, seven unique consumer type groups are identified, in which their behaviours characterise each group regarding shopping frequency, channel choice, store choice, and basket type choice.

Chapter 6 provides context to the modelling approach developed for this thesis. A hybrid approach was established which incorporates features from ABM, MSM, and decision trees. Due to the model being the first of its kind in this application and the use of a significant dataset, the “keep it simple, stupid” approach was followed, adding complexity as and when required in the model building process (Edmonds and Moss, 2005). The goal of this chapter was to establish an intuitive framework for an IBM that simulates individual customers transactional behaviours that directly related to which consumer type group they belong to. The chapter covers incorporating the temporal and spatial behaviours of consumers into the IBM, discussing the unique challenges and achievements found along the way. A short example case study is presented which discusses how the model could be used to perform scenario-tests in the future. This example illustrates the potential uses of IBM in location planning analytics in such ways that are not possible in current methods.

Finally, Chapter 7 discusses the findings and limitations of the study presented in this thesis and provides recommendations for future work. The chapter concludes by discussing the extent to which IBMs can be used within grocery location analytics, the expected challenges, and the contributions this work has made via the attainment of the thesis’ aims and objectives.

## 1.4 Thesis contributions

The outcomes of this thesis exemplify the importance of collaborative research between academic institutions and the private sector. The access to proprietary data provided by the collaborator allowed for an in-depth study of the true behaviours of known grocery customers, a rarely available opportunity. In exchange, along with their industry knowledge, study collaborators are provided with a wealth of knowledge from the research output and a bespoke modelling framework ready for expansion and implementation. The research presented in this thesis is the first documented attempt at building an IBM framework of this calibre for the application of modelling consumer behaviour, both temporally and spatially, in grocery retail. Due to the study being the first attempt, unique challenges were faced, and are documented, identifying specific areas of development required to enhance the model, notably regarding modelling spatiality. The findings produced have the potential to support location planning experts in expanding their current toolset for modelling consumers and are already being used to develop bespoke IBMs for clients with loyalty card data.

The research presented also benefits the academic literature, filling the research gaps in grocery retail modelling. Specifically, this thesis provides insight into a notable gap in current spatial models: the inability to model customers who exemplify different transactional behaviours; current models assume that all customers behave homogeneously. The findings of this research identify key consumer typologies and their unique consumer behaviours regarding where, when, how, and why they transact for groceries. These consumer type groups were then implemented into an IBM, accurately simulating their past behaviour by adding behavioural stochasticity and heterogeneity and their relationships to store choice. The output from this study provides a deeper insight into how customers of a specific retailer behaved over a 3-

month period and can be used to better understand the temporal relationships between individuals and stores, which has been notoriously difficult in this area of research. These insights can be used to perform various scenario tests, from changing populations to changing consumer behaviours, changes in spatial demand, and supply-side changes. This research enhances our understanding of grocery consumer behaviour, providing a framework to repeat the study for other geographical areas, customers, and time periods.

Thus far, the first journal article has been written with the support of Andy Newing and Tim Rains, focusing on inferring consumer type from loyalty card data, and is awaiting submission to the Journal of Retailing and Consumer Services. A second paper is being written regarding the development of the modelling framework using IBMs for simulating consumer spatiotemporal behaviours. The journals considered for submission are the Journal of Artificial Societies and Social Simulation (JASS) and the International Review of Retail, Distribution and Consumer Research. Presentations were given at the GISRUK conferences during and prior to the Covid-19 pandemic, focusing on the visualisation of flow data, and the segmentation of consumers based on store and choice behaviours. Outside of academia, these novel research findings have been presented at in-house company events and will likely be presented at the Society for Location Analysis, a respected group of industry practitioners, and potentially at the next Spatial Data Science Conference, a high-profile event focused on commercial sector spatial data science.

This thesis has contributed to a variety of areas of research, including retail consumer behaviour and expanding retail location models. It has generated interest from the industry, including the study collaborator and more. The research has identified areas for further development, particularly in enhancing the spatial element, in which the strengths of other modelling methodologies can be implemented, such as those in SIMs. Additionally, more extensive scenario tests can be modelled, further demonstrating the true strengths this research provides.

# Chapter 2 Evolving consumer behaviours

The overarching goal of this research project is to refine spatial grocery models of consumer behaviour using an individual-based approach. To undertake this study, a thorough understanding of all facets of the study is required, notably what those spatiotemporal behaviours of grocery consumers are and how they have evolved over the years, what models have been developed and are being currently used to analyse the spatiality of grocery consumer behaviours, and how an individual-level modelling approach can be utilised in this application.

The following sections of this chapter, along with Chapter 3, aim to achieve the first aim outlined in section 1.2:

*To explore the literature around British grocery retail over the past decades, identifying key changes that impacted consumer behaviour and the key indicators of consumer behaviour.*

To achieve the study's first aim, this chapter is split into three key sections and attains objective 1:

*To explore the literature around British grocery retail over the past decades, identifying key changes that impacted consumer behaviour and the key indicators of consumer behaviour.*

Section 2.1 provides a context of the UK grocery sector, outlining what it supplies to consumers and how it has changed over the years. Section 2.2 explores the demand side changes observed in British grocery retailing. It focuses on how consumer behaviours have evolved regarding shopping missions, the channels they use to transact, and how shopping frequency varies. Section 2.3 identifies the key indicators of consumer behaviour associated with grocery retail and how these impact the models designed for retail location analytics.

## 2.1 The UK grocery sector: then and now

To model consumer store and channel choice behaviours, a thorough understanding of both the supply and demand sides of grocery retail is required. As Burt et al. (2010: p.174) state, “Any attempt to understand the UK retail scene should start by recognising the distinctive features of this market”. Therefore, the following paragraphs of this section outline the major events that occurred in the grocery retail market over the past few decades in a UK context. The grocery industry is one of the UK’s most critical and thriving sectors, worth approximately ~£216bn in 2022 (IGD, 2022). It is one of the largest private sector employers and property owners represented by some of the largest international retail firms (Cuthbertson et al., 2012). The UK had once been described as “a nation of shopkeepers” by Napoleon Bonaparte, but this characterisation is inapplicable to today’s grocery retail climate (Free, 2008). The UK grocery sector has undergone significant changes over the past decades, with various supply-side factors and customer shopping behaviours shaping the industry (Burt and Sparks, 1994; De Kervenoael et al., 2006; Elms et al., 2010). The following paragraphs briefly overview the key changes in the UK grocery market over the past few decades and explore the modern grocery retail landscape. The first few sections cover what is commonly referred to as the ‘Golden Age’ of UK grocery retailing (Jones, 1981; Akehurst, 1984; Wrigley, 1987; Sparks, 1990; Burt and Sparks, 1994), in which the grocery sector went through most of its radical developments between the 1960s and 1990s. During this period, notable food retailers dominated the market in terms of their concentration and market power, which are still prominent today. Pommering (1979 cited in Harris and Ogbonna, 2001) identified three notable shifts, recognising ‘manufacturers as king’ in the 1950s, to ‘consumers as kings’ in the 1960s, to finally ‘trade is king’ in the 1970s. These moving times are primarily due to the Resale Price Maintenance Act (RPM) abolishment (Harris and Ogbonna, 2001), followed by the increase in competition (Langston et al., 1997) and the growth of major multiples (Akehurst, 1984), which are further discussed below.

### 2.1.1 The ‘golden age’ of UK grocery retail

#### 2.1.1.1 From independent greengrocers to supermarket enclosures

Before the advent of supermarkets, the UK grocery industry was primarily dominated by small independent retailers and food co-operative stores, such as butchers, bakers, and greengrocers, who largely operated on a cash basis (Akehurst, 1984; Clarke, 2000). By 1948, supermarkets began to emerge in post-war Britain, offering a more comprehensive range of products, often at lower prices. In 1950, supermarkets accounted for 20% of the grocery market share, with the remaining 80% owned by a mixture of food co-operative and independent shops (Blythman, 2004). The introduction of supermarkets allowed customers to have a one-stop shop where they could purchase not only meats, breads, and fresh produce, but also general merchandise without having to visit multiple stores (Harris and Ogbonna, 2001). At this time, supermarkets developed to become more commonly self-service based, where customers can pick the items they desire off a shelf and pay at a till with a cashier (Poole et al., 2002b). Instead of only being able to purchase food items stores from specialised stores, this is the first time customers were provided options and *choices* regarding where they purchased their groceries.

#### 2.1.1.2 Price wars: may the cheapest retailer win!

Since modern supermarkets had become more common, supermarket chains and smaller independent grocers, entered a period of intense competition to attract the most customers (Akehurst, 1984; Wrigley, 1987; Seth et al., 1999). The competition was further proliferated by the abolishment of the RPM in 1964, impacting all stakeholders in the retail industry, from retailers and suppliers and more (Wrigley, 1987; Harris and Ogbonna, 2001). The act was originally in place to control the UK market artificially, in which retailers did not have the power to alter the price of the goods they sold, thus reducing their competitive advantage. However, as the act was repealed, retailers now took advantage of using product pricing as a competitive tool to enable growth, resulting in the expansion of numerous early retail grocery multiples. During this time, grocery retailers managed their product pricing to gain market share, which is still observed today as the market became dominated by a few select retailers, each commanding a sizeable market share (Harris and Ogbonna, 2001; Poole et al., 2002a). Largely due to the abolishment of RPM, the UK grocery sector entered a period coined as the ‘price wars’ (Wrigley, 1992), in which there was increasing competition between retailers mostly focused on price and value from the consumer perspective (Burt and Sparks, 2003). Another key driver of the grocery price wars began in the 1970s when there was increasing popularity of British discount retailers such as Kwik Save and Fine Fare (Sparks, 1990; Burt and Sparks, 1995). These retailers could offer customers significantly lower prices than their competitors by adopting the strategies observed in successful European markets. These strategies included cost-cutting in their stores, such as no-frills stores with narrow aisles, selling their own brand of food items at an extremely low cost, and selling a limited range of goods on favourable payment terms with their suppliers (Sparks, 1990). In reaction to the low-cost competition, other larger established retailers began to compete in their prices, engaging in aggressive price-cutting strategies. This resulted in a significant decrease in profit margins for many retailers, with some smaller independent stores being forced out of business, largely due to the competition from larger multiples (Clarke, 2000; Burt and Sparks, 2001).

Whilst grocery retailers faced negative impacts due to competition, customers now had a wider variety of supermarkets to visit and a *choice* regarding how much they were willing to spend. Customers could purchase well-known brands at a reasonable price or opt for discount retailers and their budget offers depending on their lifestyles (Knox and Denison, 2000). The price wars also resulted in the further development of new grocery retail formats, such as hypermarkets which combined large grocery stores with other retail products, such as clothing and electronic goods (Jones, 1981; Langston et al., 1997; Bevan, 2006). This allowed customers to purchase groceries from a supermarket and other items in a one-stop-shop shopping trip (Pacione, 1979). The increased competition between retailers contributed to the growth of private label or store brand products, with most major grocery retailers now offering their own budget version of foods alongside those offered by major manufacturers. With most supermarket chains now offering a similar variety of food and drinks that have become the standard and at a reasonably similar price, customers require more incentive to choose a particular retailer rather than competitors. Therefore, customer loyalty schemes become more prominent amongst grocery retailers, as further discussed in section 3.3.1.

#### 2.1.1.3 If you can’t beat them, buy them.

Since the introduction of modern supermarkets, the grocery market has become increasingly consolidated, with a small number of large players dominating the sector. This has been partly driven by mergers and acquisitions (M&A), with larger grocery retailers acquiring smaller rivals, which has a major impact on the industry. With large-scale grocery chains reducing their prices and increasing in popularity and number of stores, smaller supermarket chains and independent retailers struggled to keep up with competition and reducing profit margins. Many smaller chains began to be acquired or merged with larger supermarket chains, leading to an increased market concentration and reduced competition in the grocery sector (Wrigley, 1987; Poole et al., 2002b). M&A activity has also been driven by changing consumer preferences and the need for retailers to adapt to changing market conditions. The consolidation of these retailers had a large positive impact on the overall grocery sector. With an increased market size, larger retailers can achieve economies of scale and increase efficiency, potentially leading to cost savings passed onto consumers through lower prices and shopping incentives. M&A can also drive innovation and diversification within the grocery sector, with retailers acquiring companies to expand their offerings to markets they have not entered before or which their previous stores could not reach (Burt, 2000).

#### 2.1.1.4 Location, location, location

Alongside the abolishment of the RPM, living standards had also been increasing as the UK entered a post-war recovery period. Consumers across Britain became more affluent, allowing those to upgrade their appliances and become owners of personal vehicles. The increase in car ownership allowed consumers to expand their choices in *where* to shop and were no longer restricted to visiting their closest supermarkets and grocery stores (Guy, 1997). Retailers then faced a new problem of being the most attractive store regarding *prices* but also in terms of *product variety* and *accessibility* (Elms et al., 2010). As consumer demands became more versatile, retailers began to develop even larger store formats with more diverse product ranges and started to locate these stores in out-of-town locations (Harris and Ogbonna, 2001; Burt and Sparks, 2003). As discussed in Wrigley (1992), placing these stores away from the city was a strategic choice by retailers as more consumers began to move to the suburbs. The acknowledgement of the importance of *location* created a new type of battle for retailers, this time a 'race for space' (Wood and McCarthy, 2014), also known as the 'store wars' (Wrigley, 1991). Major retailers began to reduce the number of their smaller stores and began to build larger, more profitable supermarkets and hypermarkets. In many cases this resulted in the closure of their town and city centre supermarkets in favour of larger replacement supermarkets located out of town. During the 1980s, the UK government approved the creation of 'Enterprise Zones', which encouraged the development of businesses across Britain in specific areas by combining tax breaks, lighter regulation, and more straightforward planning rules for retailers. Grocery retailers aggressively tried to purchase the best locations to build their stores, along with aggressive 'land banking' strategies for future expansion and felt the pressure of other retailers to increase their market share. Both the price wars and store wars together resulted in an incredibly concentrated UK grocery market, with only five grocery retailers owning over 50% of the market share (Langston et al., 1998).

### 2.1.2 From the 'Golden Age' heydays to the dog days

Upon entering the 1990s, those within the British grocery retail sector began to face difficulties from all sides, from property and planning policy changes to market saturation and new competitors. During the 'golden age', the grocery sector experienced the emergence of major grocery retailers who strengthened their market power via store network expansions, resulting in large profit increases. These major supermarket chains caused the sector to be highly concentrated, producing a grocery market share split between only a few retailers. The previous period of intense growth of major multiples was feared to be unsustainable, as that growth resulted from multiple market factors that produced an increasingly competitive environment (Wrigley and Lowe, 1996). The following subsections discuss the prominent issues the UK grocery retail market faced from the 1990s and their impacts on consumer behaviour. These issues include covering the new restrictions placed in planning policy, the fear of retail saturation, and the arrival of European discount retailers.

#### 2.1.2.1 Deep discounters from Europe

By the early 1990s, additional competitors entered the UK grocery retail landscape, notably the German (Aldi and Lidl) and Danish (Netto) discount retailers (Burt and Sparks, 1994). These retailers predominantly established themselves in areas of deprivation in the north of England and the West Midlands. They have since expanded across the UK, increasing their grocery market share and geographical coverage (Thompson et al., 2012), including expansion into more affluent areas. Their main aims were to fill two critical retail supply gaps: value by providing highly competitive low prices, and spatiality, by locating in deprived inner city locations. By 1996, deep discounters became a notable part of the UK grocery market, directly competing with Kiwksave’s business strengths (Poole et al., 2002b) and their smaller profit margins compared to the leading grocery players. The introduction and permanence of these deep discounters increased competition between retailers, as customers now had even more choices between stores. The combination of store location and consumer budgets now played an even more prominent role in grocery competition.

#### 2.1.2.2 Planning Policy Guidance, the government is ruining all the fun!

To build new store developments in the UK, local planning authorities must adhere to national planning policy guidance (PPGs), which state the requirements for locating new store sites. During the 1980s, planning policies were lax, allowing retailers to locate where they desired during the space race. However, in 1998, Planning Policy Guidance 6 (PPG6) was introduced, creating much stricter planning regulations (Wrigley, 1994). At first, the PPG6 allowed and encouraged grocery retailers to build extremely large supermarkets out-of-town to reduce traffic into town centres and provide accessible food stores from multiple areas (Wrigley, 1987). However, by 1993, the PPG6 was revised to protect and promote town centres by disallowing such a high number of superstores to be located out-of-town to increase retailer market share. Essentially, building these large format grocery stores out-of-town was a last resort, and retailers should focus on building closer to town centres utilising ‘within-centre’ or ‘edge-of-centre' sites, uplifting consumer traffic to towns, and reducing car use (Guy, 1997; Wood et al., 2006). The encouragement of building within town centres coincided with consumer behavioural changes, as convenience began to play more significant role in customer behaviour.

#### 2.1.2.3 Now that’s a conveniently placed grocery store

Due to policy changes the development of out-of-town stores slowed down by the 2000s, and grocery retailers began to diversify their store types (Wood et al., 2006). Retailers found benefit in developing new, smaller store formats and placing them in places of consumer convenience, such as within town and city centres, near train stations, and within residential neighbourhoods (Wood and Browne, 2007). By locating in these areas, these stores were expected to experience distinct types of consumer demand at varying points in the day (Waddington et al., 2018). Tesco and Sainsbury's were key grocers at the forefront of convenience retailing and focused on developing convenience-focused stores whilst maintaining their larger superstores (Guy, 2007). These smaller, convenience-based stores provided customers with limited product ranges and an in-store environment more suitable for 'urban markets' (Burt et al., 2010). Wood et al. (2006) note that customer expectations of store standards were raised in previously 'fragmented' markets characterised by poor services, unreasonable prices, and limited ranges typically offered by independent convenience stores. Around the 2000s, many consumers began to prefer shopping for groceries locally and more frequently, identifiable as 'top up' shopping. Not only did retailers now compete with store locations, pricing, and customer accessibility, but new forms of consumer preferences were also emerging, notably around convenience (further discussed in section 2.2).

#### 2.1.2.4 Sainsbury's join the convenience market

In 1998, Sainsbury's introduced their "Local" brand stores, their first venture into the newly recognised convenience market, catering to 'on-the-go' and 'top-up' basket types (Sainsbury's Archive, 2019). Sainsbury's Locals are generally located in urban and suburban areas, providing those in town centres with access to convenience-based foods. The variety of foods available in Local stores is much smaller than in supermarkets, focusing on the types of items that customers need in-between 'main' shops. To expand their geographical coverage in the convenience market, Sainsbury's acquired a variety of convenience shop chains in 2004 and 2005, rivalling Tesco's expansion of Tesco Express stores (their convenience-based stores) in 2002. Sainsbury's acquired Bells Stores for northeast England coverage, Jacksons Stores in Yorkshire and north Midlands, JB Beaumont in Nottingham, and SL Shaw stores in southeast England (RetailWeek, 2005). These stores initially had a variety of brand names but were all rebranded to Sainsbury's Local by 2007 (conveniencestore.co.uk, 2007).

In 1999 Sainsbury's announced another new alternative store format named "Sainsbury's Central", carrying the same foods found in their Local stores but with an additional selection of food from their supermarkets (Campaign Magazine, 1999). The Central stores were mid-sized and were aimed to be located in town centres and commuter areas, essentially rivalling Tesco's "Metro" store format launched in 1994 (Owen, 2003). These stores were aimed to provide a more comprehensive range of foods, allowing those who live within the city to purchase larger shopping baskets than Sainsbury's Local's could offer. However, these stores were later rebranded as 'Local' stores as Sainsbury's made efforts in to streamline their store strategy by making it clear they offer two store choices: supermarkets, catering for all basket types (especially 'main' baskets), and convenience stores for any basket type that is not considered a 'main' basket. The same is true of the Tesco Metro format, which was subsequently rebranded as part of their Tesco Express proposition.

More recently, in 2020, Sainsbury's Supermarkets launched a store format categorised as 'Neighbourhood Hub', providing a "one-stop offer" to neighbourhood communities (Sainsbury's.co.uk, 2020). These stores are around 5,000 ft$^2$ and 7,000 ft$^2$, being a middle-ground between their Local format types and Supermarket format stores (10,000 ft$^2$ to over 60,000 ft$^2$). The larger Sainsbury's Local stores have been rebranded and expanded to fit the Neighbourhood Hub format, with other smaller stores being completely refurbished. Neighbourhood hub stores are aimed to be accessible to neighbourhood and suburban areas, offering more services that are not currently offered in their Local branded stores (Sainsbury's.co.uk, 2023c). These services include Argos collection and Tu clothing collection points, value-added foods, and other household items outside groceries (Sainsbury's.co.uk, 2020).

Sainsbury's has become one of the major players in the convenience sector and is continually changing their store network, especially with the recent introduction of Neighbourhood Hub stores. Whether Neighbourhood Hub stores will be as successful as their Local brand is still yet to be determined, but with their increased services, and a wider variety of food products and services, Sainsbury's might take the lead over other retailers in the convenience market. Sainsbury's provides a variety of store types from Locals, Neighbourhood Hubs, Supermarkets, and their extra-large supermarkets, sometimes referred to as Superstores. Sainsbury's has a comprehensive store network, including e-commerce services in which they both deliver groceries and allow customers to use click and collect. Although not thoroughly discussed in this review, Sainsbury's provides non-food products in their stores thanks to their acquisition of Argos, a catalogue retailer, and Habitat, a furnishings retailer, in 2016 (Guo and Wang, 2019). Their non-food service provision is integral to their attractiveness to customers; however, the data provided in this study only accounts for grocery purchases. Future considerations of modelling such stores should include other services provided, as these will attract non-grocery consumers to the brand who may become grocery customers as footfall into the stores increases.

### 2.1.3 You can buy groceries *without* leaving the house?

By 2010, grocery retailers began incorporating e-commerce into their channels for consumers. Grocers realised that via the internet, they could provide goods to their customers straight from the store to the home using delivery vans. Providing a new shopping channel gave retailers insight into expanding their customer base and market share without opening new stores and making the most of current ones (Clarke et al., 2015). However, providing a new channel presented new challenges. These challenges included where to set delivery boundaries from stores (i.e. determining the spatial extent a given store is able to serve), deciding which supermarkets have the capacity to offer online orders, and whether current stores are suitable for doing so, and questioning the need for dark stores and online fulfilment centres (warehouse based stores used for online order fulfilment) (Kirby-Hawkins et al., 2018).

Additionally, how could the retailer provide online orders for far away customers, and who should pay for delivery costs and within what range? Retailers had to decide whether e-commerce would benefit the company overall in terms of profit and consider multiple aspects such as demand, competition, capacity and cost. Anderson et al. (2003) highlight that customers who reside in rural areas and away from supermarkets are most likely to demand e-commerce channels; how can retailers balance the scales between distance, cost, and provision to such areas? Many retailers have been reluctant to provide e-commerce grocery deliveries to rural areas, as they base their e-commerce network around their existing (predominantly urban) supermarket store estate store sites (Newing et al., 2022).

Online grocery delivery provision is commonplace today (Newing et al., 2020) and grew significantly during the Covid-19 pandemic (Meister et al., 2023). The unprecedented demand for online caused strain on multiple grocery retailers, as orders could not be fulfilled or delivered fast enough due to lack of capacity within the order fulfilment system (Pantano et al., 2020). The increase in on-demand grocery during the UK lockdowns had become known as ‘quick commerce’ (Q-commerce), where around 150 hubs were set up directly by food suppliers to provide customers with on-demand food deliveries (Sleeman, 2022). However, since the UK’s cost-of-living crisis, analysis has found that customers are now less likely to choose the extreme convenience of quick commerce and return to old money-conscious shopping habits (Rigby, 2022).

Online grocery shopping has been a growing channel since its introduction to the UK grocery market and is expected to continue its rising increase in consumer demand (IBISWorld, 2023). The demand for online delivery peaked during the Covid-19 pandemic but has since returned to regular growth rates as customers return to everyday life without the restrictions of the UK lockdown (Meister et al., 2023). The unprecedented need for online during the pandemic brought an entirely new challenge to grocery retailers, as they faced issues with product availability, staffing, and scheduling of deliveries (Boyle et al., 2022). For grocery retailers to be better prepared for such unique circumstances, a thorough understanding of their consumers, and rigorous scenario testing that includes the temporal and spatial aspects of their customers is required.

A summary of the above sections on grocery supply changes is provided in section 2.4; the following section of this chapter focuses on the demand side changes.

## 2.2 Demand side of consumer behaviours

Section 2.1 provided context of the supply side developments in the UK grocery sector since post-war Britain. The following sections are concerned with the demand side of consumer behaviour in grocery retail, which has evolved over recent years due to societal and supply-side changes. By setting the scene for both the demand and supply side of grocery retail, we can gauge an understanding of why customers behave the way they do, including their store and channel choice behaviours.

Over the last century, the experience of grocery shopping has changed greatly, from shopping at specialised food shops (butchers, greengrocers, and bakeries, for example), to self-service supermarkets, and the evolution of the habitual weekly food shop - often undertaken at the same retailer - thus expressing brand loyalty (Wood and Browne, 2007). The assumption that retailers will perform a weekly shop at the same retailer enabled location planners to efficiently make store location decisions. However, over the past decade, with the emergence of online shopping and cultural reforms (including the impacts of Covid-19 on mobility patterns), consumers have been performing unpredictable shopping habits. Consumers today may not partake in the weekly shop and instead shop as and when it is most convenient for that individual or shop online (Jackson et al., 2006; Hood et al., 2016). This shift in consumer behaviour has increased the complexity of grocery location planning, thus posing model calibration challenges for spatial interaction modelling and retail analytics (discussed further in section 3.2). In response to these challenges, there has been development the of SIMs in retail applications, such as model disaggregation that incorporates e-commerce, discount stores, and convenience retailing (Newing et al., 2015; Waddington et al., 2018; Hood et al., 2021). Despite these developments, it is integral that retail location planners consider alternative spatial modelling methods that reap the benefits of spatial interaction modelling and can capture consumer behaviour at the individual level, especially with the abundance of consumer data being captured each day (section 3.3). The following sections provide an overview of the changes in grocery consumer behaviour over time and space.

### 2.2.1 The relationship between time and geography

Consumer grocery shopping behaviour is complex and influenced by various factors, including convenience, price, product quality and variety, social and cultural factors, and technology. Understanding these factors is essential for retailers and marketers to develop effective marketing strategies and product offerings that meet the needs and preferences of consumers.

Many studies within grocery location analytics focus primarily on the spatial aspect of a store network, especially studies using modelling techniques. At the top level, a store's location is one of the most important factors to determine how well it can perform in terms of sales and accessibility (Birkin and Clarke, 2012; Clarke and Birkin, 2018). Other factors include the store's product range, size, whether it has a car park, a café, etc. In the tools and models used within location analytics, an element of temporality is often considered via the incorporation of different demand layers, signifying the potential consumer population around a store at different times (Newing et al., 2014; Newing et al., 2015; Waddington et al., 2019; Rowe et al., 2022).

The incorporation of time-based demand into location analytical tools can be linked to the classical theoretical framework of 'time geography', first coined by Hägerstrand (1970). Time geography is a constraints-oriented approach to understanding human phenomena and aims to understand and analyse the spatial and temporal dimensions of human activities and movements (Miller, 2008). In time geography, the concept states that what we do as humans in space is inherently constrained by time. As we live our lives, we make choices about where we go and when we go, which are all linked to purpose. Our daily activities are constrained by time, and Hägerstrand (1970) identified three primary constraints that shape our behaviour and movement: capability constraints, coupling constraints, and authority constraints. These three types of constraints impact our lives as consumers, as time and place impact our ability to purchase goods. Particularly, 'authority constraints', in which our access to a location is only available at specific times, impacts us as consumers as we can only make purchases if a store is open (Miller, 2008). If a store is not open, then we are limited in both time and space to perform an activity. ‘Capability constraints’ are related to an individual’s physical limitations, such as age, health or disabilities, impacting their ability to travel long distances, essentially impacting their mobility and accessibility. ‘Coupling constraints’ are related to the temporal co-dependencies of different activities; for example, a person may need to catch a train to travel to work; therefore, that action is coupled with the train’s schedule, meaning the person may have to leave their home at a specific time to catch that train to then reach work.

#### 2.2.1.1 Trip chaining and convenience culture

In the past, literature on consumer behaviour primarily focused on the principle of least effort, as in people always chose the closest available store regardless of their location. This principle was heavily interlinked with the Central Place Theory (CPT) (Getis and Getis, 1966) and Tobler’s First Law of Geography, in which he states “Everything is related to everything else, but near things are more related than distant things” (Tobler, 1987). However, those in the field of location analytics acknowledge that individuals do not always perform actions locally and are likely to perform multiple trips in a journey, known as trip chaining. Trip chaining is the practice of linking multiple trips to form a single journey, creating a chain of activities connected in time and space (Primerano et al., 2008). Trip chaining helps to explain how individuals organise and coordinate their activities to optimise their time. For example, a person may plan to visit a friend after work and would like to purchase dinner on their way over. To perform this sequence of actions, the individual will decide how to chain their trip events, integrally linked to both space and time. They might choose a different route from work to their friend’s house to purposefully pass a particular grocery store to buy dinner. Alternatively, they might take their usual route to their friend’s due to bus times and purchase dinner from any store close to their destination.

Understanding the concept of time geography and trip chaining together is essential in transport studies and heavily impacts other areas, such as urban planning and consumer store choices (Thill and Thomas, 1987). A closely linked concept to trip chaining and time geography is ‘tours’. Tours are sets of linked trips taken by individuals or groups to complete specific activities; trip chaining can be seen as a set of tours. Common tours are home-to-home loops and are often studied in transport to understand an individual’s actions outside the home (Daisy et al., 2018). An example of a home-to-home loop is a person travelling to work from home and then returning straight home from work. These simple tours occur often, but more complex trip chains, such as trip chaining between work, a grocery store, and a friend’s home, occur. Simple tours can be modelled relatively easily when a person’s home and work are known, but spatially modelled tours of non-workers are incredibly complex, as highlighted in Daisy et al. (2018), due to their flexible activity schedules and unique behaviours.

#### 2.2.1.2 Shopping missions, store choices and time

Along with trip chaining and complex tours come multi-purpose shopping trips. Before the early 2000s, research found that customers mainly purchased their groceries in a single weekly shop on a habitual basis and often at the same retailer (Popkowski Leszczyc et al., 2004). The big weekly shop, often done at the weekend, became a common habit of most British families to purchase enough groceries for the whole family (Humphery, 1998). The increase in large out-of-town supermarkets, built due to policy guidance (section 2.1.2), allowed customers to undertake these large transactions at one store that provided everything the customer needed. These predictable behaviours allowed retailers to understand when demand was due in their stores and the types of transactions that would take place. As time passed, consumer lifestyles shifted as convenience became a significant part of our lives (Elms et al., 2010). From a demand perspective, the increase in consumers exhibiting convenience-based grocery transactions is often attributed to the rise in young, single-person households of professionals who work longer hours and are considered 'money-rich, time-poor' (Hood et al., 2016). Alongside these changes came an increase in female workforces, a move to more time-centric lifestyles, and increased personal mobility (Vanhaverbeke and Macharis, 2011), further proliferating the need for convenience (De Kervenoael et al., 2006).

The increase in convenience culture can also be attributed to the supply-side changes by the "big four" grocery retailers (Tesco, Sainsbury's, ASDA and Morrisons). As part of the PPG6 (section 2.1.2), the 'Town Centre First Policy' was introduced (DoE, 1996), which restricted the opening of new retail and other traditional town centre activities to town centre locations. This policy was deemed necessary to reduce the decay of the British high streets and maintain busyness after the proliferation of out-of-town supermarkets that were previously encouraged. Thus, many retailers began to acquire smaller independent stores within town centres to open small-format supermarkets. This resulted in a significant increase in convenience stores opening between 1999 and 2008 across England, providing customers with more shopping destinations and increased accessibility (Cheshire et al., 2022). The encouragement of building convenience stores in town centres via policy not only provided the "big four" with the opportunity to expand into new formats but also provided customers access to groceries in more convenient locations. By 2008, online grocery shopping began to grow as some customers swapped their larger in-store transactions for online transactions alongside their in-store convenience-based transactions (Harris et al., 2017).

The increase in the convenience-based lifestyle saw an increase in various shopping missions (also referred to as basket types throughout this thesis). The large weekly shopping transactions, also known as “main” baskets, were still prevalent in consumer behaviour, but “top-up” baskets became more common. Top-up baskets are grocery transactions that contain food for a few days, containing some fresh foods and shelf-stable foods. These transactions primarily occur at supermarkets or larger convenience stores due to the variety of available foods. Kahn and Schmittlein (1989) found a correlation between shop trip timing and store choice regarding basket type. Main transactions tend to occur on weekends at a customer's favourite supermarket, whereas top-up shops occur anytime at any store. The relationship between basket types and transaction temporality provides great insight into consumer behaviours of customers and their store choice (Popkowski Leszczyc et al., 2004). The third type of transaction is identified as "impulse" buys, or as referred to in this thesis, "for now" baskets. These transactions occur out of convenience, usually containing food to eat on-the-go or to be eaten soon. These convenience-based transactions often occur in inner-city shopping areas (Borgers and Timmermans, 1986). 'For now' transactions occur around places of work and are often associated with workers buying lunches such as meal deals.

There is a complex relationship between consumer grocery transactions and temporality, especially as consumer behaviours become more individualised. Accounting for residential-based demand, particularly ‘main’ baskets, are well understood by retailers. Workplace grocery demand is also acknowledged by retailers and incorporated into their store location models. However, non-residential or work-based grocery demand is difficult to capture, especially with the rise of convenience, trip chaining, and multi-purpose shopping trips. There is an element of stochasticity in consumer behaviour as non-habitual journeys occur, which are incredibly difficult to predict. The increase in convenience-based demand translates into a need for more sophisticated modelling tools for location analytics, as the current models today find difficulty in incorporating individualised demand (Vanhaverbeke and Macharis, 2011; Rowe et al., 2022).

#### 2.2.1.3 Money, society, and culture

Another important factor that has influenced differences in consumer grocery shopping behaviour is money. Consumers are often price-sensitive and shop at stores that offer competitive pricing. With the increase of discount retailers (section 2.1.2), consumers are now provided with a choice regarding whether they shop and the price they are willing to pay. This increased competition has impacted grocery retailers too, with supermarkets such as Sainsbury's and Tesco advertising price matches against Aldi (Leyland, 2022). Research has found that many customers are willing to switch brands if they perceive to be receiving a better deal elsewhere (Retail Connections, 2023). The deep discount brands Aldi and Lidl have recently experienced an influx of new customers. Considering the cost of living crisis impacting all areas of life, it is no surprise that customers have been changing their grocery purchasing habits (Consumer Intelligence, 2022). Around 83% of grocery shoppers are part of more than one grocery loyalty card scheme, reaping monetary benefits at most supermarkets they visit (Gonçalves and Nott, 2022). With money and budget becoming increasingly important in daily life, customers are willing to drop brand loyalty for a cheaper competitor (Retail Connections, 2023). The lack of loyalty to a particular retailer creates challenges for location planners as demand becomes more variable.

In addition to changing budgets, culture and society play an integral part in consumer store and channel choices. Over recent years, specialised food products have increased, such as vegan and vegetarian ranges and organic food preferences (Trewern et al., 2021). As customer preferences have changed, retailers have been adapting the variety of groceries they offer and promise to do so for a more sustainable future (Hughes, 2019). Some retailers now have a competitive edge over others by providing a more comprehensive range of foods to cater for multiple customers. Amongst these changes, there has been an increased preference for international foods, particularly in mixed ethnic areas. Research by (ABPL Group (2016) found that customers of south-east Asian descent will often transact at British supermarkets for their weekly main baskets but will go to local international food stores for the foods not stocked in supermarkets. This results in a proportion of transactions and resultant spend not being captured at specific retailers due to their limited range of foods.

#### 2.2.1.4 E-commerce

Finally, the number of consumers using online for their transacting has increased, with more customers engaging in the variety of alternative channels provided by retailers, such as online and click & collect (Kirby-Hawkins et al., 2018; Dolega and Lord, 2020). This increase in online transacting has come about with the increase in convenience culture and the Covid-19 pandemic. With instant access to stock checks, zero travel costs, and price comparison across multiple retailers, online provides customers with the easiest and least intensive form of grocery shopping (Hamad and Schmitz, 2019). Grocery e-commerce is one of the fastest-growing sectors in the UK, and is expected to continue growing (Birkin et al., 2017; Beckers et al., 2022). During the height of the Covid-19 pandemic, online shopping for groceries had grown unprecedently, causing disruptions in the online service provision of many grocers (Pantano et al., 2020). Since the pandemic, consumer channel choice behaviours have shifted, with more customers turning to online after avoiding retail premises during lockdowns (East, 2022). E-commerce will likely continue garnering attention from new customers as online becomes more convenient. Grocery retailers who provide multiple shopping channels must consider the future of online provision, as their store delivery networks may need adapting as new customers join. The concept of the ‘last mile’ in grocery delivery is costly for retailers, and with increased competition, they must be prepared to both adjust their delivery networks and consider how to afford delivery costs (Urquhart et al., 2022)

## 2.3 Key indicators of customer transaction behaviour

At the surface level, purchasing groceries can be as simple as a thought of 'I need to buy food, I will go to the supermarket'. However, a few factors are involved in making that transaction based on the customer's needs, preferences, and choices. Customer transaction behaviour refers to the patterns of behaviour and the decision-making process we go through to make a purchase, and these can be summarised into key indicators.

A customer will determine a need to transact before performing any transaction, whether that transaction is for a snack on-the-go, to pick up a few things from the shop such as bread and milk, or a large transaction containing food for multiple days. Alongside a need to transact, a budget is also considered, shaping the customer's choices. Some brands are too expensive, and the customer will seek elsewhere to make their purchase, or they have a limit on the amount to spend and will buy as much as they require within that budget. Alternatively, the customer has no budget limit and will buy the items they want. With the transaction need, there is a connected time to the transaction. If the customer needs their items imminently, then they will transact if no constraints are in place. If the customer needs the items but knows that they should wait until later for a better shopping experience, then their transaction timing or channel choice.

In some cases, the customer wants to purchase a food item, but there is no store nearby to visit or perhaps the closest store is not within walking distance and the consumer does not own a car. Therefore, the customer may not transact at all or will wait until another opportunity. All these considerations of transaction purpose interplay with choices that are both temporally and spatially linked. The grocery purchases we make are always based on choice, whether to transact or not to transact. If we do a transaction, then when, how, what, and where?

These individual factors make up the key indicators of consumer store choice: When we buy, what channel we use to buy, what basket type we buy, and where do we buy the products from. These key indicators and their relationships to one another provide businesses with insights into the consumer behaviours of their customers and help them tailor their strategies to better meet the needs and preferences of their target audiences. By analysing these indicators in conjunction with their customers' behaviours, grocery retailers can identify shopping trends, patterns, and areas of improvement in their business and can help make location-based decisions.

## 2.4 Chapter 2 summary

This first section of this chapter has focused on the notable events that occurred in the supply side of grocery retail, such as the increase in traditional supermarkets, the emergence of top players, the increase in competition from abroad, the race for space, the increase in convenience store provision, and the extension of grocery shopping to online. Other events also happened during this time; however, the most relevant for this thesis has been discussed to place the context for this study. Today, the UK grocery market provides consumers with various choices, providing access to food in areas of convenience (town centres, train stations, residential neighbourhoods) and large format superstores located along accessible road networks as guided by the PPG6. Customers now have more choice than ever regarding the brand of the supermarket they can purchase groceries from; from the big name brands that dominate the market share to more affluent retailers to the deep discounters, there is something for everyone regarding grocery expenditure. The addition of e-commerce has been a large development for grocery retail, providing unique challenges and questions regarding how to best approach the delivery of groceries and assign orders from stores for efficiency.

The increase in convenience provision by retailers has allowed consumers to access grocery stores from a place of convenience, inherently adapting the behaviour of customers. The second part of this chapter focused on the demand side of grocery retail and how these factors have changed over recent years. This section highlighted the interconnections between time, space, and consumer behaviours and their relationships to channel choice and basket types. As mentioned in this chapter, consumer behaviours have been heavily impacted by the rise of convenience culture, the proliferation of e-commerce, and the cost of living crisis. This chapter presented the key indicators of consumer behaviour which are used throughout this study. The literature discussed highlights some of the most important factors of grocery purchasing behaviour and is used to provide context during the analysis of the loyalty card-linked transaction dataset in Chapter 4.

# Chapter 3 Location Planning Analytics

This chapter focuses on the location planning analytics of grocery retail, highlighting the methods used in this sector, the theories involved, and the available data to help improve these methods. The following sections of this chapter, along with Chapter 2, achieve the first aim outlined in section 1.2:

*To present a review of the historical and recent changes in consumer behaviour in the context of grocery retail planning and the tools used in location planning analytics.*

To achieve this aim, this chapter attains objective 2:

*To articulate the benefits and limitations of current grocery location planning analytical tools, including the potential data sources available.*

In addition to exploring the literature, my professional experience in these areas of study is also discussed for a thorough and rounded understanding of spatial analytics in grocery retail. This chapter beings with an overview of the theories involved in spatial analytics before moving onto the commonly used modelling methods used by location planners. Section 3.2.3 introduces an alternative computational-based approach, agent-based modelling, that is rarely used in grocery location analytics. Section 3.3 concludes the chapter by suggesting data sources that could help build these individual-based models, especially for spatiality.

## 3.1 Background of store location planning

Since the 1950s, grocery retailers have been developing techniques for store location planning using sales forecasts and experience to identify optimal site locations (Guy, 1980; Clarkson et al., 1996; Clarke, 1998; Wood and Browne, 2007). These techniques were initially designed to understand where to open, close, and reformat stores to gain revenue, market share and estimate store performance. Over time, site location techniques have progressed due to improvements in technology and analytical tools, the growth of data, and an increase in budget size as grocery retailers recognise the importance of location planning.

These improvements have transformed from the reliance on intuition, checklists and comparative methods to using scorecards (regression-based models) and advanced spatial analytical tools such as Geographical Information Systems (GIS) and Spatial Interaction Models (SIM) (Reynolds and Wood, 2010). The progression of these techniques is primarily influenced by the incredible global growth of data volume and an exponential increase in computational power. According to the market intelligence company IDC, the ‘Global Datasphere’ had read approximately 18 zettabytes in 2018 (Reinsel et al., 2018). Most of these data have been generated in the last few years and are predicted to grow to around 175 zettabytes by 2025. These information-rich datasets often contain location data on population (size), people (purchase history and income), property (residential land or commercial), places of interest (tourist destinations, commuter neighbourhoods), phone data (insight into movement or person location), and many other areas. The synthesis and analysis of these large datasets are essential for location planners to understand the population dynamics of an area to make location-based decisions.

Along with increased computational power and data, the development of these techniques is often a result of collaborative work between grocery retailers and retail geographers. In this relationship, academics are often partly responsible for designing and developing a technique whilst the retailer provides rarely accessed (in the academic sector) datasets such as consumer transaction data. This partnership is essential as large retailer datasets contain location data on consumers at the individual level instead of an aggregate level, thus allowing the creation of spatial models at the individual level.

The evolution of retail location decision-making can be divided into four periods, where theoretical approaches, models, and data sources are used to support the decision-making process. In early location decision-making, choices were based on four key “cornerstone” theories of retail location (Brown, 1993), the Central Place Theory (CPT) (Getis and Getis, 1966), SIMs (Huff, 1963) the Bid Rent theory (BD), and Hotelling’s Law (Working and Hotelling, 1929). The first two theories, CPT and SIM, focus largely on the spatial behaviour of consumers, while the latter theories are strategy focused on choosing suitable retail locations.

According to CPT, towns are referred to as central places as they were attractive places to locate retail stores, drawing in consumers from surrounding areas. This theory assumes that all consumers perform identical behaviours and that these consumers will only perform single-purpose trips to the closest shopping centre (Clarke, 1998). The key element that CPT does not incorporate is the notion of multi-purpose trips, a behaviour that is very apparent today (as discussed in Chapter 2). CPT is based on the two assumptions that customers always shop at the closest place available, and if a service or good is in high demand, it will be offered close to where the population are based (Brown, 1993). Progressing from CPT and the 1960s quantitative revolution, modelling methods were developed in the discipline of geography, one notable method being Spatial Interaction Modelling (SIM). SIM acknowledges that the attractiveness of stores plays an important role in consumer store choices as much as distance (Roy and Thill, 2004; Clarke and Birkin, 2018; Birkin and Clarke, 2019); customer makes a trade-off between distance and how attractive the store is (such as its sales area, car park offering, closeness to other retail points). The idea is that these retail points draw in consumers from different areas based on these two factors and allow retailers to analyse the market competition. In 1963, Huff recognised that the likelihood of a consumer visiting a store is probabilistic and based on several factors. Therefore, the Huff spatial interaction model was created, a slight advancement of the SIM as Huff identified the need to incorporate a variety of factors that store attractiveness relies on. The SIM method provided retailers with a model that could be used to make location-based decisions. However, due to the lack of available data and computational power at the time, retailers did not begin to entertain the use of SIM in location-based analysis until the 1980s after large developments and progress were made by Wilson (1970; 1974).

The collaborative work between retailers and academics has also resulted not only in the method of SIM but also in the development of in-house site researchers for various supermarkets and the emergence of businesses that utilise spatial modelling techniques (Wood and Browne, 2007; Rogers, 2016). A few examples of these businesses include GMAP, a planning consultancy using spatial interaction models for retail modelling and forecasting. CACI, a location planning consultancy combining customer data with geographic data for location analysis, and Geolytix, another location planning consultancy, provide comprehensive datasets for retail planning and are expanding their top-down modelling tools to incorporate bottom-up methods. Retailers and consultancies acknowledge the importance of location planning, especially considering the vast changes in consumer behaviour over the last few decades (Chapter 2). The following subsections outline and describe grocery retailers' most commonly used methods to support store location decision-making. These techniques have been categorised into two analytical types “comparative methods” and “predictive methods”. Comparative methods include analogue techniques, checklists, and intuition; these methods rely on comparing potential site locations and using experience to make spatial decisions. Alternatively, regression analysis and SIMs are classified as predictive methods, a more advanced technique in which data is extracted, and statistics are used to predict an outcome. In this literature review section, the “knowledge-based methods” analytical category has been omitted due to the low usage amongst UK supermarket retailers (Reynolds and Wood, 2010). Although knowledge-based methods such as expert systems and neural networks were not commonly used a decade ago, there is a current rise of interest in using forms of artificial intelligence due to its substantial development and potential use of incorporating behavioural frameworks in spatial planning.

In addition to supporting location-based decisions, statistical models play a crucial role in conducting Retail Impact Assessments (RIA), as mandated by UK local authorities under the Planning Policy Statement 6 (PPS6) (ODPM, 2005). These assessments aim to gauge the repercussions of new retail developments on existing urban areas, encompassing various dimensions such as economic, social, and environmental impacts (Khawaldah et al., 2012).

Traditionally, two approaches have been used for RIAs: spatial interaction models (SIM) (section 3.2.2) and a “step-by-step” RIA approach. SIM, which gained prominence in the 1960s, leverages disaggregated data to forecast consumer demand for new retail developments, such as Haydock in 1964 (Foot, 1981). However, discrepancies among agencies’ results and criticisms of statistical-based models led to the Department of the Environment putting caution against SIM usage in impact studies since the 1980s (Khawaldah et al., 2012). Nevertheless, studies have highlighted SIM’s significance in accurately allocating expenditure-based geographies and populations, making them indispensable tools for major UK supermarkets such as Tesco and Sainsbury’s (Birkin et al., 2010). Understanding geographical demand is essential for a RIA and produces more coherent results than a “step-by-step” approach as outlined in (England, 2000). Despite SIM’s efficacy, its sensitivity to specific components, such as attractiveness, underscores the importance of meticulous calibration. While SIM has been a more accurate approach for the RIA process, there is growing recognition that individual-based methodologies might offer more nuanced insights (Wilkinson, 2023). These methods account for diverse consumer behaviours, including brand preference, demographics, and shopping habits, providing a more granular understanding of demand variations within geographical areas. Consequently, the modelling framework proposed in this thesis presents a valuable tool for enhancing the accuracy of Retail Impact Assessments. By capturing consumer behavioural changes in response to new retail developments, the model offers insights into consumer behaviour’s spatial and temporal dynamics, thus aiding decision-makers in anticipating the impacts of new developments on local areas.

As elaborated later in this thesis, the model’s ability to simulate diverse consumer behaviours in different locations and times enhances its utility for conducting robust RIAs, empowering stakeholders with a more comprehensive understanding of the potential ramifications of retail developments and local communities.

## 3.2 Modelling methods in grocery location analytics

### 3.2.1 Analogues, checklists and intuition

Retail location planning in the 1970s largely relied on the combination of intuition, analogue techniques, and checklists when deciding where to open, close or reformat their stores (Simkin, 1990; Clarke, 1998). The most simplistic method commonly used by small companies is intuition, where the director embarks on a site visit and makes an on-site decision based on ‘gut feeling’ (Davies and Rogers, 1984). The director will take into consideration the competition surrounding the location and get a ‘feel’ for the area and how well a store could perform there. Whilst the director may make a sound judgement call (ensuring the location meets retail requirements such as size, catchment area, and idea of competition) of the location based on experience, this method is subjective, time-consuming, and expensive (Clarke, 1998). Today, site visits are still an essential part of location planning even in an area of more sophisticated techniques, notably to supplement model-based techniques (Wood and Tasker, 2008); however, it is advised that decisions should not be solely based on gut instinct as there is a lack of validity.

Along with intuition, checklist approaches were commonly used before the 1980s. A list of location characteristics that are considered to have a positive influence on store performance is used to make store location decisions. These characteristics include the potential store floor space, local neighbourhood population size, demographics, and expected consumer catchment area. Potential store locations are measured against the checklist and ranked on their suitability as a site (Simkin, 1990). Since the development of GIS, the checklist approach has become more accessible due to the ability to geocode stores and household locations, calculate drivetimes and identify potential consumer catchment areas using buffers. Using a GIS, users can overlay numerous datasets and create a variety of buffers to identify the most suitable store locations based on a set of user-defined criteria. However, buffer analysis in GIS is subject to drawbacks, as competitors are assumed to have equal market share, thus misrepresenting reality (Beaumont, 1991; Benoit and Clarke, 1997; Clarke, 1998). Other analytical tools have been embedded within many GIS programs that go beyond the scope of buffer analysis, such as incorporating spatial interaction models, discussed later in this literature review.

The final comparative method commonly used today by location planners is analogue modelling. In this method, store sales forecasts are calculated by drawing analogies with other similar existing stores that are physically alike in terms of location, floor space, demographic area, and trade area circumstances (Clarke and Hayes, 2013). If a store performs successfully in an area, it is assumed that a similar size and location elsewhere would also be successful. Analogue techniques are popular today due to their low cost and low requirements regarding data and staff knowledge and have been used successfully in the past (Hernández and Bennison, 2000). One of the central critiques that these techniques carry is the assumption that stores in similar areas will perform the same; this can be a considerable risk for some retailers if the analogous store is currently over or underperforming, and it is unknown whether the new store would perform the same (Clarke, 1998).

Additionally, other aspects may be overlooked that contribute to store performance, such as accessibility via transport networks, changing demographics, or the economic downturn of an area (Birkin et al., 1999). Today, consumers are exhibiting complex behaviours and have shown to be alluding to convenience shopping; thus, the transport infrastructure plays a more prominent role (for example, the closer a store is to a train station, the more likely a customer will partake in a multi-purpose trip) (section 2.2). These factors may be overlooked in an analogue application, as commuters who visit a store may not reside in that catchment area, only pass through. Whereas in the analogous area, the retailer may attract more residential customers from the area than public transport users. Therefore, more sophisticated predictive methods should also be consulted due to their statistical nature and objectiveness. Despite these drawbacks, analogue methods are still commonly used by location consultancies today, providing a client with an affordable model that can be used to loosely indicate potential new store sites.

### 3.2.2 Predictive Methods: Regression Analysis and Spatial Interaction Models

Unlike comparative methods, predictive methods such as regression analysis (*also known as Scorecards*) entail using “cumulative data on past store performances to ascertain future ones” (Hernandez et al., 1998, p.305). Reynolds and Wood (2010) found that most of their grocery retailer participants relied on the use of multiple regression; this involves analysing the correlation between a dependent variable (such as turnover) and a set of independent/exploratory variables such as store size, the distance between customer home addresses and store location, and population demographics. This method has successfully explained the variation of store performance for some studies (Morphet, 1991) and can be a viable tool for location planning. As with all location-based decision-making techniques, regression analysis must be used with some caution as they provide a snapshot of store performance (Ciari et al., 2008), and there is an assumption that the exploratory variables are uncorrelated (Clarke, 1998). However, regression analysis and similar techniques have been proven to be hugely important for convenience store location planning, therefore, are still a valuable method to be implemented in retail location analysis today (Wood and Browne, 2007).

Progressing into the 1980s, as the competition for store sites and consumer patronage heightened during the store wars (section 2.1.1.4), retailers began to adopt more sophisticated modelling methodologies for their location-based decisions. SIMs, *also known as gravity models*, were first only used by a few retailers due to the cost requirements for access to data and software to run these models (Clarke and Birkin, 2018); some early models used survey data but were difficult to calibrate and validate. As data became increasingly available and computational technologies advanced in the early 1990s, SIMs produced significant value for early adopters of the method in their location-based decisions (Clarke and Birkin, 2018). SIMs assume that spatial issues play an integral role in the attractiveness of a store (Ciari et al., 2008). Benoit and Clarke (1997) explain that a SIM presents the flow of people or money from a residential area to a store, and is based on the demand in the area, the attractiveness of a store (such as floorspace, car parking availability, price), the distance to the store, and impact of competition.

Significant developments in SIM have occurred over the past few decades, particularly by those in academia. Researchers acknowledged that as consumer behaviours changes, the models used for location planning must also change. A key factor in store-based SIMs is the demand layers; for each store, different types of grocery demand are available at different time points. As opposed to only accounting for residential-based demand in the models, researchers have enhanced models to incorporate other demand layers such as tourism-based demand (Newing et al., 2013; Newing et al., 2014), daytime population demand (Hood et al., 2016; Waddington et al., 2018), and e-commerce demand (Beckers et al., 2022). Each of these developments and incorporations of new demand types has enhanced SIMs, typically using anonymised loyalty card data. With these model developments, predicted store revenues have been shown to be more accurate when accounting for the varying temporal demand.

SIMs are a well-known modelling method in location analytics and have been heavily used over the past decade. However, in today's retail environment, recent studies such as Birkin and Heppenstall (2011), Sturley et al. (2018), Rowe et al. (2022) and Wilkinson (2023) have highlighted that past studies of expanded SIMs are not able to produce consistent results today, and are no longer optimal (at least on their own) for modelling grocery store revenue. Their research indicates that newer methods are required for spatial models to continue their success in retail applications today. The main reason why SIMs are not optimal today is that they perform using a “top-down” methodology, in which data is disaggregated. By nature, SIMs cannot account for heterogenous customer behaviours and are missing an integral part of what makes a store perform well: the customer. As today’s consumer behaviour is becoming increasingly convenience-based, with more customers performing multi-purpose shopping transactions and turning to online shopping channels (as discussed in Chapter 2), SIMS must be further expanded, or newer methods are required. Location professionals have noticed the limitations in their current spatial modelling tools, especially at the author’s place of work. They are therefore seeking alternative modelling methodologies to use alongside their current ones. The following section discusses individual-based modelling methods as an alternative tool to support the development of spatial models in grocery retail.

### 3.2.3 Individual-based models: Microsimulation and Agent-based modelling

An alternative modelling method that can be utilised in grocery spatial modelling is individual-based modelling (IBM), which uses an alternative methodological approach to SIM. When modelling complex systems, two primary approaches emerge: “bottom-up” and “top-down” (Figure 3.1) (Lovelace and Dumont, 2016). Top-down models, like SIM, start with a high-level perspective and use aggregated data to identify macro-level patterns and trends. Conversely, bottom-up models use individual-level data to capture micro-level intricacies and interactions. Each approach offers unique benefits and drawbacks.

Due to their reliance on aggregated data, top-down models require less computational power and offer transparency in tracking each step of the model's logic. However, top-down models struggle to accurately capture nuanced behaviours, assuming homogeneity within modelled entities (Timmermans, 2001). In contrast, bottom-up models address these challenges by incorporating individual-level data, accounting for heterogeneity amongst modelled individuals. Furthermore, bottom-up models can accommodate complexity, yielding deeper insights by revealing nuanced patterns and relationships that may not be apparent at a more aggregate level (Bruch and Atwell, 2015). Despite their advantages, bottom-up models necessitate large, high-quality, granular datasets and pose challenges in validation due to their heightened complexity (Heppenstall et al., 2021). While top-down models excel in efficiency and transparency, bottom-up models offer a more nuanced understanding of complex systems, albeit with greater data and validation demands.

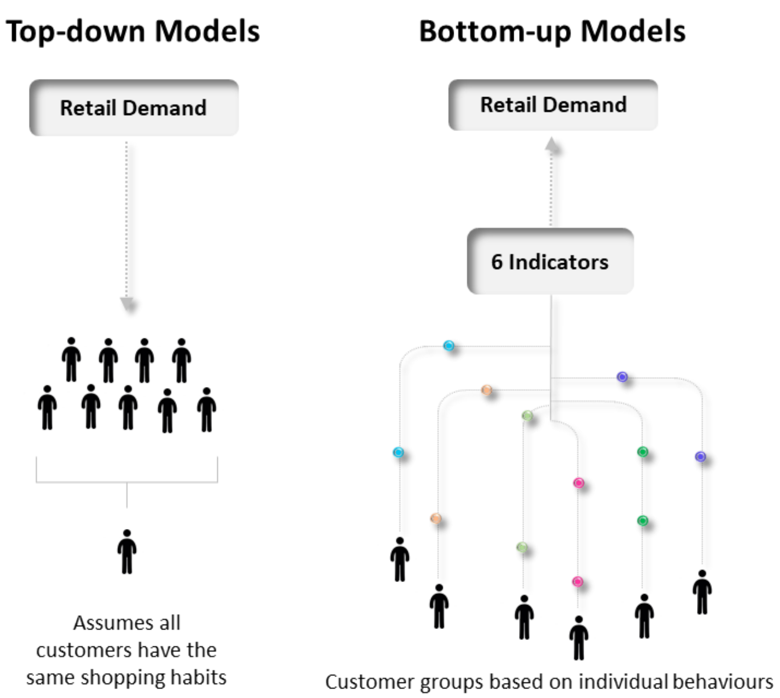



**Figure 3.1** Example structures of top-down and bottom-up models

Two prominent bottom-up methodologies that utilise an individual-based approach and have significant potential for application in spatial grocery models are microsimulation (MSM) and agent-based modelling (ABM). IBM serves as a computational technique to examine complex systems by simulating individual entities, often referred to as agents, each possessing distinct attributes and behaviours, capable of interacting with one another (Parrott, 2008). IBM finds utility across diverse disciplines, ranging from ecological and biological systems (DeAngelis and Grimm, 2014) to finance (Axtell and Farmer, 2022) and population forecasting (Puga-Gonzalez et al., 2022). The agents modelled within an IBM can represent anything, such as customers, biological cells, organisms, or abstract entities.

In an IBM simulation, an abstract environment is populated with these individual entities to incorporate spatiality and discrete-time steps are used to model temporality. During each time step, agents execute their assigned behaviours, leading to the emergence of patterns and phenomena at the macroscopic level. The critical advantage of IBM methodology lies in its capacity to model the genuine individuality of agents, thereby accommodating heterogeneity and stochasticity – a capability not readily achievable in top-down approaches (Heppenstall et al., 2016). Behaviour incorporation within an IBM can be achieved through various strategies. For instance, MSMs typically utilise rates, probabilities, and rules to simulate events and transitions. ABM employs rules, algorithms, interactions, stochastic processes and learning mechanisms to model behaviour. Each approach presents distinct benefits and is suited to study requirements and contexts.

MSM is a classic modelling method nested within the individual-based modelling framework (Willem et al., 2017). At its core, MSM applies discrete rules to a list of elements to determine an outcome (Birkin, 2021). For instance, MSMs are suitably applied to model population dynamics, characterising population behaviours through transition probabilities. These models leverage census data delineating individual-level attributes (e.g. demographics) and incorporate statistical metrics (such as mortality, fertility and migration rates) to forecast population trends amalgamating behavioural outputs (Bae et al., 2016). Beyond population dynamics, MSMs have found applications in studies on energy consumption, econometric analyses and policy evaluations.

The probabilistic nature of MSMs facilitates empirical-based investigations, rendering them highly representative of real-world scenarios. However, while MSMs excel in rate-based modelling at the individual level, they may need specific features offered by other IBMs. These include spatial interactions among individuals and other entities within the model environment, genuine autonomy, and spatial awareness (Ballas et al., 2018). Hence, depending on the model's specific requirements, alternative modelling approaches such as ABM may offer a more suitable solution. Unlike MSMs, ABMs necessitate a virtual environment populated with individual agents, which are then simulated in an artificial environment using predefined behavioural rules. ABM is a modelling approach that simulates systems by representing entities as a group of autonomous entities, commonly referred to as 'agents' (Bonabeau, 2002). These models are widely applied to social, economic and political sciences due to their minimal interference with society, their ability to test theories and model output based on 'actual' observed data, and the ability to simulate individuals more akin to reality compared to other approaches (Castle and Crooks, 2006).

ABM provides the opportunity to model phenomena from the bottom-up; All ABMs are made up of three key components; a set of agents that own attributes, characteristics and preferred behaviours, a set of agent relationships and interaction structures, and finally, an environment that the agents reside in, adapt to and interact with (North et al., 2010).

ABMs have been successfully applied in various fields of study such as behavioural sciences, (Smith and Conrey, 2007), economics (Axtell and Farmer, 2022), and pandemic modelling (Kerr et al., 2021). Few models have been applied in a retail context: Schenk et al. (2007)'s ABM of grocery behaviour using only population and survey data, Birkin and Heppenstall (2011) proposed a hybrid SIM and ABM for forecourt petrol retailing, Sturley et al. (2018) designed a proof-of-concept ABM model based on the possibility of using loyalty card data, and Bell and Mgbemena (2018) designed an ABM for mobile device sales using a decision-tree structure embedded within the model's base. These different models all utilise the ABM methodology, simulating the behaviours of individual agents.

Unlike top-down models, ABMs simulate the interactions of individual agents between one another and their environments, allowing for micro-behaviours to create aggregate output. Through modelling the individual, ABMs combat the issue of SIMs in which customers are assumed to behave the same. Agents in these models are autonomous and make their own decisions based on the information available in their environment and the rules assigned to them (North et al., 2010). Due to the agents having bounded rationality, they are inherently heterogeneous in their behaviour, allowing for the modelling of observable stochastic behaviour (Crooks and Heppenstall, 2012). In the application of grocery retail, individual customers can be modelled based on known and observable data, such as that found within loyalty card-linked transaction data. As customers become more individualised in their consumer preferences, mobility, and personal budgets (as discussed in Chapter 2), ABM and IBM methods could be an alternative spatial modelling technique for store location planning. Sturley et al. (2018) is the most recent study that attempts to design a proof-of-concept model of consumer store and channel choice and has been thoroughly critiqued later in section 6.1.3. Their work identified that ABMs could be used by location professionals to undertake site location assessments and Retail Impact Assessments through scenario testing. However, due to the software chosen, their work had a fundamental flaw, limiting their ability to accurately simulate consumer temporality.

Designing an ABM poses significant challenges, particularly as models can become overly complex. In response, the field of computation modelling advocates for the principle of “Keep it Simple, Stupid” (KISS) (Edmonds et al., 2019). This approach advises starting with a simple model and gradually introducing complexity when only necessary. By Adhering to this principle, modellers can maintain maximum control and comprehension of the model’s development.

ABMs are notorious for being data-intensive and demanding substantial computational resources and execution time (van der Ploeg et al., 2014; Heppenstall et al., 2016). Therefore, utilising as much information from a single dataset as possible is recommended before considering additional complexities. Additionally, analysing ABM outputs can be challenging, particularly when limited data is available for validation (Lee et al., 2015), often leading to the formulation of inferences rather than definitive conclusions. While both ABM and MSM share similarities as individual-based approaches, they possess distinct characteristics and individual descriptions. ABM emphasises modelling interactions between individuals, thereby incorporating spatial elements into the model. In contrast, MSM facilitates the integration of behavioural data based on discrete time intervals (Bae et al., 2016).

Recognising these differences, a hybrid approach combining MSM and ABM methodologies can be advantageous. Such a hybrid approach leverages MSM to depict individual stochastic behaviours based on actual real data, while ABM accounts for individual interactions and environmental dynamics (Richiardi, 2014). This integrated approach offers a comprehensive framework for modelling complex systems, incorporating individual-level behaviours and interactions within a spatial context.

In grocery retail, ABMs hold promise as spatial modelling tools due to their ability to simulate individual consumer behaviours (Sturley et al., 2018). Leveraging the abundance of consumer data collected through loyalty card programmes, ABMs can closely resemble the observed behaviours of known customers. By employing decision-tree structures within the modelling framework (Lipowski and Lipowska, 2012; Kotsiantis, 2013; Bell and Mgbemena, 2018), an ABM can capture the temporal aspects of consumer behaviour, incorporating bounded stochasticity while accounting for both independent and similar behaviours among individuals (Cooley and Solano, 2023). The decision-tree structure is used within the modelling framework presented in this thesis and is further discussed in section 6.1.4.

The modelling framework presented in this thesis adopts a hybrid MSM-ABM approach, integrating decision-tree structures to simulate individual customers' store and channel choices, considering spatial, temporal and basket-type variations. This framework captures observed customers' stochastic and heterogeneous behaviours, addressing limitations in traditional SIMs. This thesis meticulously documents each model development step, providing insight into the challenges encountered and the potential applications of loyalty card data within the context of grocery retail analytics. The resulting IBM offers a reproducible framework for scenario testing and enhancing our understanding of consumer behaviour in grocery retail environments, which is ideal for analysing retail impacts.

Subsequent sections explore the potential data sources that could further enhance spatial models in grocery retail and their potential implementation within the IBM framework.

## 3.3 Data in spatial grocery analytics

Along with changing consumer behaviours, technology has also evolved, leading to the 'Big Data' field – where large sets of data are collected and analysed using specialised methods, as the data are too complex to be examined using traditional data-processing software. Today, personal data is constantly being generated and stored from the use of personal mobile devices, purchases using debit or credit cards, online activity, loyalty card use, transport cards, social media activity, and even from your television-watching habits (Birkin, 2018). These datasets contain data beneficial to retailers and location planners, spatial data, otherwise referred to as 'Spatial Big Data' (SBD). Like Big Data, SBD data is characterised by the velocity of data collection, their large volume, and the wide variety of sources. SBD contain various information on humans, notably information about their spatial mobility and consumption behaviour (Lee and Kang, 2015).

### 3.3.1 Loyalty card data

Loyalty card data is a notable dataset collected by some grocery retailers that is incredibly insightful in retail location decision-making. Loyalty card schemes are commonly used by British supermarkets, providing customers with accounts in which points are accumulated on each recorded transaction made at that retailer in exchange for rewards. For example, there are Sainsbury's Nectar card, Tesco

Clubcard, Morrisons More, Sparks by M&S, and ASDA Rewards among others. Tesco was the first grocer to launch a loyalty programme for UK customers in 1995 (Humby et al., 2004), shortly followed by Sainsbury's and their launch of the Nectar card scheme in 2002 (Hassan and Parvez, 2013). Today, Tesco has ~21 million Clubcard users (Statistica.com, 2023) compared to ~18 million Sainsbury's Nectar card users (Sainsbury's.co.uk, 2023b). As Tesco and Sainsbury's are direct competitors, the lower number of Nectar card users could be an incentive for why Sainsbury's seeks to better understand their consumer base, improve their services to current customers, and attract new ones.

Based on consumer shopping frequency and items purchased, rewards are tailored and personalised to the individual as an incentive to shop at that supermarket chain (Bombaij and Dekimpe, 2020). Whilst beneficial for the consumer, loyalty programs offer retailers a wealth of data regarding their customers' shopping behaviours (Jackson et al., 2006). Each transaction recorded with a loyalty card provides grocers with a summary of each unique cardholder, collecting information on which stores they visit, when they visit, the items purchased, what shopping channel was used, and how much money was spent (Byrom et al., 2001). These data provide location planners with the ability to better inform location-based decisions (Waddington et al., 2018; Hood et al., 2021), especially concerning customer store and channel choice. These data have been strategically used in the past to identify 'gaps' in a retailer's store estate, provide insight into customer lifestyles, and can be used to support store location decision-making through the estimation of impact or sales cannibalisation of new stores (Wood and Browne, 2007).

Recent studies investigate the richness of grocery retail data regarding customers' transactional behaviours to further enhance location planning models (Sturley et al., 2018; Waddington et al., 2018; Rains, 2019). Each study explores an integral area of customer behaviours using loyalty card transaction data, but each has limitations. Waddington et al. (2018) analysed grocery transaction data to cluster stores into types based on in-store customer transactions instead of clustering customers into types based on their observed behaviours. Sturley et al. (2018) clustered grocery consumers to support the development of a simple agent-based model but used synthetic data. Rains (2019) provides a methodology for clustering grocery customers into types based on behaviours; however, due to using a dataset covering the entirety of the UK, any nuances in consumer behaviour based on regional differences may be obscured.

Loyalty card data is a valuable tool that provides insight into the spatial and temporal behaviours of known customers of a grocery retailer. However, not everyone who owns a loyalty card is a “loyal” customer, as they may shop at various brands for different reasons, as discussed in Chapter 2. Furthermore, loyalty card data only captures the behaviours of those who are part of the loyalty scheme, and those with a card may not always scan it for each transaction (Lloyd and Cheshire, 2019; Rains and Longley, 2021). Additionally, whilst these data provide information on the customer’s place of residence and store choice locations, no other information is provided regarding their whereabouts (Lloyd and Cheshire, 2019). Therefore, there is a void in the knowledge of the relationships between transaction spatiality and temporality. Loyalty card data can provide context for residential-based grocery demand but is limited in its ability to accurately represent other types of demand.

Whilst the loyalty card data provides invaluable insight into consumer behaviours, it only partially represents all customers. Alternative data sources, such as the Living Costs and Food Survey (LCFS), provide insight into consumers' grocery purchasing behaviour by directly asking customers about their expenses. However, the LCFS and other surveys are limited in scope as they typically have moderate sample sizes of ~5,000 to 6,000 households in England and Wales and only cover a two-week survey period (Bulman et al., 2017). In contrast, the loyalty card scheme for Sainsbury's contains data for up to 18 million customers, thus providing a vast wealth of insight about those customers and their transactional behaviours over an unlimited time period.

All available data sources have their benefits and drawbacks; therefore, a complimentary approach may be the most appropriate when researching consumer behaviour. However, the applicability of survey data in building an IBM may be insufficient, whereas loyalty card data presents a more comprehensive source of data. This thesis mainly utilises loyalty card data, exploring the extent in which it can be used to create an IBM whilst applying other data sources where necessary, for example, workplace data.

### 3.3.2 Capturing customer movement

Other novel data sources could be used to approximate the movements of Sainsbury's customers to help estimate non-residential-based grocery demand. The following sections discuss what data sources exist and could be used to further expand upon the modelling framework developed in this thesis.

### 3.3.3 Workplace population

As the loyalty card data does contain information regarding consumers' whereabouts outside of the home, supplementary data could be used to estimate where customers are located at other times of the day, e.g., the census. Since 1801, the UK government have collected a nationwide census every ten years in which all households and communal establishments in England and Wales receive a questionnaire through the post and may complete the survey online (ONS, 2012). The census gathers household data about the inhabitants, such as dependents, employment status, place of work, work hours, and educational backgrounds. The Office for National Statistics (ONS) provide access to the census data that has been anonymised to retain individual privacy and include further analytical reports. The UK census is a rich source of data providing information regarding the demographics of households.

As the UK census is collected at the household level, individuals are enumerated at their ‘place of usual residence’. Within the data, non-standard populations are also captured, such as students away from home, and persons with second residences. Using these questions regarding residential addresses, employment status, places of work, and hours worked, the ONS presents alternative population bases for England and Wales. The alternative population bases are designed to estimate the population density of an area and suggest how this may change between the workday and at night (ONS, 2012). For this thesis, the most relevant population bases from the census are the residential and workplace populations. These population bases are helpful when attempting to estimate where people are located during the day, such as work or home. For each household in the UK, the census provides a place of residence and a place of work at the output area level (OA). OAs are produced by the Office for National Statistics (ONS) and designed for analysing residential and household data. They are broadly consistent in population size, with each OA containing 40 to 250 households or 100 to 625 individuals (ONS, 2012). Therefore, the movement of people can be inferred, which is helpful for building an IBM to capture non-residential grocery demand. However, it is difficult to determine which individuals in the census are likely Sainsbury’s customers. The workplace data is helpful for modelling movements between OAs, but it is impossible to know how representative it is of the Sainsbury's customers. Nevertheless, both sources of information (census and loyalty card data) are invaluable for creating an IBM of consumer store and channel choice behaviours.

#### 3.3.3.1 Transport data

A novel data source that could be used to estimate the movements of a population includes transport datasets, particularly data collected by public transport. Each time a person alights on a public transport service such as a bus, train, tram, or subway using a smart card or pass, their journey is recorded from start to end to estimate the fare. These data are often used to analyse the spatiotemporal patterns of urban human mobility (Hasan et al., 2013), the estimation of fares (Seaborn et al., 2009), and the distribution of a population (Ma et al., 2017). Various transport services across the UK collect these data using smart cards, including Oyster card for various transport types in Greater London, Swift cards for train and bus travel in the West Midlands, and MCards for transport in West Yorkshire. Over recent years, there has been an increase in the number of areas around the UK providing these smart cards. The increase in the number of journeys being captured provides researchers and retailers insight into where people are travelling, how far, and how often. These characteristics cannot be captured in the census as journeys are not always consistent and vary day to day. The census does capture the mode of transport to and from work; however, it is not constantly updated, and individuals may move homes or places of work.

Public transport that uses smart cards is an incredibly useful source of data. Each journey is captured with a timestamp from which the journey started, and in some cases, the data presents the destination of that particular journey. Actual routes can also be identified, knowing where people move through space and time. These data can provide retailers with an insight into when people are on the move and whether they are potential consumers of their stores based on distance.

#### 3.3.3.2 Footfall Cameras

Footfall cameras are one piece of technology that can capture spatiotemporal fluctuations within a geographical area. These cameras count the number of people passing a specific geographical point, estimating how many people are within the area at a given time (Crols and Malleson, 2019). Private companies often operate these cameras to quantify how many people passing the area are potential consumers. One drawback of this method of capturing footfall counts is that the same person may enter and leave an area, thus being double counted. Therefore, footfall cameras can use other methods, such as specific target tracking. In this method, small counting devices are mounted to columns around an area, creating a virtual zone in which pedestrians who pass through are counted (spring-board.info, 2021). This technique combats the issue of double counting pedestrians; however, it is best suited for the analysis of specific areas rather than a broad coverage.

Footfall camera data have yet to be presented in many academic studies, likely due to the data's private ownership and the associated ethical concerns. Retailers, however, have the means to purchase footfall cameras and use these technologies to estimate how many people are entering their store and predict how many customers are visiting that day and perhaps not making purchases (Lansley and Longley, 2017). Crols and Malleson (2019) use footfall data for their study on quantifying the ambient population of Otley, West Yorkshire. In this research, an agent-based model was used to disaggregate mean footfall counts to estimate the size and demographic makeup of the ambient populations during the day. A similar application could be applied in a retail context; however, it may provide limited spatial coverage of the population fluctuations and would focus solely on the store location (Hildebrand, 2012). Footfall data are instrumental in providing insight into population fluctuations. However, there are concerns around geo-privacy; thus, data access is restricted. Therefore, if an IBM were to incorporate data from footfall cameras, permissions from private companies would be required, or Sainsbury's could employ their own set of footfall cameras for spatiotemporal retail demand research. Ultimately however, linking those customers to a place of residence would be almost impossible solely using these data.

#### 3.3.3.3 Geo-tagged Social Media Posts

Over recent years, there has been an increase in the variety of social media platforms in which users can share their location, such as Twitter, Facebook, Instagram, Snapchat and FourSquare. The geotagged posts provide the exact coordinates of where a user posted online if they turned on the feature. These data can be downloaded via an application programming interface (API) for free by anyone who has requested an Access Token for secure access from the social media provider. Unfortunately, only around 1% of all social media posts are geotagged and thus may not be a large enough dataset to analyse the spatiotemporal fluctuations of an area (Jurgens et al., 2015). Additionally, access to social media and posting at all times is not equal amongst all populations, not everyone has access to data away from wi-fi. Despite this, several studies have used geotagged social media posts to analyse spatiotemporal fluctuations. Tsou et al. (2018) used geotagged tweets, land use data, and dasymetric maps to estimate the hourly population distribution during 2015 in San Diego. The study found that using geotagged data to create a population model is incredibly difficult to validate and that Twitter data is not stable in terms of consistent production. Despite geotagged social media posts being openly accessible, the datasets are not substantial enough to provide robust estimates of population change in an area for 24 hours. Studies have shown, however, that there is a potential for emerging geotagged social media datasets to be used in a retail geography modelling context (Lovelace et al., 2016).

#### 3.3.3.4 Wi-Fi Sensors

Almost all mobile devices in 2019 use wireless technologies in which data is exchanged using radio signals via airwaves (Torrens, 2012). These devices connect to Access Points (AP), and every time a device connects to Wi-Fi, it connects to an AP. Today APs can be found anywhere, from homes and hospitals to universities, coffee shops and commercial properties. When located in an area with Wi-Fi accessibility and a device has Wi-Fi turned on, probe requests are emitted from the device and sent to the APs. These probe requests can be counted to estimate the footfall within a certain proximity of a Wi-Fi point. With so many devices connected to Wi-Fi, there is an abundance of data available regarding where people are during the day based on probe counts. Yang et al. (2018) studied how Wi-Fi devices can be used to identify where people are moving and in which direction. Companies such as The Local Data Company (localdatacompany.com) use Wi-Fi sensors to estimate footfall for retailers, highlighting how many pedestrians are nearby during the day, making them potential consumers. Whilst this data source is incredibly useful in counting the number of people within an area at a specific time, it does raise concerns regarding user privacy. Therefore, Wi-Fi sensor data is not publicly available or easily accessible for all studies.

## 3.4 Chapter 3 summary

Several location-based theories for grocery retail were presented in this chapter: highlighting the incremental developments and increased applicability to the changes in consumer grocery demand. Coinciding with these theories, various location-based methodologies have been developed, each becoming more sophisticated. Due to improvements in computational power and academic research, tools used by retailers have evolved from checklists and intuitive-based approaches to predictive and data-focused approaches. These tools are primarily used to support site location assessments and Retail Impact Assessments. Models such as SIMs have become industry standard due to their robustness and applicability to the retail landscape, especially over the past decade. However, as consumer behaviours have become more complex and individualised due to the rise in convenience-based demand, e-commerce, and changing working environments, SIMs may no longer be the most appropriate methodology for location-based decision-making. Due to their aggregative approach and inability to consider heterogeneous behaviours of customers, other IBM methods, such as ABM, may be more applicable.

Whilst IBMs provide the ability to simulate individual consumer behaviours, unique challenges are presented:

1. These models require abundant individual-level data to represent observed consumer behaviours. Such data is often inaccessible due to the proprietary nature and concerns regarding customer privacy.

2. Analysing such sizable datasets is time-consuming, requires computational efficiency, and must be appropriate for mining behavioural rules for simulated agents.

3. The anonymised nature of accessible individual-based data does not provide information regarding the movements of these individuals as it cannot be linked to other spatial-based data sources.

Nevertheless, supermarket loyalty card-linked transaction data provides invaluable insight into known customers' store and channel choices. Whilst these data do not provide contexts for *all* customers, they do provide a comprehensive representation of *many* customers who have preferences in their grocery shopping habits. These data are suitable for developing behavioural rules for agents within ABMs, resulting in an entirely data-driven model. To combat some of the challenges of individual-based data, supplementary data sources can be considered. For example, mobility data may support the modelling of customer transactional behaviours outside of their homes, capturing non-residential-based grocery demand. As the literature highlights, these data are often difficult to obtain or link to the same individuals in the transaction dataset. The most accessible data currently available is the ONS' census data, providing the home locations and places of work for British residents. Knowing which of these residents are the same customers captured in the loyalty card datasets is impossible. However, it provides an opportunity to develop such IBMs in the context of grocery retail.

Developing IBMs of consumer store and channel choice behaviours can provide grocery retailers and researchers with a valuable bottom-up modelling tool currently missing in their arsenal of spatial models. These IBMs, once fully validating against other datasets, can be used for scenario-testing various changes in the grocery markets, such as changes in consumer behaviours and their impact at the store-level or the changes by retailers and the impact on consumer behaviours. Ultimately, these IBMs can support the location planning decisions being considered by grocers and used as a tool for the Retail Impact Assessment process.

# Chapter 4 Preliminary analyses

Chapter 1 presented this study's key aims and objectives, introducing the key concepts and study purpose. Chapter 2 and Chapter 3 achieved the first aim of exploring and reviewing existing literature around the evolution of grocery consumer behaviours, the tools used by those in grocery location analytics, and the datasets and more recent methodologies available to further expand and refine those methods. This chapter supports the achievement of the second aim regarding the identification of key consumer segments within the loyalty card dataset. To achieve the second aim, objective 3 was attained:

*To investigate the loyalty card linked transaction dataset provided by Sainsbury's, identifying general behaviours of their customers in relation to their store choice behaviours in West Yorkshire.*

This chapter overviews the case study area and investigates the critical dataset provided by the study collaborator. This study is the first documented piece of research that investigates a loyalty card-linked dataset of this calibre in such a way for individual-level model building. Therefore, a thorough understanding of what can be derived from such data, and the drawbacks of using these data, is vital for expanding the field of knowledge. The chapter is structured as follows to investigate the spatial, temporal, and behavioural aspects of customers store choice behaviours that can be extracted from these data.

Section 4.1 offers insight into the significant features of the study area of West Yorkshire, such as population, demographics, and grocery expenditure, and how these factors may relate with consumer types. Section 4.2 explores the commercial sector loyalty card and retail supply side data provided by the study collaborator, Sainsbury's, covering variables and store information. This section includes a summary of the types of transactions that customers had made over the 12-week period, where they had taken place and the differences between day type (weekday versus weekend) and time of the day. The data provided by Sainsbury's covers an entire county, at the finest resolution of the individual consumer level, including details regarding each transaction undertaken by uniquely identifiable consumers. The transaction dataset covers the period between May and July 2018, capturing all transactions at a Sainsbury's store within West Yorkshire by only those with a loyalty-card registered household address within West Yorkshire. Restricting the study to one region allowed for better model development due to reducing complexity and better model testing. The region chosen comprises of five local authority districts (LADs), in which the cross-district transactional behaviours of customers can be observed. A smaller case study area would provide limited insight into consumer behaviours across space, notably between workplace hubs and residential areas, which form a significant part of grocery demand (Waddington et al., 2018). A wider study area, however, would have provided excessively more data and created computational difficulties when working with individual-level transaction data. Therefore, the county of West Yorkshire provided a middle-ground between enough data to analyse, and enough for modelling software to manage. Sainsbury's agreed to the study area due to the ample number of past studies based on the area in retail location analysis, meaning that we have a good baseline understanding of the type of consumer behaviours and the structure of the retail supply side in this area (Newing et al., 2014; Kirby-Hawkins et al., 2018; Waddington et al., 2018; Sturley et al., 2018; Newing et al., 2018; Waddington et al., 2019), and their collaboration with these studies.

## 4.1 Study area

The metropolitan county of West Yorkshire was chosen as the study area as it covers a range of urban and rural geodemographic segments (Office for National Statistics, 2018; Office for National Statistics, 2020a), has been used in various studies of spatial dynamics in the UK grocery sector (Waddington et al., 2018; Sturley et al., 2018; Waddington et al., 2018; Hood et al., 2020), and has a comprehensive store and online groceries coverage by Sainsbury's and key competitors. With a population of around 2.2 to 2.35 million (Office for National Statistics, 2012; Office for National Statistics, 2021), West Yorkshire consists of five Local Authority Districts (LADs) Bradford, Calderdale, Kirklees, Leeds, and Wakefield. These LADs comprise 7,131 Output Areas (OAs) (the smallest geographical unit for the release of small-area population, housing, expenditures and commuting data) located in various urban and rural areas. OAs are produced by the Office for National Statistics (ONS) and designed for analysing residential and household data. They are broadly consistent in terms of population size with each OA containing between 40 to 250 households or 100 to 625 individuals (Office for National Statistics, 2012). The ONS also created their Output Area Classification (OAC) using the 2011 Census data, which classifies each OA in England and Wales (based on its dwelling type and population composition) to determine an overall insight into the demographics and composition of neighbourhoods (Gale et al., 2016, p.201). The OAC was produced using a top-down classification system and segments the population into 8 supergroups, 26 groups and 76 subgroups (Office for National Statistics, 2018). Figure 4.1 presents the 2011 OAC for West Yorkshire at the supergroup level, showing the spread of different demographic groups across the county at the OA-level (Office for National Statistics, 2018). The most common demographic group by OA in the study areas are suburbanites (22%), followed by urbanites (20%), hard-pressed living (20%) and multicultural metropolitans (19%). Previous studies analysing the geodemographic profiles of grocery retailers found that for Sainsbury's the majority of their customers reside in areas considered as 'typical traits', 'prospering suburbs', and 'blue collar communities' (Thompson et al., 2012). Therefore, it is expected that majority of customers in this study, using the loyalty card-linked dataset, reside in areas categorised as 'suburbanites' or 'urbanites', this is explored in section 5.4.4. The pen portraits for the OAC groups can be found in Office for National Statistics (2018), which further describes the characteristics of these populations. The OAC is a crucial input to this study as it allows for the identification of any relationships between customer behaviours and the type of areas they reside in later in the study.

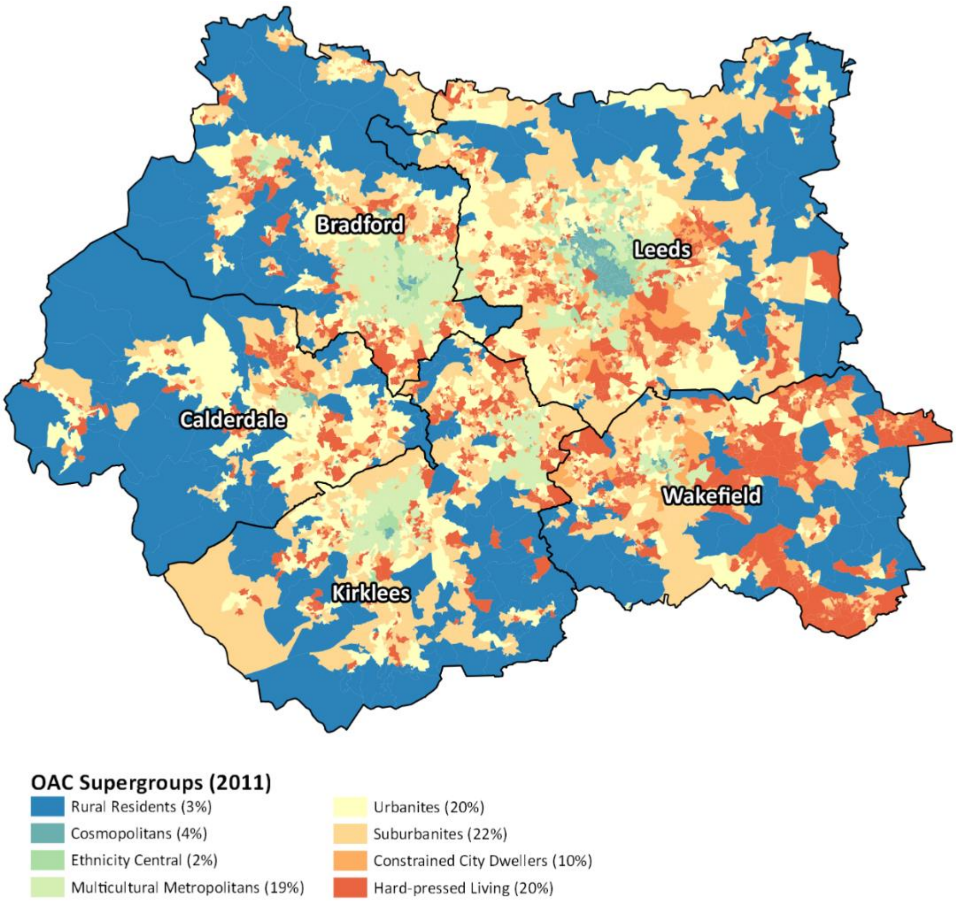



**Figure 4.1** OAC Super Groups across West Yorkshire districts (percentage of total OAC groups within West Yorkshire by OA), (Office for National Statistics, 2017).

To better understand the relationship between demographic groups and the amount of money spent on groceries per week, the Living Costs and Food Survey (LCFS) dataset was analysed. The LCFS is an annual survey administered by the ONS in which a sample size of 5,000 to 6,000 households in England and Wales are asked to record their daily expenditure over two weeks (Bulman et al., 2017). These expenses include household items, food and drink purchases, and larger purchases such as holidays and cars. Information on household composition, characteristics, and income are also collected, with results given that account for the characteristics relative to all households. The LCFS is one of the few data sources that provide an idea of a household's total expenditure ready for research (Rains and Longley, 2021). In this study, the 2017–18 survey period is used, containing data for 5,480 households (Office for National Statistics, 2020b) which ties in with the period from the Sainsbury's transaction data, linking the average likely behaviours by OAC group to those who transacted in the time period.

The LCFS is later used in this chapter to understand the share of expenditure by OAC group in relationship to the loyalty car dataset; allowing us to analyse whether customers likely perform most or all their transactions at a Sainsbury's store, thus suggesting whether they are 'loyal' customers. The relationships between store locations, customers locations and 'loyal' customers are integral for retailers to understand to support location-based decision-making, which can be implemented in their location planning tools and models. Future model expansions beyond this thesis can incorporate the aspect of consumer loyalty once competitor data has been obtained. As the loyalty card dataset used in this project only provides insights into the Sainsbury's stores that customers visited, other data are required to link customers to other supermarket brands. For those customers who are not 'loyal', we can infer what other retailers they are likely to transact at using competitor data or other data types such as banking data – though there are unique challenges with linking customers transactional behaviours across different retailers.

It is commonly accepted that retailers will target distinct types of consumers (Lansley and Cheshire, 2018), where ‘high-end’ brands and retailers often target more affluent consumers, and ‘budget’ and ‘affordable’ brands and retailers will target those less affluent (Shukla et al., 2013). In this study, Sainsbury’s is considered a mid-range supermarket due to the variety of their own brands that they offer from, ‘SO Organic’ and ‘Taste the Difference’, which tend to be more expensive products, to ‘Sainsbury’s basics’, which are their value products that are often the cheapest (Sainsbury’s.co.uk, 2022). Data from Which.co.uk’s supermarket price comparison tracker (2023) consistently indicate that Sainsbury’s is often the cheapest or second cheapest supermarket out of the UK’s ‘big four’. Therefore, it is expected in section 5.4 that customers within this study area are likely to reside across various area types (as defined by their geodemographic classification) as they target both affluent and less affluent customer types. However, because West Yorkshire is predominantly urban and suburban (Figure 4.1), it is also anticipated that most Sainsbury’s customers will reside those areas.

Assessing each LAD individually and referring to Office for National Statistics (2018), Table 4.1 highlights the proportions of each OAC supergroup using population counts. 40% of Bradford’s population falls within the OAC supergroup ‘multicultural metropolitans’, predominantly located between urban centres and suburbia, and have an unemployment rate above the national average. Bradford is one of the UKs largest multicultural cities, with 32.1% of inhabitants identifying as “Asian, Asian British or “Asian Welsh” (ONS, 2023). Research found that Asian British customers tend to do visit a supermarket for a large shop on a weekly basis, but struggle to find specific food items, therefore rely on visiting local international food shops instead (ABPL Group, 2016). In this study, it is expected that customers in the loyalty card dataset will make up a smaller proportion of those who belong to the multicultural metropolitan OAC group. Those who are part of that OAC group, may be more likely to shop at Sainsbury’s for their ‘main’ basket purchases only. In contrast, one-third of Calderdale’s population is classified as living in the ‘urbanites’ group, where unemployment is lower than the national average, and people are likely to live in more urban areas. Kirklees has a more varied population with 28% belonging to the ‘multicultural metropolitans’ group and 24% in ‘suburbanites’. The suburbanites are usually located on the outskirts of urban areas, are more affluent, and are most likely to own a car. It is expected that the majority of Sainsbury’s customers in this study belong to the suburbanite OAC group due to Sainsbury’s provision of both convenience and supermarket stores located on city edges, and within neighbourhoods. Similarly, 22% of Leeds’ population is categorised as ‘suburbanites’, and 21% as ‘urbanites’. Conversely, Wakefield has 40% of its population living within ‘hard-pressed living’ areas, where people usually live in urban areas where unemployment is higher than the national average. This initial understanding of the variation in demographics between LADs will be used to assess the customer segments identified in section 5.4.

**Table 4.1** Population breakdown by LAD and OAC supergroup in which the green values are the higher percentages of OAC groups for that LAD. Red represents the lower percentages of OAC groups for that LAD. Data source: (Office for National Statistics, 2015).

| OAC Supergroups | Bradford | Calderdale | Kirklees | Leeds | Wakefield |
|---|---|---|---|---|---|
| 1 - Rural residents | 2% | 6% | 5% | 2% | 5% |
| 2 - Cosmopolitans | 1% | 1% | 1% | 9% | 0% |
| 3 - Ethnicity central | 2% | 0% | 1% | 3% | 0% |
| 4 - Multicultural metropolitans | 40% | 10% | 28% | 17% | 5% |
| 5 - Urbanites | 19% | 33% | 21% | 21% | 12% |
| 6 - Suburbanites | 18% | 17% | 24% | 22% | 27% |
| 7 - Constrained city dwellers | 6% | 10% | 4% | 9% | 10% |
| 8 - Hard-pressed living | 12% | 22% | 17% | 16% | 41% |

Table 4.2 presents each demographic group's expected weekly grocery (food and non-alcoholic drinks) expenditure. As the LCFS data has been linked to the OAC data, the relationship between population, demographics and grocery expenditures can be assessed. On average, in 2017/18, the more affluent OAC groups such as suburbanites, urbanites and rural residents all spent the most each week on food and non-alcoholic drinks (between £61.70 and £70.30 per household per week), whereas those less affluent such as those living in neighbourhoods classed as cosmopolitans, ethnicity central and constrained city dwellers spent the least amount (between £41.40 and £54.80 per household per week). The LCFS data will be used to assess how many customers in each consumer type group identified in Chapter 5 spent the expected weekly amount of money on food and drink. This forms an important indication as to whether that customer is likely to purchase most of their groceries from a Sainsbury's store on a week-by-week basis, capturing a measure of customer loyalty and a notion of the share of their expenditure captured by our study retailer.

**Table 4.2** Expected weekly expenditure on food and non-alcoholic drinks by OAC supergroup. Green values represent hight amounts of expenditure per week. Red represents lower expected expenditure per week. Data Source: (Office for National Statistics, 2020b).

| OAC Supergroups | Weekly Expected Expenditure (LCFS) |
|---|---|
| 1 - Rural residents | £61.70 - £68.60 |
| 2 - Cosmopolitans | £41.40 - £54.80 |
| 3 - Ethnicity central | £47.00 - £59.80 |
| 4 - Multicultural metropolitans | £56.90 - £68.40 |
| 5 - Urbanites | £62.90 - £64.20 |
| 6 - Suburbanites | £64.00 - £70.30 |
| 7 - Constrained city dwellers | £43.30 - £46.80 |
| 8 - Hard-pressed living | £56.70 - £60.30 |

As discussed in section 2.2, these insights into the demographics of the study area LADs help for a better understanding of where customers with certain behaviours are likely to reside. For example, those who live in less affluent areas, such as constrained city dwellers, are less likely to travel far to undertake their 'main' basket grocery transactions due to lower private car ownership (Office for National Statistics, 2018). Whereas those residing in more rural or suburban areas are likely to own a car, travel further afar, at stores that target a more affluent audience or shop online (Kirby-Hawkins et al., 2018; Hood et al., 2020). Rural dwellers are often forced to travel further to undertake their food shops due to the lower density of food store provision in those rural areas.

Additionally, the LCFS data allows for a better estimate of the amount of money likely to be spent by customers, depending on their location. It is acknowledged that not everyone who resides within a particular area expresses the same behaviours as those of their demographic segment. However, it does offer an indication of what could be expected. The following section explores the transaction dataset provided by Sainsbury's, analysing the dataset from the individual level and store level.

## 4.2 Sainsbury's loyalty card-linked transaction dataset

Loyalty programs provided by grocery retailers have become a profound data source for expanding the tools used to support location-based decisions, as summarised in section 3.3.1. Sainsbury's have provided access to a comprehensive subset of their Nectar card transaction-level dataset to support the development of the model built in this thesis. This section introduces that data and explores, summarises, and explains the variable creation process to mine consumer behaviours from these data for subsequent individual-based model development. A 12-week analytical overview is conducted of the customer's behaviours regarding the transaction's temporality, basket type, and spatiality from both the customer's perspective and store level. Understanding transactional behaviour from both the demand and supply side is important to gain insight how both sides influence the other. For example, the location of Sainsbury's stores and their store type will impact the opportunities that customers have to interact with the stores, i.e., stores located in train stations are expected to receive custom from customers from a variety of locations due to the store location and will likely provide smaller basket type transactions. Larger supermarkets located closer to residential neighbourhoods or city edges are expected to be visited by customers for a variety of basket type transactions, at all times of the day and week. Within those transactions however, unique consumer type groups are expected to be proliferating specific types of transactions scenarios for specific store types. The following section provides an overview of the loyalty card-linked transactional dataset and analyses the spatiotemporal and channel choice behaviours of those customers.

### 4.2.1 Dataset overview and analysis

The loyalty card dataset contains 2.7 million transactions, by 216 thousand customers, in 55 stores, over 12 weeks. Sainsbury's 'Nectar' loyalty scheme collected these loyalty card-linked transactions and extracted them from the spring and summer months between May and July 2018, avoiding seasonal product purchases at Christmas and Easter, which could skew habitual behaviours. The dataset includes all in-store and online transactions at a Sainsbury's store (online orders were attributed to the store where the order was packed, typically the closest large format store to the delivery address) by customers whose home address linked to their loyalty card account is within West Yorkshire. Table 4.3 summarises the transactional dataset provided by Sainsbury's. Each customer has a unique ID, enabling us to group all transactions individual customers made whilst presenting their loyalty card. We know nothing about transactions made by those customers or other members of their household for which a loyalty card was not presented. Every transaction recorded also has a date and time stamp, essential to explore the temporality of shopping. This enables us to link individual customer transactions to a specific basket type. The basket type variable was created by Sainsbury's and classified each transaction based on the range of goods purchased and the total transaction value. It uses a confidential in-house methodology that combines transaction and customer survey data. This study categorised transactions into three key basket types: 'food for now', 'top-up', and 'main'. Transactions considered as 'food for now' basket types contain perishable foods that are ready to eat or are expected to be eaten within a few hours in the same day (meal deal items, ready meals, and drinks) and are low in sales value. 'Top-up' transactions contain a small number of items that are a mix of perishable and shelf-stable foods and are expected to last the customer a few days (milk, bread, fruit and vegetables, etc.). 'Main' basket types contain food that will last many days and over a week, these are large value transactions that contain a variety of food item types, including frozen items, meats, and the items found in ‘for now’ and ‘top-up’ baskets. The basket type variable infers the purpose of a transaction without invading customer privacy by limiting insight into the exact items purchased. This crucial variable also allows the linkage between basket type and time of day, allowing the assessment of the relationships between types of transactions with time and geography.

The home addresses of each customer were aggregated into OA geographies to preserve anonymity. These addresses are the ones the customer used to sign up to Nectar or are updated if the customers have a saved home delivery address when transacting online. Using the population-weighted centroid (PWC) of each customer’s OA with coordinates of the store they visited, a distance variable was created indicating the likely distance the customer travelled to purchase their groceries in-store. The distance variable is measured in kilometres and was calculated using the shortest possible route via the road network, providing a more accurate measurement than using a straight-line distance or as the crow flies. The inclusion of store distance in this study is integral, as it forms a key indicator of customer shopping behaviour and incorporates spatiality. Studies such as Waddington et al, (2018; 2019) and Wood and Browne (2007) highlight how customers perform grocery transactions in various areas, at various times of the day, and for different shopping missions. Analysing how distance interplays with shopping mission and store choice will provide insight into whether these customers are likely shopping locally or further away, perhaps due to multi-purpose shopping trips or picking food up on the way to/from elsewhere, such as a workplace or leisure activities. The OAC data discussed in section 4.1 has also been joined to the dataset to explore purchasing patterns by underlying area type and the relationship between customer segments and geodemographics. All the variables mentioned, coupled with the transaction date and time, allow for a thorough summary of how cardholders transacted individually, across space and time, and using various shopping channels for different shopping baskets.

**Table 4.3** Data structure of the transaction dataset provided by Sainsbury's Nectar prior to pre-processing.

| Variable | Description |
|---|---|
| **Transaction ID** | A unique numeric ID number assigned to the transaction |
| **Customer ID** | An anonymised loyalty card ID number linked to the cardholder. |
| **Transaction date and time** | Date and time of the transaction. |
| **Channel** | The channel used for the transaction (in-store or online). |
| **Sales value** | The total cost of the transaction (£). |
| **Basket type** | The shopping mission classification, supplied by Sainsbury's. |
| **Store ID** | Identifies the store in which the transaction took place (in-store) or at which the order was packed (online). |
| **Output Area** | The 2011 Output Area of the customer's home address. |
| **Distance via Road Network** | The distance (kilometres) between the customer's OA population weighted centroid and the store the transaction took place following the road infrastructure. |
| **Store ID** | A unique numeric ID capturing the store at which the order took place. |
| **Store type** | Defines whether a store is a Sainsbury's Local (smaller, convenience stores) or Supermarket. |
| **Store coordinates** | The longitude and latitude of the store. |
| **Floorspace** | An indicator of the size of the store (m2) at which the transaction took place. |

## 4.2.2 Sainsbury's supply-side data

Sainsbury's also provided a list of their West Yorkshire stores, for a time period concurrent with the demand side transaction records. For each store, these data included information on store type (supermarket or convenience), coordinates, and floorspace Table 4.3). A total of 55 Sainsbury's stores are located within West Yorkshire, with 50% of those stores located in the Leeds LAD (Table 4.4). Over 70% of these stores are classified as 'convenience' branded under the 'JS Local' fascia.

Essentially, these stores are small at less than 3,000 sqm, offer a reduced variety of food products, and cater to smaller shopping baskets. Additionally, the number of stores where online transactions were picked and packed was calculated (6 supermarkets), finding that at least one supermarket in each LAD was used for order fulfilment (Figure 4.2). The locations of these varying store types will be used to build the supply side representation in the individual-based model and to capture the simulated transactions that took place at the stores within the model. Figure 4.2 provides an overview of Sainsbury's store network coverage in West Yorkshire, highlighting where each store is located, including the type of store it is. These stores have been mapped against the urban/rural classification (Office for National Statistics, 2016) at OA-level for context regarding the type of areas that surround the stores. The urban/rural classification has been used for a clearer visual representation of the store locations whilst indicating the types of areas within which they are located. Most Sainsbury's stores are in urban areas, notably convenience stores which are most tailored to customers performing smaller basket type transactions. Supermarkets that fulfil online orders tend to be located on the outskirts of town and city centres; this is done purposefully due to better road network accessibility and are the types of stores that were planning policy compliant during the store wars (section 2.1.2). Some convenience stores are located close to rural villages, which most likely supply goods to locals within the area. Supermarkets that do not provide online order fulfilment are in a variety of locations both in town and city centres and in suburban areas. One unique store is the Leeds Station Local, located inside the train station. In this case, the store is in close proximity to other Sainsbury's convenience stores, but is likely to serve a distinct purpose, with  most of the transactions that occur here are expected to be performed by those who are commuting via train, either stopping off in Leeds or transacting as part of a longer journey.

**Table 4.4** Count of Sainsbury's stores by type and by Local Authority District. Only a select few stores provide online order fulfilment and are always supermarket store types.

| Local Authority District | Supermarket | Convenience | Total Stores | Online Fulfilment |
|---|---|---|---|---|
| **Bradford** | 2 | 4 | 6 | 1 |
| **Calderdale** | 2 | 2 | 4 | 1 |
| **Kirklees** | 4 | 5 | 9 | 2 |
| **Leeds** | 6 | 21 | 27 | 1 |
| **Wakefield** | 2 | 7 | 9 | 1 |
| **Totals:** | 16 | 39 | 55 | 6 |

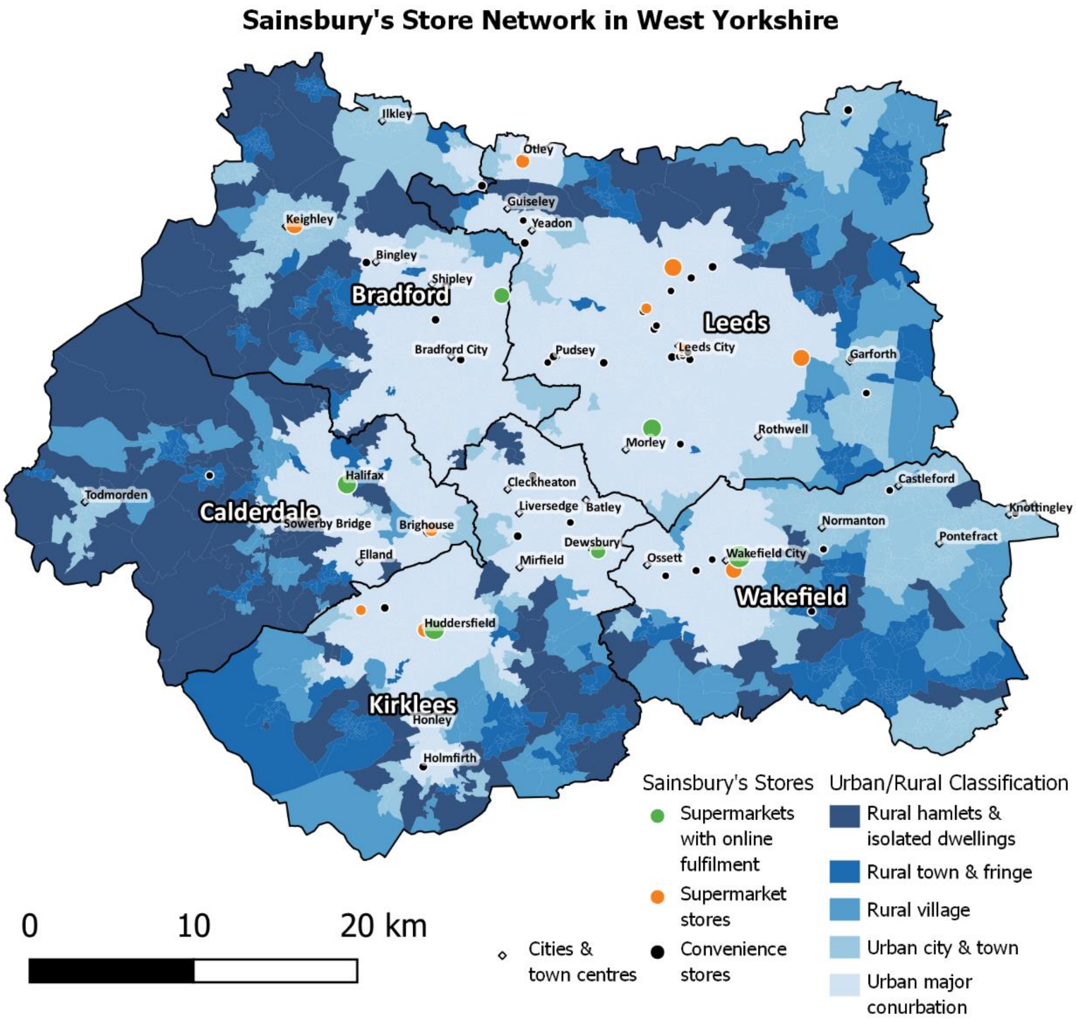



**Figure 4.2** Sainsbury's store network in West Yorkshire with urban/rural classifications at OA-level mapped for geographical context. Data Source: Sainsbury's Ltd., and (Office for National Statistics, 2016).

### 4.2.3 Data pre-processing and variable aggregation

This section of the thesis focuses on the preparation work for segmenting customers in the loyalty card-linked dataset. Consumer typologies are inferred based on their similar transactional behaviours at Sainsbury's stores regarding transaction frequency, the time of day they transacted, the shopping channels they used, and the basket types purchased.

To prepare the loyalty card data for clustering, all data were cleaned and aggregated into variables which best represent the key indicators of customer transaction behaviour, as discussed in section 2.3. The first step was to separate customers by their channel choices: those who solely shopped in-store, those who only shopped online, and those who performed multi-channel transactions. The customer behaviours between these channel groups vary and would skew the cluster behaviours if all were included in the same clustering process later. For example, whilst the author was working at Sainsbury's (see Chapter 1), it was found that all online transactions are time-stamped the moment the order is picked and packed for delivery, which was almost always in the morning of delivery. However, the delivery itself may take place at any time that day. If all online transactions were included in the clustering process of customer behaviour, it would assume that specific customers have an exceedingly high likelihood of shopping in the morning. The usual shopping behaviours of transacting online are vastly different to in-store, where for online, the time of day is not insightful, and the shopping basket is most often classified as 'main' due to the minimum order value and delivery cost (Sainsbury's.co.uk, 2023a). Additionally, distance does not pertain to the distance the customer travelled, but more so what a Sainsbury's driver had travelled between fulfilment store and customer address. Although distance for online transactions were excluded for the clustering calculations, they have been included for data analysis from this section onwards as it provides insight into the store assignment behaviours of online orders by Sainsbury's.

Therefore, using an automated procedure for each channel group (online, in-store or multi-channel), transactions were aggregated temporally into day type (weekday or weekend), time of day (morning, afternoon, or evening), and labelled each transaction by channel (in-store or online), and basket type (for now, top-up or main). The day variables were aggregated for a more general temporal overview, as customer behaviours were similar over the five weekdays and the two weekend days. The time of each transaction was aggregated to the time of day to provide more valuable insights. The time of day variable aggregated each transaction into three segments; morning covers the hours from 12am until 12pm, afternoon is from 12pm until 6pm, and evening is from 6pm until 12am. A nighttime segment was considered, which groups transactions from 12am until 6am; however, these transactions accounted for less than 0.1% of all transactions in the dataset, since most stores are closed during this period. Therefore these transactions were assigned to the morning period. The average number of weekly transactions for each customer was calculated using the date and time variables to identify shopping frequency. All variables used and generated are presented in Table 4.5, which were carefully chosen as they represent the most suitable indicators of shopper behaviour based on domain knowledge, experience of the location planning sector, and insight from Sainsbury's.

**Table 4.5** Transaction dataset variables after data processing and cleansing.

| Variable | Mean values for 12-weeks |
|---|---|
| Total number of transactions. | 12.3 |
| Total number of transactions per week. | 1.8 |
| Average distance between customer home and store(s) visited. | 5.5 kilometres |
| Number of transactions that took place in-store or online. | In-store: 12.1<br>Online: 5.5 |
| Number of transactions that took place on a weekday or weekend. | Weekday: 9.3<br>Weekend: 4 |
| Number of transactions that took place during a morning, afternoon, or evening. | Morning: 5<br>Afternoon: 3.9<br>Evening: 7 |
| Number of transactions classified under basket type categories as ‘for now’, ‘top-up’, and ‘main’. | For now: 6.4<br>Top-up: 5.5<br>Main: 4.4 |

In the transaction dataset, in-store shoppers comprised 94% of all shoppers, 4% were multi-channel shoppers, and 2% were online-only shoppers. For all LADs besides Leeds, online shoppers encompassed 2% of all Sainsbury’s customers, 5% were multi-channel shoppers, and 93% shopped in-store only. The Leeds LAD, however, had 1% of its customers who shopped online only, 4% that shopped both online and in-store, and 95% of its customers shopped in-store only. These slight differences are potentially due to Leeds having more Sainsbury’s stores in the accessible city centre and within suburban areas, thus boosting in-store trade (Figure 4.2).

Before preparing the dataset for clustering, overall consumer behaviours are analysed to understand better what transactions Sainsbury’s customers in West Yorkshire performed between May and July 2018. The following section provides an overview of consumer behaviours at the aggregate level, followed by a section focusing on analysing those consumer behaviours at store level. The analysis of these data is used to support the identification of individual consumer type groups generated via the clustering process in section 5.5.

### 4.2.4 Transaction data: customer overview

To begin with the top-level overview of customer behaviours, Table 4.6 summarises of the total number of transactions that took place over the 12 weeks, split by day type and channel.

**Table 4.6** Total number of transactions split by day type and channel over the 12-week period.

| Day Type | Channel | Total | |
|---|---|---|---|
| | In-Store | Online | In-store |
| Weekday | 1,936,123 | 49,950 | 1,986,073 |
| Weekend | 653,474 | 19,406 | 672,880 |
| Total | 2,589,597 | 69,356 | 2,658,953 |

As shown in the table, 2.59 thousand (97.4%) of all transactions took place in-store, with the remaining being performed using the online channel. At the individual transaction level, in-store transacting is a more common method of purchasing groceries. Of the 12,600 online customers, 18.6% had performed only one online purchase over the 12 weeks. These customers have no observable pattern in their online purchasing and are 'by chance' customers. Approximately 17.7% had consistently transacted online at least once a week or more; these are consistent customers who habitually purchase their groceries online every week. Due to online transacting often being high value, it is sensible that these make up a small proportion of all observed transactions. Although, during 2018, when this data was recorded, online was one of the fastest growing shopping channels in the UK grocery market and made up 7% of all grocery transactions (Mintel, 2019). In the same year, the Leeds LAD was found to have a lower usage of online transacting compared to other areas in Yorkshire and the Humber despite having a considerably large number of grocery stores (Kirby-Hawkins et al., 2018; Hood et al., 2020). As Leeds contained around 35.8% of all Sainsbury's customers, this could be why so few transactions took place online compared to the UK average. Other factors may impact this, such as the time of year the dataset is from; seasonality impacts are not captured which could present alternative results, customers transacting outside the West Yorkshire boundary, or Sainsbury's customers transacting at other grocery retailers for their online food shops. For in-store and online transactions, most transactions occurred on a weekday at 74.8% and 72%, respectively. As the weekend only comprises of 2-days in a 7-day week, fewer transactions during the weekend are expected. Overall, most transactions took place in-store and on weekdays.

Considering that the proportions of transactions split by day type and channel vary, Figure 4.3 provides context regarding when those transactions took place using the time of day variable. The time of day purchasing patterns for in-store transactions are similar between weekdays and weekends, with over 50% of transactions occurring during the afternoon. Despite the afternoon period only covering 6 hours, the dataset suggests that this is the most common time for customers to perform grocery transactions. The morning time of day period accounted for almost 1/3rd of in-store transactions, and the evening period experienced much less for both weekdays and weekends. Considering these data are at the transaction-level, it is integral to analyse the type of baskets purchased at these varying times of the day. The high volume of afternoon transactions is from customers buying food 'for now' or 'top-up' baskets due to their lower monetary value and containing food to be eaten soon.

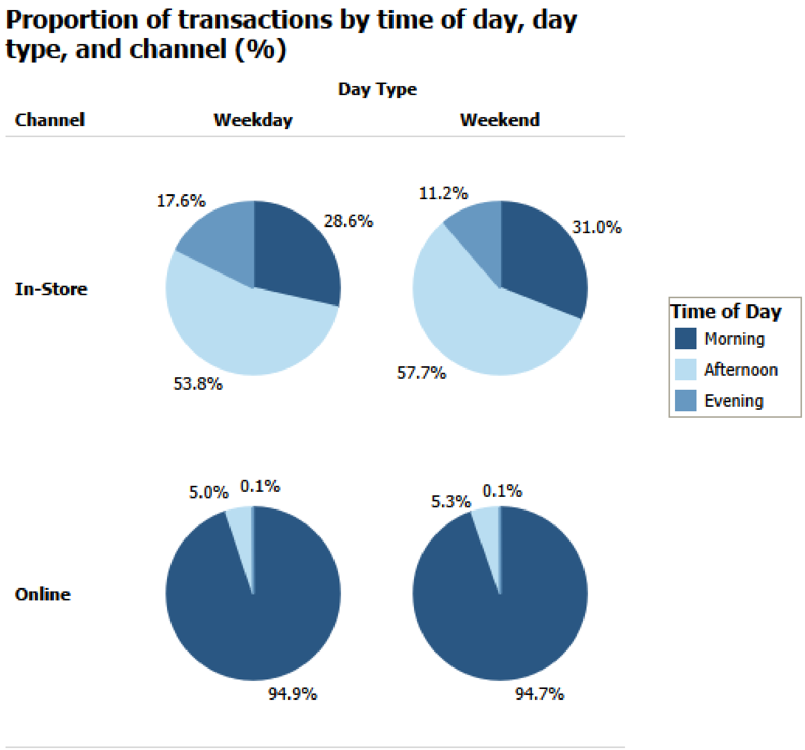



**Figure 4.3** Proportional breakdown of transactions by day type, channel and time of day (%).

Some general assumptions can be made from the data presented in Figure 4.3 and academic literature regarding the temporality of consumer purchasing (Elms et al., 2010; Thompson et al., 2012; Berry et al., 2016). Firstly, focusing on weekends, it is essential to acknowledge that all supermarkets on Sundays have much shorter and stricter opening hours due to UK Sunday trading restrictions in England (gov.uk, 2022). Therefore, fewer transactions would be observed during the weekend evening and instead would take place in the morning or afternoon. Secondly, focusing on weekdays, it is vital to acknowledge the typical working hours of employed customers. Sainsbury's does not hold personal data for their customers; therefore, assumptions can only be made regarding where customers are located during different time periods. On weekdays, many customers are most likely to be at places of work for half of the morning and most of the afternoon, on the assumption of a standard 9-5 working day. A significant proportion of transactions during these hours are most likely frequently purchased ‘for now’ and ‘top-up’ baskets in which customers purchase food for their breakfasts or lunches, or purchase food to be taken home to be eaten later that day.

In comparison, around ~94% of online transactions occurred during the morning regardless of day type due to how Sainsbury’s records online purchases in their database. During the internship at Sainsbury’s, it was found that almost all online transactions are processed on the morning of the delivery day once payment has been received. Although the grocery delivery may happen anytime during the day until 11pm, the payment is almost always taken in the morning or once the order has been picked and packed for delivery. From this analysis, it can be determined that day type and time of day variables are less significant for online transactions but are important for in-store transacting.

Despite having rich information for these transactions, little information can be retrieved for online transactions. Online is an ever-growing channel, notably during the Covid-19 pandemic, yet it is not easy to mine unique observations about these transactions and the types of customers who perform them. Further data regarding who these customers are such as demographics would be helpful, or understanding how online transactions are allocated to stores for picking and packing.

Temporally, consumer behaviours differ between weekdays and weekends, and with time of day, notably for in-store transactions. Table 4.7 provides a breakdown of in-store transactions, including the basket type variable and the time of day variable, for a temporal insight into the types of transactions made. The basket type variable indicates the purpose of the transaction with, ‘for now’ being food products often purchased ‘on the go’. ‘Top-up’ baskets contain food that will be consumed within a few days or a mix of ‘on the go’ and ready to eat food products, and ingredients that form parts of larger meals. ‘Main’ baskets contain food for many days and are high-value transactions that are often performed on a more habitual basis. Focusing primarily on weekday mornings, most transactions were categorised as ‘for now’ at 13.3%, followed by ‘top-up’ baskets at 10.2% (Table 4.7). As stated, these transactions are most likely to be food purchased for on then go or to be consumed later in the day. It is assumed that these transactions predominantly take place near the customer’s place of work.

Similarly, most afternoon transactions were categorised as ‘for now’ and ‘top-up’ baskets, at a much higher percentage compared to mornings. These frequent transactions comprise most of the transactions in the dataset and occur during the typical lunch break between 12p and 2pm. Evening transactions took place less often in general and still followed a similar trend regarding basket type. The top-level analysis of this dataset primarily shows that the most common time for a transaction to take place is during an afternoon, for ‘for now’ and ‘top-up’ baskets. Similar findings were found when analysing weekend only in-store transactions regarding basket size and time of day.

**Table 4.7** Proportion of in-store transactions for basket type split by day type and time of day variables.

| Basket Type | Weekday In-store | | | Total |
|---|---|---|---|---|
| | Morning | Afternoon | Evening | |
| For now | 13.3% | 23.0% | 7.9% | 44.2% |
| Top-up | 10.2% | 22.0% | 7.2% | 39.4% |
| Main | 5.1% | 8.9% | 2.4% | 16.4% |
| Total | 28.6% | 53.8% | 17.6% | 100% |

These frequent transactions will make up a large part of consumer behaviours; however, it is hypothesised that there are clear segments of customers who perform these common transactions alongside other types, based not only on temporality but also spatiality. These observations will be considered when segmenting customers into consumer type groups in Chapter 5.

To incorporate spatiality in the analysis, Figure 4.4 presents the variation in the number of transactions over distance (km) between customer homes (using PWC OA points) and the stores they visited split by day type and time of day. The maximum distance observed in the dataset is 52km; however, these are not included in the figure as the same pattern continues from 20km onwards. The three most common distances observed in the dataset for all transaction types occurred at 0.3km, followed by 0.5km and 1.4km. Customers tend to travel locally for their groceries regardless of when they transact, and their transactions further away from home become less frequent. All transactions follow a similar pattern, with peaks at 2km, 4.5km, and become less dramatic from 7.4km onwards. The distance variation becomes much less drastic from 2km onwards for weekend evening transactions, though these transactions occur the least often. As highlighted in the literature (Byrom et al., 2001;

Waddington et al., 2018), customers tend to purchase groceries at their most local store, or favourite brand. This is observed in the dataset as customers mostly shop locally at all times of the day. The weekday afternoon transactions, however, have a much further distance overall, correlating with the hypothesis that customers are making smaller value transactions while at work and not near home. The increase in transactions at 4.5km could highlight a group of customers who reside within a specific area who all tend to travel to a specific city or place for their afternoon baskets. The clustering process will likely highlight these groups in the next chapter. At the aggregate level, it is challenging to identify nuances in the dataset due to the overbearing number of weekday afternoon transactions. Therefore, the following section considers the dataset from the store-level perspective. These analyses will be used to better understand the consumer type groups identified later in this thesis.

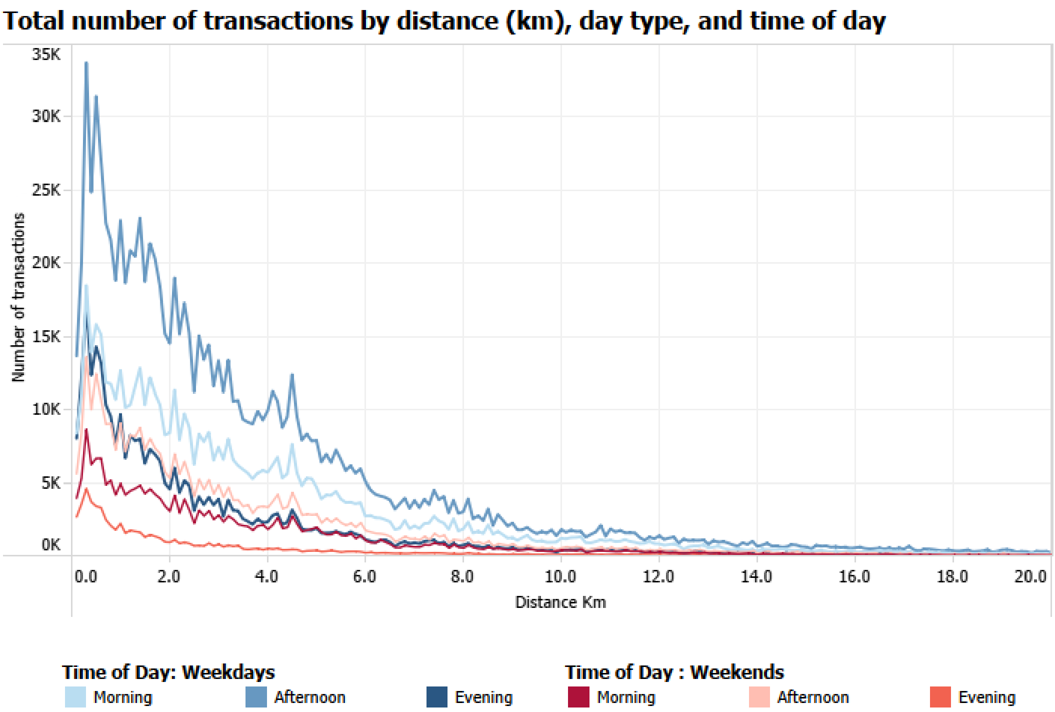



**Figure 4.4** Total number of transactions by distance (km) split by day type and time of day.

### 4.2.5 Transaction data: store-level overview

To grasp the relationship between store performance (i.e., the number of transactions that took place) over the 12 weeks, Figure 4.5 presents an overview of supermarkets that fulfil online orders. Around 99% of the transactions that took place online were categorised as basket type ‘main’; this is likely due to the minimum order value of online orders and delivery cost increasing the transaction price (Sainsbury’s.co.uk, 2023a). Removing the delivery cost from the online transactions to calculate the true value of the basket was considered, however, Sainsbury’s use variable pricing for its delivery slots, hindering further insight. Figure 4.5 shows that more transactions occur on a weekday than a weekend because there are only two weekend days within the 7-day week. Despite there being over half the number of weekdays, weekends still experience a relatively large number of transactions. Two stores that stand out the most amongst the rest in terms of the total number of transactions (weekday and weekend together): White Rose in Leeds’ White Rose Shopping Centre, which delivered 19,029 transactions, and Shorehead in Kirklees (next to Huddersfield town centre label in Figure 4.2) which delivered 17,632 transactions. These stores, and the Halifax town centre store in Calderdale delivered the farthest transactions compared to all other stores. These three stores act as hubs for online delivery to households within West Yorkshire, as they are also some of the largest stores by sales area (Figure 4.2). Shorehead stands at 63,200 $ft^2$, followed by Halifax at 52,600 $ft^2$, and White Rose at 50,600 $ft^2$.

Wakefield Marsh Way has the largest store size within the dataset, with an enormous 78,000 $ft^2$ sales area, yet it does not deliver the most online transactions. Perhaps due to being located relatively close to White Rose, Wakefield prioritises home deliveries on the outskirts of West Yorkshire, in which those transactions are not captured in the dataset. White Rose and Shorehead stores generated the most revenue during the 12-week period in terms of online Nectar card linked sales, representing 52.9% of online revenue, with the rest split across four stores. The type of customers that shop at these stores for online orders will be identified in Chapter 5, linking their transactions to stores spatially. The analysis of this dataset is used to support the model development process in Chapter 6 when assigning online transactions to stores based on time of day and for assigning transaction value.

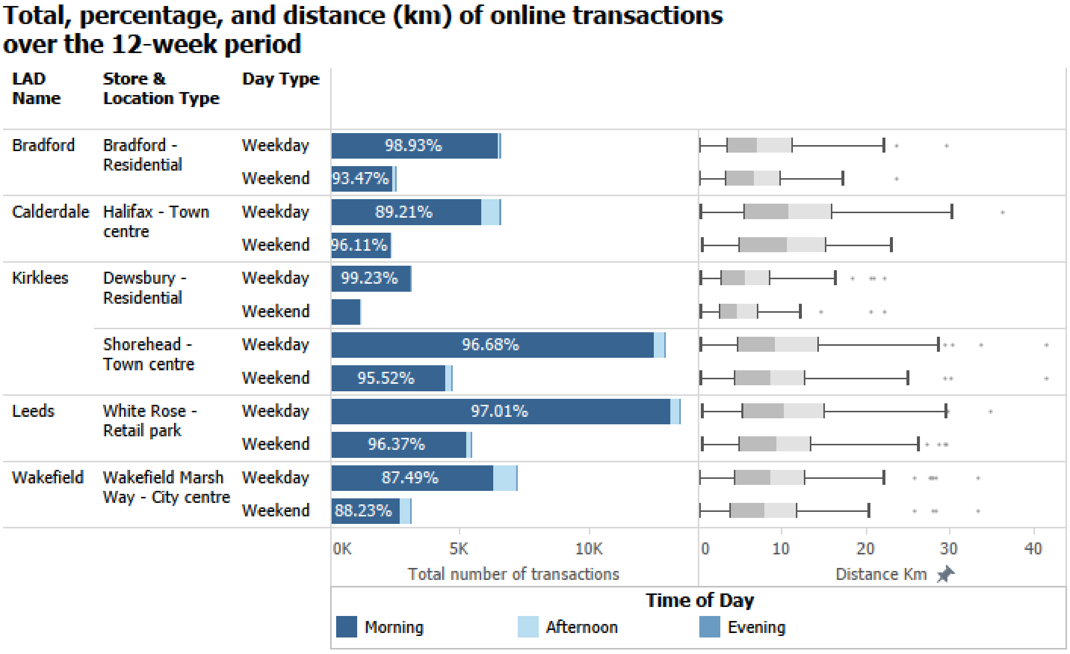



**Figure 4.5** Total number and percentage of transactions that took place online, including distances between the customer's home address and the store delivered their online order (km). Bar length shows the total number of transactions, bar text and colour show the proportion regarding time of day.

Focusing primarily on in-store transactions, Figure 4.6 captures the proportion of transactions in different LADs. For all LADs, at least 76.9% of all transactions occurred at a store within the same LAD, with most districts having a much higher proportion. Customers in Bradford transacted in a broader range of districts compared to others, with their second highest transacted district being Leeds at 13%, followed by

Calderdale at 8.6%; this is likely due to the location of Sainsbury's stores in Leeds and Calderdale along Bradford's district boundary (Figure 4.2). Bradford has many urban areas without a Sainsbury's store, which may be why these customers seek further afield for their groceries. The least interacted districts were Wakefield and Bradford, mostly likely due to these districts not sharing a boundary, therefore not capturing local cross-boundary interactions. Cross-district transactions are frequent in this dataset due to a mix of store locations, store attractiveness, and the fact that district boundaries are arbitrary, with no physical boundary on the ground preventing customer flows. For example, Leeds is a large city that draws customers from all different districts, acting as a hub. 93.6% of Leeds' customers transacted within Leeds, most likely due to the abundance of Sainsbury's stores and having a higher population count. Other districts will likely have more local stores that are competitors to Sainsbury's, drawing customers there instead. Unfortunately, this study cannot access competitor data; therefore, these predictions cannot be analysed but only theorised. The second highest out-of-district flow of transactions is from customers in Wakefield transacting in Leeds at 9.2%, followed by Calderdale customers transacting in Kirklees at 8.5%. These transactions are likely smaller value baskets performed by workers and students inside these small towns.

**Proportion of in-store transactions between customer and store LADs (%)**

| | Customer LAD | | | | |
|---|---|---|---|---|---|
| **Store LAD** | **Bradford** | **Calderdale** | **Kirklees** | **Leeds** | **Wakefield** |
| **Bradford** | 76.9% | 0.9% | 0.3% | 3.7% | 0.1% |
| **Calderdale** | 8.6% | 87.6% | 3.7% | 0.2% | 0.3% |
| **Kirklees** | 1.1% | 8.5% | 88.6% | 0.8% | 4.7% |
| **Leeds** | 13.0% | 2.7% | 4.7% | 93.6% | 9.2% |
| **Wakefield** | 0.5% | 0.2% | 2.6% | 1.6% | 85.6% |
| **Grand Total** | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**% of total transactions**

0.1% – 93.6%

**Figure 4.6** Proportion of in-store transactions by customers in each Local Authority District into stores in each Local Authority District (%).

Figure 4.7 presents an overview of all transactions that took place in-store in non-residential locations, including city and town centre stores and retail park stores. Over the 12 weeks, Shorehead supermarket in Kirklees experienced the highest % of transactions at 17.2%. Shorehead, on average, also had the most extensive distance travelled between the customer's home (PWC OA point) and the store and had the highest proportion of 'main' basket types for both weekdays and weekends; this is likely due to the store being a hub for Kirklees customers and its large sales area (discussed later in Figure 5.5). In general, all stores had a shorter distance travelled for weekend transactions, reflecting fewer workers travelling into the city compared to weekdays. The most notable finding of Figure 4.7 is the dominance of 'for now' basket type transactions in the dataset. For all convenience stores within city and town centres, between 80% and 94% of transactions were baskets categorised as 'for now', with the other transactions being 'top-up' baskets. The number of 'for now' and 'top-up' baskets reflect the nature of convenience stores with a considerably smaller product assortment and are likely to have smaller provision of car parking, making it difficult for customers to transport 'main' baskets home. Additionally, these stores also have a higher proportion of transactions that take place on weekdays reflecting the influx of worker purchases.

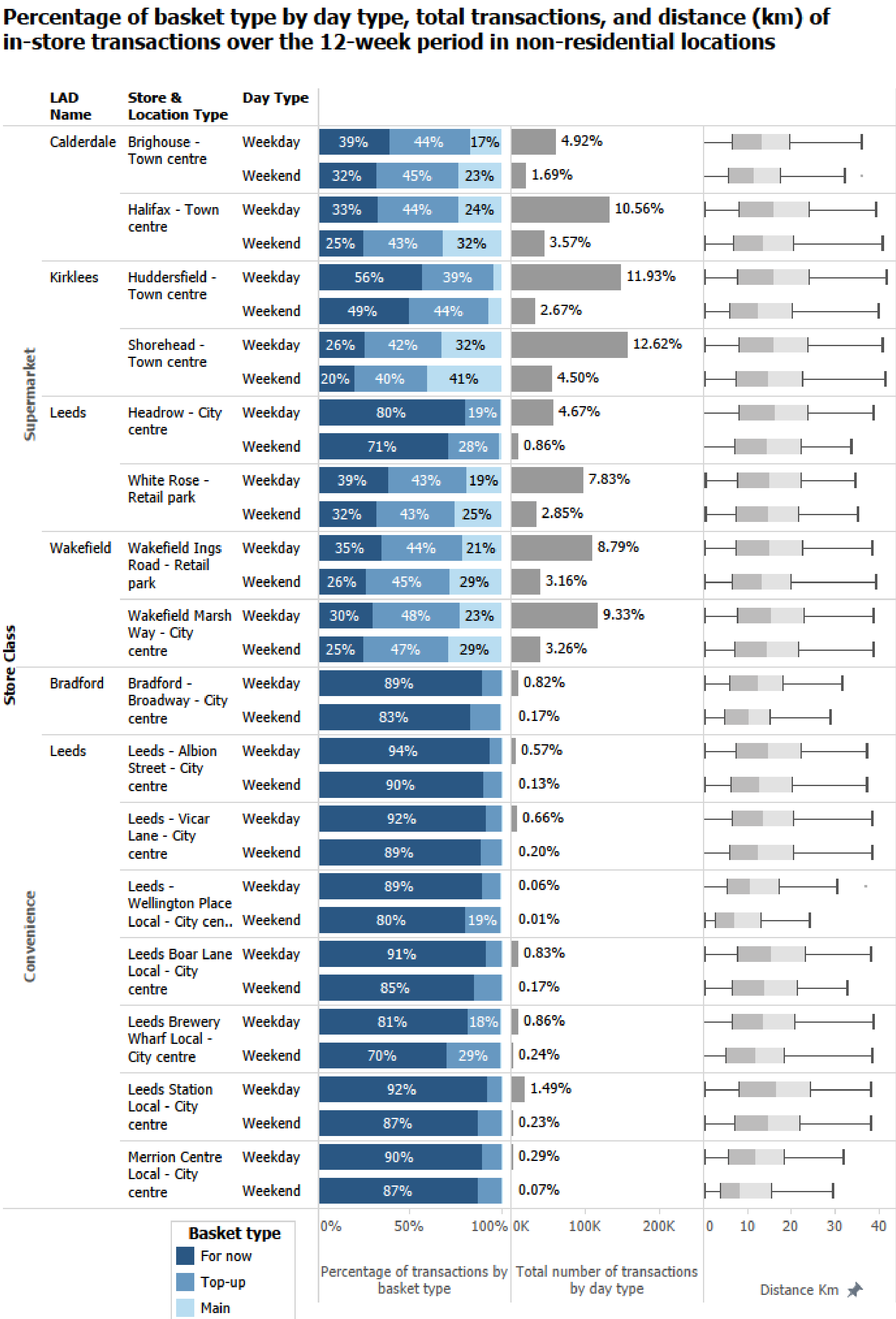



**Figure 4.7** Summary of in-store transactions in non-residential locations split by day type, basket type. Percentages are of basket type by day type, followed by total number of transactions over the 12-week period and finally distance boxplots.

Leeds train station is the busiest train station in the north of England and England's second-busiest outside of London (Office of Rail and Road, 2018). The inflow of customers to this station significantly benefits Sainsbury's as customers looking to buy food 'on the go' have limited store choices (Karamshuk et al., 2013). Customers arriving at Leeds train station will come from areas around Leeds, West Yorkshire, and the entire country. Fortunately, Sainsbury's has a local store within the Leeds train station and has the highest proportion of all transactions out of convenience stores on weekdays (Figure 4.7). The Leeds Station store is unique because it is one of the few stores in the dataset mainly visited during a journey rather than a store close to a customer's home or destination (Hood et al., 2016). Stores like Leeds Station create difficulties when creating an individual-based model, as it is notoriously difficult to simulate individual customers' movements in general, especially when related to consumer behaviour (Torrens, 2023). Figure 4.8 provides a context of where customers of the Leeds Station store resided at the PWC OA-level. Over the 12 weeks, 21,647 individual transactions occurred at this store, coming from 3,281 different OAs across West Yorkshire. Incorporating such a unique store within an individual-based model is not easy, and the difficulties of this are further discussed in Chapter 7.

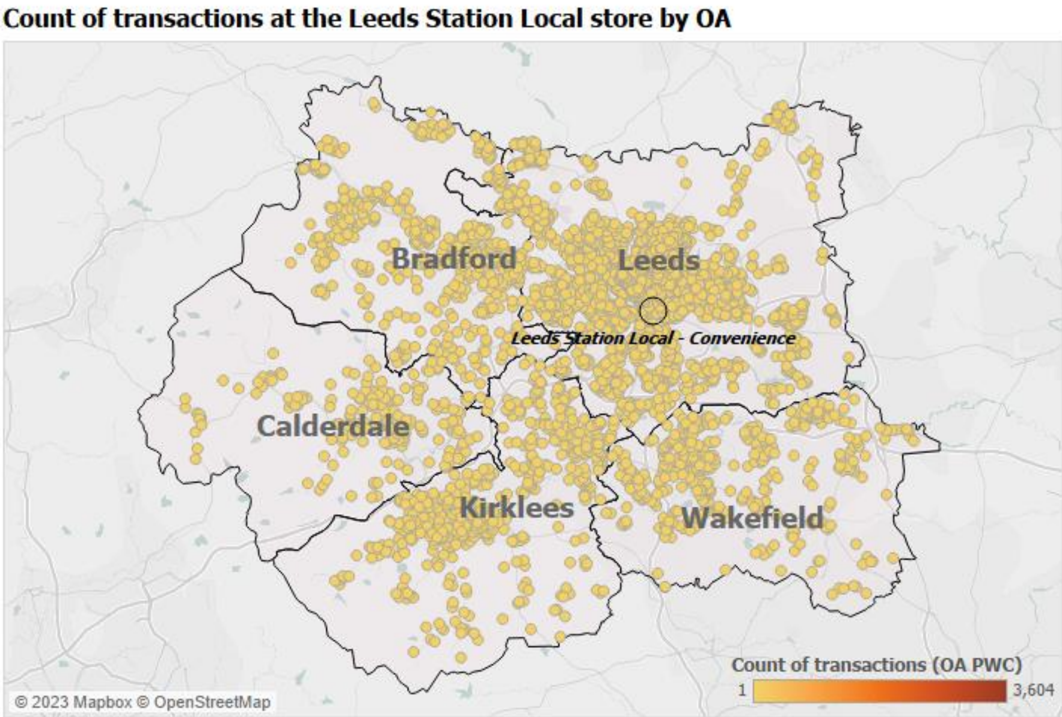



**Figure 4.8** Map of customer transactions counts between OAs and the Leeds Station store. The black outlined circle is the store, and each smaller dot represents PWC OAs where customers reside.

Another key feature of this data is the basket type breakdown for Leeds' Headrow supermarket in the city centre. Despite being a supermarket with a comprehensive product range, this store predominantly sold 'for now' baskets (80%) and only 1% of 'main' baskets. These observations are likely due to the store's location, on a busy street in the city centre with many customers. Similarly to the Leeds Station store, the Headrow Sainsbury's had customers visiting from 3,846 unique OAs across all districts of West Yorkshire. Many of these transactions are performed by different customers and occur once in the 12 weeks at an OA. While customers travel from afar and shop at these stores, this is not necessarily a shopping habit but more of a 'by chance' transaction. These observations would ideally be incorporated in an individual-based model, but deciding customers are making those transactions is challenging.

The final observation of interest is the proportion of ‘main’ baskets at Halifax’s town centre supermarket and Shorehead in Kirklees. Despite being in town centres, both these stores experienced high proportions of ‘main’ transactions on weekdays and weekends. This is likely because these large stores are Sainsbury’s supermarket hubs for the small towns and are located around a cluster of Sainsbury’s customers (Figure 5.5), provide a full range of in-store services, and substantial parking space despite being located in town centre locations. From this analysis, understanding store class is insightful but is paramount in conjunction with store locations and the types of areas they are located. City centre stores are a large target for drawing in workers for ‘for now’ and ‘top-up’ transactions for convenience stores and supermarkets. Supermarkets in city and town centres follow a similar pattern, with some stores capturing more ‘main’ transactions, primarily due to their location and product range.

## 4.3 Chapter 4 summary

The loyalty card-linked transaction data provided by Sainsbury’s are rich regarding where, when, and how and potentially why these particular customers have transacted in West Yorkshire over the three months. Based on those factors, we have been able to make some inferences about why, though we should treat these interpretations with some caution. Despite their richness, these data have some notable limitations. It is acknowledged that the information the loyalty card dataset provides is for only those who transacted at a Sainsbury’s store, and their actions elsewhere at other retailers are unknown. Those customers captured in the dataset will shop at other retailers for groceries, and the profiles of those transactions may vary greatly compared to their behaviours at a Sainsbury’s store. Additionally, the customer’s transactions at other Sainsbury’s stores outside the case study area are not captured, even if they transacted at a store that falls just outside West Yorkshire’s boundaries. The modifiable aerial unit problem (MAUP) likely skews customers assumed behaviours as customers who made few transactions over the 12 weeks may also transact at a Sainsbury’s store just outside the county boundary. To combat the issue of MAUP, stores located within a buffer zone around West Yorkshire may have been incorporated; however, the study collaborator did not provide this data. It was agreed that a closed study of these customers and stores was sufficient for this study and the collaborator’s needs. Finally, as discussed in Rains and Longley (2021) the loyalty card transaction dataset only captures those transactions in which the customer scanned their Nectar card. Not all customers use their Nectar cards consistently, whether due to the scanner being broken, saving time by not scanning their cards, or not feeling like it is worth scanning when purchasing lower value baskets such as ‘for now’ basket types (as discussed in section 3.3.1).

The transaction dataset used in this study has clear benefits and insights into known customer behaviours, linking each transaction to an individual cardholder. The limitations of these data are considered throughout this thesis’s analysis, notably during model building and discussion. This novel dataset allows researchers to explore what *we can* observe and use these empirical data to create empirically validated models. This concludes chapter 4 and ultimately achieves aim 2 and objective 3.

# Chapter 5 Inferring consumer typologies

The following chapter discusses the methodological approach to segment the customers observed in the loyalty card-linked transaction dataset into distinct consumer type groups based on their shopping behaviours. This chapter particularly supports the attainment of the second aim outlined in section 1.2; via the completion of objective 4:

*To segment the observed customers into consumer type groups based on their key behaviours regarding transaction frequency, store and channel choice, and purchase purpose, i.e., basket type.*

As discussed in section 2.3, customer store choices regarding store location, the shopping channel used, and the basket type purchased are key indicators of consumer behaviour, all in conjunction with time. Customer behaviours vary from day to day as shopping needs differ. However, every week, there tends to be a pattern in habitual shopping behaviour, such as the 'main' weekly shop, whether online or in-store. The segmentation of the customers in the dataset aims to identify distinct grocery consumer type groups, with each differing in their consumer behaviour, whether that distinction is in the channels used, the times they make their purchases, how often they purchase groceries, or how far they travel between home and stores.

The attainment of this aim allows for direct progression toward the completion of aim 5:

*To perform data mining on the customer segments, identifying the probability of making a transaction at any given time based on their linked loyalty card transactions.*

The customer type groups identified in section 5.5 are analysed to derive probabilities for the individual-based model.

## 5.1 Preparing for customer segmentation

The different customer groups by channel (in store versus online) were first investigated to prepare the loyalty card-linked transaction data for clustering. As previously mentioned, online-only customers made up 2% of the 216 thousand unique customers in the dataset, and multi-channel shoppers made up 4%. As these groups both represent such small percentages of the customer database, each group was deemed an individual cluster in their own right. These groups were still clustered for testing to explore whether there are distinct groups within them; although, little variation was found and were ultimately deemed as individual clusters. Online-only shoppers had variables/shopping indicators removed before clustering due to needing to be more suitable and accurate. The variables not included for online shoppers were the distance between home and store(s) visited as these customers did not travel themselves (this information does provide insight into how Sainsbury's may assign their online picking and packing stores to different OAs, but this is beyond the scope of this thesis). The time of day variable was also not included, as mentioned previously in section 4.2.3; these data are almost always assigned as a morning transaction regardless of when the order was delivered. The multi-channel shoppers had all variables included from Table 4.5 for clustering, and for transactions that took place online, distance and time of day variables were set to *null* to not be included in the clustering as this would distort the results due to the prevalence of both online and in-store transactions by each customer. Finally, the in-store-only shoppers had all shopping indicators included, as each one applies to every transaction made.

Using Pearson's coefficient, the relationships between variables were explored to ensure the data were logical (Gautheir, 2001). For example, positive correlations were found between 'main' baskets (large value transactions containing ingredients for large meals) and online transactions and negative for ‘for now’ baskets and online transactions. Due to the minimum spend of online shopping baskets, a transaction should not be classified as an online and a ‘for now’ shopping basket unless captured as a user error or another unknown reason. The relationship between weekend and evening transactions was understandably weak due to the UK’s trading hour restrictions on Sunday (gov.uk, 2022). No significant peculiarities were found within the data; all variables were carried over to the clustering process.

## 5.2 Customer segmentation using k-means clustering

A popular method of identifying clusters within data is k-means clustering. K-means clustering is an unsupervised algorithm that makes inferences from data using only input vectors, grouping similar data points to discover underlying patterns (Macqueen, 1967; Lloyd, 1982; Kanungo et al., 2000). A set number of clusters must be pre-determined for this method; then, each data point is classified by the computed distance in feature space between all other data points and cluster group centres using the mean averages; the point is assigned to a cluster group depending on which centre point is the closest. Each group centre is then recomputed by taking the mean of all vectors in the group. This is repeated over several iterations until comparable results are continually reproduced. K-means clustering has proven to be successful with a variety of geodemographic datasets (Gale et al., 2016; Hincks et al., 2018; Burns et al., 2018), including customer segmentation (Sturley et al., 2018; Syakur et al., 2018; Kansal et al., 2018), and was therefore used within this study. Alternative clustering methods were considered, such as hierarchical clustering and expectation-maximization Gaussian mixture models, which both have unique benefits. However, they both have major drawbacks regarding time complexity, inefficiency, and inability to operate with large datasets (Qi et al., 2016). K-means clustering offers a much shorter execution time, much tighter clusters, and requires a lower amount of computational power for vast datasets (Kanungo et al., 2000). Considering the transaction dataset contains 2.6 million transactions, k-means was deemed the most suitable method. One of the most important user-driven decisions when applying k-means is the determination of the optimal number of clusters (k). The next section outlines how the number of clusters was determined in this application.

## 5.3 Determining the number of clusters

The k-means clustering algorithm was applied to each customer group: online customers, multi-channel, and in-store. Due to the small number of customers in the online group, the k-means clustering algorithm struggled to identify distinct and unique groups; therefore it has been considered as a single cluster. The algorithm also struggled to cluster the multi-channel customers for the same reasons, identifying minor variations in behaviours between those who use both shopping channels. Consequently, the in-store-only group was put through the clustering algorithm, which successfully identified multiple groups that did express differing shopping behaviours.

The k-means clustering algorithm was executed to consider 2 to 10 clusters and were analysed using both the Adjusted Rand Index (ARI) (Yeung and Ruzzo, 2001) and Calinski-Harabasz Index (CHI) (Klein et al., 2018) methods as set out in (Putler and Krider, 2012). Before executing the algorithm, z-scores were calculated and used for each variable to standardise the data, as the variables are a mix of percentages and counts. The k-means clustering algorithm was executed seven times, each run using 100 bootstrap samples of the dataset. Each run of the algorithm created multiple replicates of the dataset to infer the uncertainty and confidence intervals related to the data to determine the number of clusters to use in the k-means clustering algorithm. More samples and executions of the algorithm will result in a higher accuracy of the chosen cluster groups; however, it does cause exponential growth in processing time. After running the algorithm, it was found that 100 bootstrap samples at seven repetitions resulted in an ideal trade-off between processing time and tight clusters for this dataset after calculating the ARI measure (Figure 5.1). To assess the output of the k-means algorithm, box plots were created for the ARI and CHI, presented in Figure 5.1. The Adjusted Rand Index measures the similarity of data points within each cluster, indicating how stable and reproducible the clusters are using various samples and starting seeds (Santos and Embrechts, 2009). Values closer to zero indicate more dissimilar clustering; therefore, the higher the value, the more similar the data are within the cluster. The clustering output suggests that there are two distinct segments of customers within the dataset at means of 0.97 and 0.94 and that the next best solution is segmenting into 5 clusters with a mean of 0.73. To further support the decision of choosing the number of clusters, the Calinski-Harabasz Indices are considered. The CHI measures how separated the clusters are; the higher the value, the more distinct the clusters are (Klein et al., 2018). Like the ARI analysis, the 2-cluster solutions produced the most considerable distinctions between the clusters, with the following best being 5 clusters (Figure 5.1).

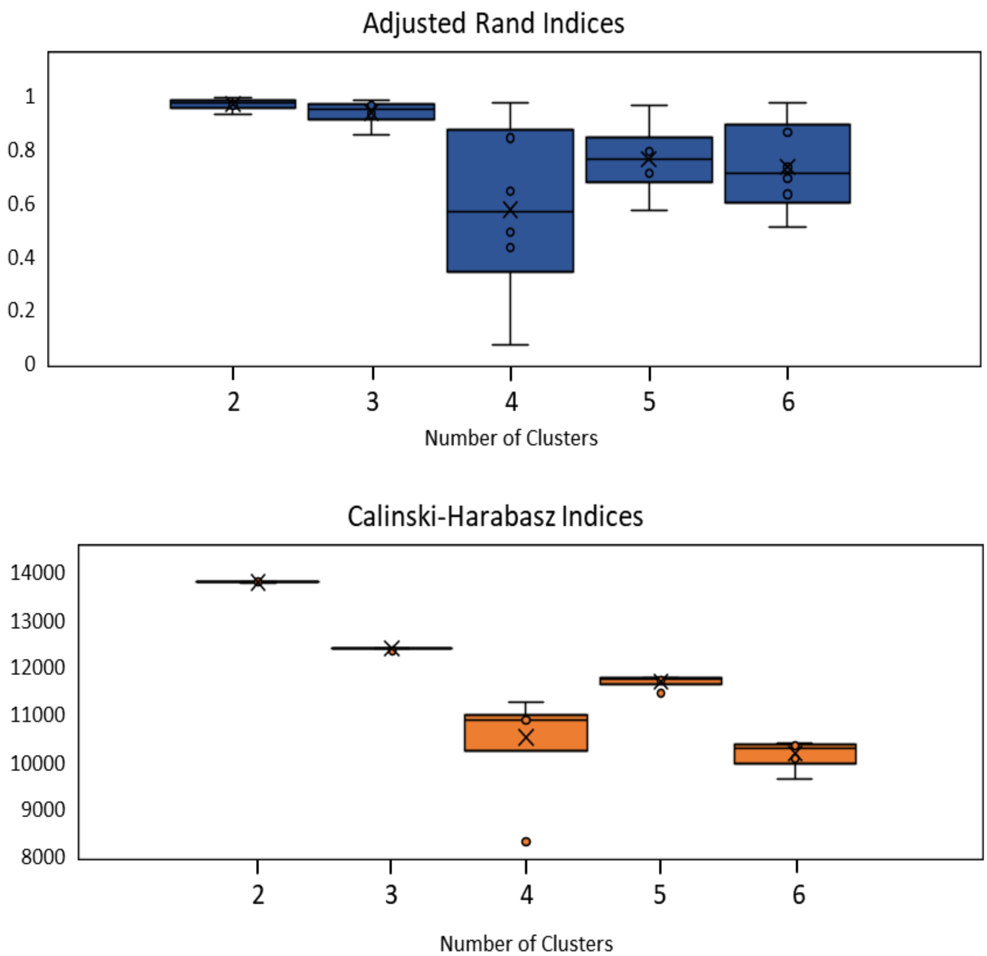



**Figure 5.1** Adjusted Rand Index and Calinski-Harabasz Indices Boxplots for consumer type classification. These are for in-store customers only as multi-channel and online clusters have been established separately.

Whilst the 2-cluster solution is suggested as the best solution by the ARI and CHI results, running this solution of the entire dataset and investigating the results merely points to this approach being able to separate customers who shop at Sainsbury's more and further away from home. This finding highlights that store distance and perhaps brand preference are notable factors in distinguishing differences between customers' shopping behaviours and will make up key segments of customers in this dataset. However, this separation provides little additional colour to the variety of customer behaviours that the study wishes to identify and model. That is, to discover behavioural traits amongst store distance and shopping frequency, channel choice, basket type and time. Further investigation of the results from each cluster runs from 3 clusters to 10, resulting in choosing a 5-cluster solution. It was found that when using less than 5 clusters, slight variation was captured regarding time-related and basket type variables. Output using 6 clusters resulted in some segments with very few customers and therefore did not capture the generalised behaviours of groups of customers but rather the atypical behaviours of unique individuals. In doing so, this meant that the clusters were easily interpretable and valuable for examining shopper behaviour across the input variables. Coupled with the 5-cluster solution being amongst the best scores for the AR and CH indices, it was decided that this would be the best choice of clusters for our modelling purposes, which specifically sought large-enough customer groups to be meaningful in a location planning context.

## 5.4 Customer segmentation output overview

All 216 thousand customers have been successfully segmented into a consumer type group based on their grocery purchasing behaviours over the 12-week period. Figure 5.2 summarises the overall behaviours of the segmented Sainsbury's customers, presenting the proportions of transactions under the different combinations of key consumer shopping indicators. At first glance, there is evident variation between consumer type groups, with group 1 being online-only shoppers, group 2 being multi-channel shoppers, and the rest being in-store-only shoppers. The k-means clustering algorithm successfully identified varying behaviours between those who transacted in-store, especially against all key shopping indicators.

**Summary of all consumer type groups broken down by shopping indicators over the 12-week period (%)**

| Day Type | Time of Day | Channel | Basket Type | Consumer Type Group 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| Weekday | Morning | In-Store | For now | | 5.81% | 8.95% | 3.15% | 3.83% | 5.26% | 19.66% |
| | | | Top-up | | 4.74% | 8.53% | 3.14% | 3.87% | 11.69% | 4.95% |
| | | | Main | | 2.91% | 1.63% | 1.81% | 2.09% | 24.34% | 0.85% |
| | | Online | For now | 0.00% | 0.00% | | | | | |
| | | | Top-up | 1.06% | 0.15% | | | | | |
| | | | Main | 68.88% | 22.00% | | | | | |
| | Afternoon | In-Store | For now | | 10.70% | 17.46% | 7.15% | 8.65% | 6.04% | 28.25% |
| | | | Top-up | | 11.26% | 18.76% | 11.67% | 15.13% | 12.31% | 12.51% |
| | | | Main | | 5.58% | 4.48% | 9.39% | 10.07% | 22.85% | 2.74% |
| | | Online | For now | | 0.00% | | | | | |
| | | | Top-up | 0.02% | 0.01% | | | | | |
| | | | Main | 2.89% | 1.34% | | | | | |
| | Evening | In-Store | For now | | 3.32% | 6.76% | 5.81% | 3.26% | 0.94% | 6.86% |
| | | | Top-up | | 3.52% | 5.88% | 14.72% | 5.42% | 1.24% | 3.46% |
| | | | Main | | 1.62% | 0.75% | 20.25% | 2.14% | 0.80% | 0.54% |
| | | Online | Main | 0.07% | 0.03% | | | | | |
| Weekend | Morning | In-Store | For now | | 1.53% | 3.16% | 1.14% | 0.99% | 1.70% | 3.83% |
| | | | Top-up | | 2.01% | 3.60% | 1.84% | 1.99% | 2.65% | 1.70% |
| | | | Main | | 1.72% | 1.88% | 2.31% | 4.60% | 2.91% | 0.76% |
| | | Online | For now | | 0.00% | | | | | |
| | | | Top-up | 0.30% | 0.06% | | | | | |
| | | | Main | 25.42% | 8.86% | | | | | |
| | Afternoon | In-Store | For now | | 2.95% | 5.72% | 2.74% | 3.00% | 1.63% | 6.29% |
| | | | Top-up | | 4.45% | 7.25% | 6.10% | 8.34% | 2.85% | 4.06% |
| | | | Main | | 3.18% | 1.91% | 4.61% | 24.20% | 2.18% | 1.15% |
| | | Online | For now | | 0.00% | | | | | |
| | | | Top-up | 0.01% | 0.01% | | | | | |
| | | | Main | 1.32% | 0.51% | | | | | |
| | Evening | In-Store | For now | | 0.74% | 1.68% | 0.88% | 0.50% | 0.21% | 1.46% |
| | | | Top-up | | 0.69% | 1.34% | 1.84% | 0.83% | 0.27% | 0.76% |
| | | | Main | | 0.30% | 0.26% | 1.46% | 1.09% | 0.13% | 0.16% |
| | | Online | Top-up | 0.00% | | | | | | |
| | | | Main | 0.02% | 0.01% | | | | | |
| Tot. | | | | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% |

**Figure 5.2** A high-level overview of the 7 consumer type group segments. Group 1 are online only customers, group 2 are multi-channel customers, and groups 3 to 7 are in-store only customers. Proportions are by day type, time of day, channel and basket type. The percentages are to be read by column.

### 5.4.1 Transaction frequency

The k-means clustering algorithm incorporated both the total number of transactions that took place by each customer and the distances they travelled (distance from registered postcode to store location) for each transaction. These were further broken down into day type, time of day, channel and basket type. As shown in Figure 5.3, customers in differing groups transact in varying amounts. Transactions that took place by online customers are less frequent, mainly due to the reasons discussed in section 2.2; on average, customers purchased online groceries once per week, or every 8.6 days, with some customers performing them less often. Customers in consumer type group 2 are multi-channel shoppers, and on average, transacted online less than once a week and purchased in-store once or twice a week. These customers had an average of 4.8 days between each transaction. Those in consumer type group 5 had transacted the most frequently, between 3 to 5 times a week on average. Consumer type group 3 represents 67.56% in the dataset. Due to the high proportion of customers, the average days between transactions is skewed, and is difficult to interpret on an individual level. On average, these customers transacted every 5.8 days. Despite transacting the most frequently, consumer type group 5 only represents 3.45% of all customers, highlighting a small cluster of Sainsbury's customers that frequently transact with the retailer. Those in groups 4 and 6 transacted every 4.1 days, and those in group 7 transacted the most days often at an average of every 3 days.

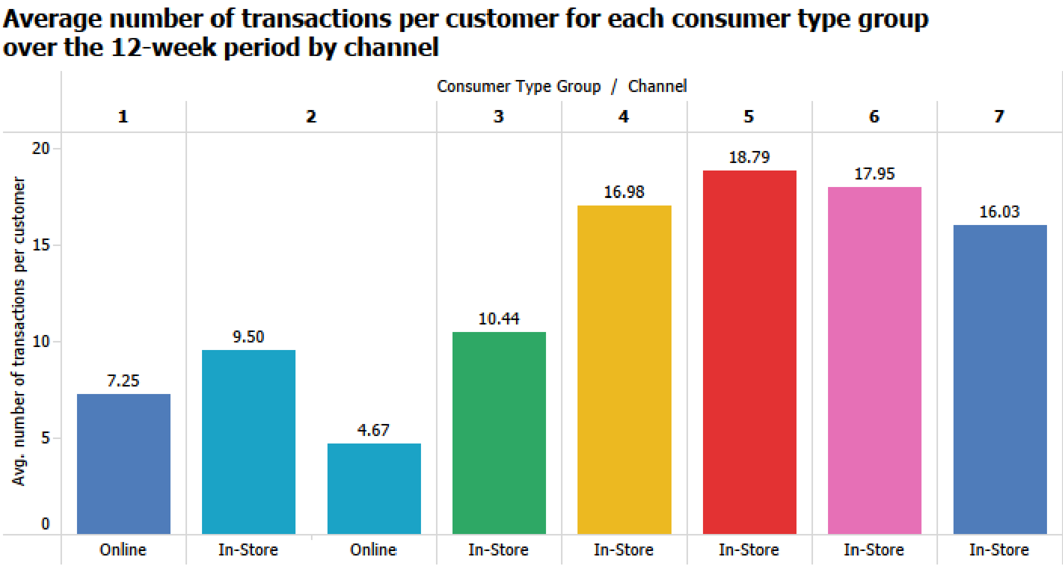



**Figure 5.3** Average number of transactions per customer within each consumer type group over the 12-week period split by shopping channel.

## 5.4.2 Transaction distances

For context regarding the distances travelled between customers' homes and the stores they visited, Figure 5.4 provides a summary broken down into basket types. In all instances, customers in consumer type group 7 had travelled the furthest for their groceries, at least 7 kilometres for their 'for now' baskets. On average, these customers travelled 10.48km for 'main' purchases. These customers travel far from home and to reside in more rural areas. Customers in group 1 did not travel to receive their online groceries themselves, but their assigned stores were around 7km from their homes. Ideally, the 'for now' basket should not be in the dataset as only one transaction of that type took place in the whole 1.2 million transactions; this may be an error in the dataset. For all other in-store customers, most transacted around 3km away from home for most basket types. Our general understanding of grocery purchasing behaviour is that customers will swipe their loyalty card at the stores they most frequently purchase from (Waddington et al., 2018). These transactions are often close to an area where a customer often visits, such as homes, workplace, or points of interest. For most in-store consumer type groups in this data, customers transact locally for ‘top-up’ baskets. Larger ‘main’ baskets are generally purchased further away from the customers’ homes; this could be due to customers living further away from their favourite grocery retailer, or that their closest supermarket happens to be a Sainsbury’s, something we cannot derive from the data. Unique behaviours are captured though, such as those in consumer type group 7 who travel the furthest. These customers transact the most frequent (Figure 5.3) yet travel the furthest away( Figure 5.4). These customers are highly likely to be visiting a particular location each week due to routines associated with work, education, leisure or other habitual behaviour, and purchase from a particular Sainsbury’s store close to that location, which may not be the closest store to home.

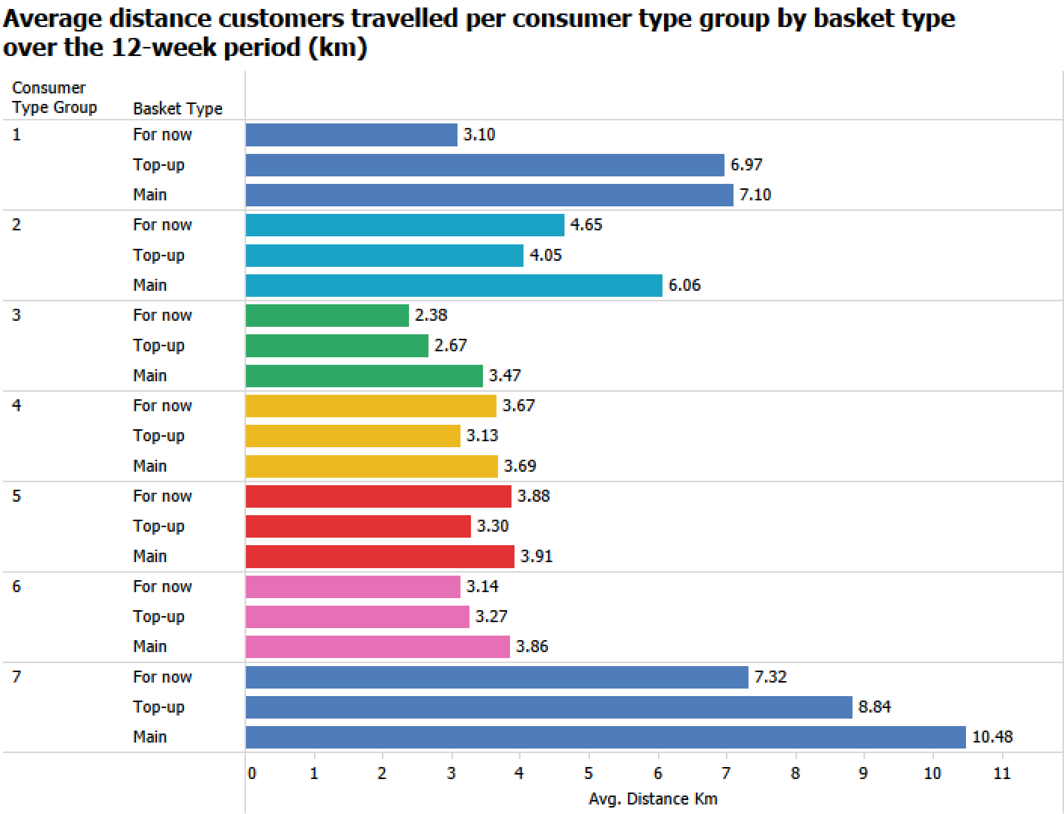



**Figure 5.4** Average distance travelled by basket type for each consumer type group over the 12-week period (km).

### 5.4.3 Transaction scenario proportions

Table 5.1 provides a breakdown of the combined variables by consumer type group for a micro-view of the consumer behaviours and how they differ. As discussed in section 4.2.1, insight into online consumer behaviours is limited due to how Sainsbury's records online transactions. Analysing all groups, weekday transactions during the afternoon in-store tend to be the most common across most consumer type groups besides groups 4 and 5. Those in group 4 predominantly transact for 'main' baskets during a weekday evening or for 'top-up' shops during a weekend afternoon. For those in consumer type group 5, over one third of their transactions occurred on a weekend afternoon for 'main' basket types. These two consumer segments potentially use Sainsbury's as their key grocery retailer to purchase the bulk of their groceries. Those in group 6 primarily perform ‘main’ transactions during weekday mornings and afternoons, not fitting into the proportion of customers who transact outside of traditional work week hours. Consumer type group 3 contains most of the observed customers in the loyalty card-linked transaction dataset and has the widest variety of transactions across all in-store transaction scenarios. Although overall, customers in this group tend to seldom purchase their ‘main’ baskets from Sainsbury’s during the evenings, this will be explored further in the following sections. The final consumer type group, 7, experienced the most unique behaviours compared to other groups. These customers predominantly transacted for ‘for now’ transactions during weekday afternoons and evenings. They also transacted frequently for ‘for now’ transactions during the weekend afternoon. These observations, combined with distance data, should support in identifying unique group behaviours that are expected of grocery consumers. For example, it is expected that those in consumer type group 7 are commuters who purchase at Sainsbury’s stores near their place of work. Though data on customer workplace locations are not available, inferences can be made.

**Table 5.1** Consumer type group breakdown by day type, time of day, channel and basket type. The percentages represent the proportion of total transactions for each consumer type group, reflecting which transaction scenario was the most common.

| Transaction Scenario by Consumer Type Group | | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|---|---|
| Weekday | Morning | Online | For Now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Morning | Online | Top-Up | 0.8% | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Morning | Online | Main | 49.0% | 15.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Afternoon | Online | For Now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Afternoon | Online | Top-Up | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Afternoon | Online | Main | 2.1% | 1.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Evening | Online | Main | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Morning | Online | For Now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Morning | Online | Top-Up | 0.5% | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Morning | Online | Main | 45.2% | 15.8% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Afternoon | Online | For Now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Afternoon | Online | Top-Up | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Afternoon | Online | Main | 2.3% | 0.9% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Evening | Online | Top-Up | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekend | Evening | Online | Main | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| Weekday | Morning | Instore | For Now | 0.0% | 4.1% | 6.4% | 2.3% | 2.3% | 4.3% | 15.1% |
| Weekday | Morning | Instore | Top-Up | 0.0% | 3.4% | 6.1% | 2.3% | 2.3% | 9.6% | 3.8% |
| Weekday | Morning | Instore | Main | 0.0% | 2.1% | 1.2% | 1.3% | 1.2% | 20.0% | 0.7% |
| Weekday | Afternoon | Instore | For Now | 0.0% | 7.6% | 12.5% | 5.3% | 5.1% | 5.0% | 21.7% |
| Weekday | Afternoon | Instore | Top-Up | 0.0% | 8.0% | 13.4% | 8.7% | 9.0% | 10.1% | 9.6% |
| Weekday | Afternoon | Instore | Main | 0.0% | 4.0% | 3.2% | 7.0% | 6.0% | 18.8% | 2.1% |
| Weekday | Evening | Instore | For Now | 0.0% | 2.4% | 4.8% | 4.3% | 1.9% | 0.8% | 5.3% |
| Weekday | Evening | Instore | Top-Up | 0.0% | 2.5% | 4.2% | 11.0% | 3.2% | 1.0% | 2.7% |
| Weekday | Evening | Instore | Main | 0.0% | 1.2% | 0.5% | 15.1% | 1.3% | 0.7% | 0.4% |
| Weekend | Morning | Instore | For Now | 0.0% | 2.7% | 5.6% | 2.1% | 1.5% | 3.5% | 7.3% |
| Weekend | Morning | Instore | Top-Up | 0.0% | 3.6% | 6.4% | 3.4% | 3.0% | 5.4% | 3.3% |
| Weekend | Morning | Instore | Main | 0.0% | 3.1% | 3.4% | 4.3% | 6.8% | 6.0% | 1.5% |
| Weekend | Afternoon | Instore | For Now | 0.0% | 5.2% | 10.2% | 5.1% | 4.5% | 3.3% | 12.1% |
| Weekend | Afternoon | Instore | Top-Up | 0.0% | 7.9% | 12.9% | 11.3% | 12.4% | 5.8% | 7.8% |
| Weekend | Afternoon | Instore | Main | 0.0% | 5.7% | 3.4% | 8.6% | 36.0% | 4.5% | 2.2% |
| Weekend | Evening | Instore | For Now | 0.0% | 1.3% | 3.0% | 1.6% | 0.7% | 0.4% | 2.8% |
| Weekend | Evening | Instore | Top-Up | 0.0% | 1.2% | 2.4% | 3.4% | 1.2% | 0.6% | 1.5% |
| Weekend | Evening | Instore | Main | 0.0% | 0.5% | 0.5% | 2.7% | 1.6% | 0.3% | 0.3% |

### 5.4.4 Demographic breakdown of consumer type groups

The group-level OAC dataset was used to explore the demographic relationships between customer segments (Table 5.2). From this, we can determine if certain demographic groups have diverse types of shopping behaviour as reported in our consumer type groups which are explored further in the next section. The most prominent finding of Table 5.2 is that Sainsbury's customers in West Yorkshire are primarily customers who belong to groups in OAC 5 and 6, which are households located in urban and suburban areas that are considered mixed in affluence, reflecting the general branding of the supermarket and its typical store locations in this region. Slight demographic variation is found in consumer type group 7 compared to other groups; some in-store customers are also located in 'rural areas' (group 1b) and in 'industrious communities' (8a). The OAC data presented here will support the writing of pen portraits of the consumer type groups in the following section: 5.4.5.

**Table 5.2** Proportion of consumer type groups by ONS Output Area Classification. Data Source: (Office for National Statistics, 2018).

| Output Area Classification Group Code | Output Area Classification Group Labels | Consumer Type Group | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1a | Farming Communities | 1% | 1% | 0% | 0% | 0% | 0% | 1% |
| 1b | Rural Tenants | 4% | 5% | 2% | 4% | 4% | 5% | 7% |
| 1c | Ageing Rural Dwellers | 1% | 0% | 0% | 0% | 1% | 1% | 1% |
| 2a | Students Around Campus | 2% | 2% | 2% | 1% | 1% | 0% | 1% |
| 2b | Inner-City Students | 1% | 1% | 1% | 0% | 0% | 0% | 0% |
| 2c | Comfortable Cosmopolitans | 1% | 1% | 1% | 1% | 1% | 0% | 1% |
| 2d | Aspiring and Affluent | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 3a | Ethnic Family Life | 0% | 0% | 1% | 0% | 0% | 0% | 0% |
| 3b | Endeavouring Ethnic Mix | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 3c | Ethnic Dynamics | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 3d | Aspirational Techies | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 4a | Rented Family Living | 4% | 4% | 6% | 5% | 4% | 3% | 3% |
| 4b | Challenged Asian Terraces | 2% | 2% | 4% | 3% | 3% | 2% | 3% |
| 4c | Asian Traits | 2% | 3% | 3% | 3% | 3% | 2% | 2% |
| 5a | Urban Professionals and Families | 18% | 20% | 18% | 19% | 20% | 16% | 19% |
| 5b | Ageing Urban Living | 11% | 10% | 9% | 10% | 10% | 11% | 8% |
| 6a | Suburban Achievers | 11% | 11% | 9% | 11% | 12% | 15% | 7% |
| 6b | Semi-Detached Suburbia | 23% | 25% | 22% | 24% | 26% | 29% | 23% |
| 7a | Challenged Diversity | 3% | 2% | 3% | 3% | 2% | 2% | 2% |
| 7b | Constrained Flat Dwellers | 0% | 0% | 0% | 0% | 0% | 0% | 0% |
| 7c | White Communities | 1% | 1% | 1% | 1% | 1% | 1% | 1% |
| 7d | Ageing City Dwellers | 1% | 1% | 1% | 1% | 0% | 1% | 1% |
| 8a | Industrious Communities | 6% | 5% | 5% | 4% | 5% | 4% | 8% |
| 8b | Challenged Terraced Workers | 2% | 2% | 2% | 2% | 2% | 1% | 2% |
| 8c | Hard-Pressed Ageing Workers | 3% | 3% | 4% | 3% | 3% | 3% | 4% |
| 8d | Migration and Churn | 3% | 3% | 4% | 3% | 3% | 2% | 4% |

### 5.4.5 Spatial context of grocery consumers

Figure 5.5 provides an insight into the locations where Sainsbury's customers resided (registered address for their loyalty card) during the 12 weeks for which Nectar card data have been collected. Ideally, total count numbers would not be used to create a choropleth map; however, the purpose of this map is to present the general geographical overview in which the customers within the loyalty card dataset lived. In conjunction with mapped store locations, the map indicates that more customers tend to be clustered around Sainsbury's Supermarket stores (Figure 5.5). This is logical and coincides with other studies, such as Waddington et al. (2018) found that loyalty card customers tend to shop at their most local stores. Stores within the city centres of Bradford and Leeds seem not to have as many customers nearby; this is potentially due to multiple factors, such as the cities containing many competitors in the area and not as many people residing in city centres.

On the other hand, Bradford does not have many Sainsbury's supermarkets in its city centre, and only Sainsbury's Locals (convenience stores), which provide a smaller selection of products. From these data, Sainsbury's customers observed within the dataset appear to reside near the larger supermarkets and further away from the Sainsbury's Locals. At this point, transactions that take place at Sainsbury's Local stores in city and town centres are expected to be performed by customers who live far away and transact during the working week. Data regarding where the observed customers work is unavailable; only logical assumptions can be made based on the evidence presented.

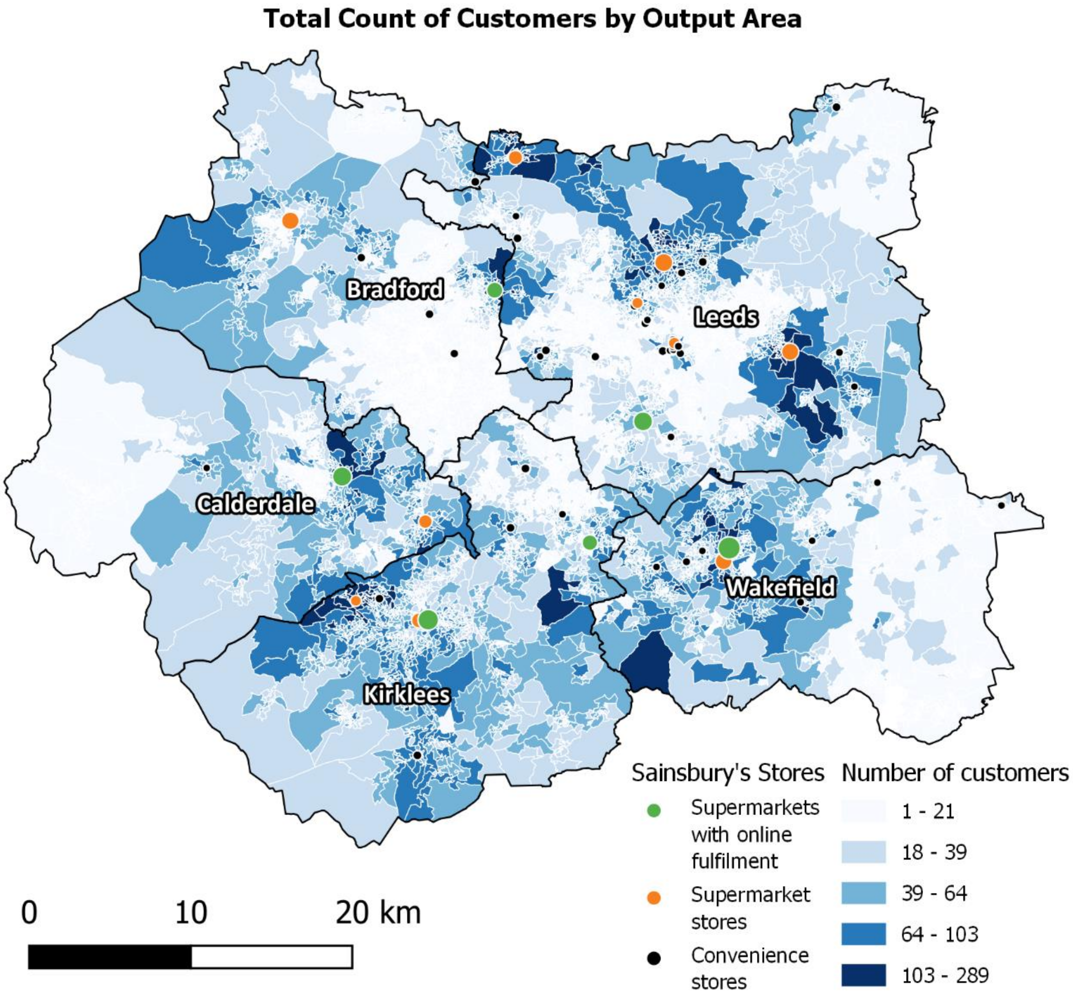



**Figure** 5.5 Count of total number of customers in the loyalty card-linked transactional dataset provided by Sainsbury's using their Nectar loyalty card scheme. Sainsbury's store points sizes vary based on floorspace, with larger symbols indicated a higher sales area.

The following sequence of figures and paragraphs goes through each unique consumer group, providing summaries about these groups, including their geographical locations, transaction scenarios, and links to their likely demographic supergroups. The subsequent series of maps show the percentage of Sainsbury's customers in the dataset that belong to each consumer type group by OA. Some OAs have a very low count of customers within our dataset, and so the maps presented in section 5.5 should be interpreted with reference to figure 5.5.

## 5.5 Pen portraits of consumer type groups

When creating geodemographic classifications, a standard approach is to interpret each unique consumer type group and provide a label related to the observations. Each consumer type group is presented in further detail in this section and is provided with geographical context. Radial charts are often created with classifications to visualise each variable's average values. However, as the research presented here contains numerous variables, these charts are notoriously difficult to create and interpret. Therefore, the following sections provide a descriptive analysis of each consumer type group regarding each variable used in the clustering process, the distances customers travelled, the socio-demographic characteristics of the areas they reside within, and maps to visualise their presence in West Yorkshire.

### 5.5.1 Consumer type group 1: Weekly Clickers

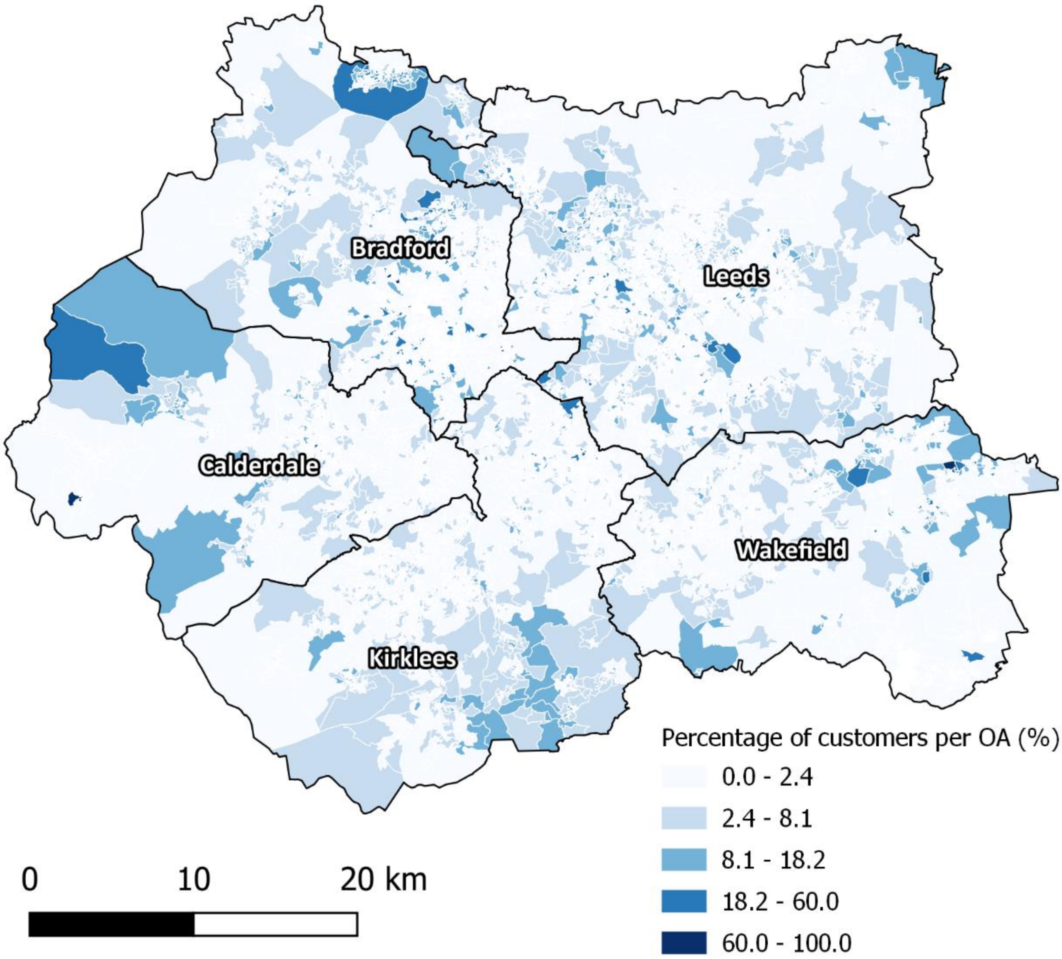



**Figure 5.6** Percent of Consumer Type Group 1 customers by Output Area.

This consumer type group represents only 2% of all customers, who, on average, transact once at a week, with 99% of their transactions categorised as a ‘main’ shop (Figure 5.2). Unfortunately, little insight is provided regarding the temporality of these consumers’ behaviours due to online transactions being assigned to the morning period by Sainsbury’s. However, this data can be helpful for Sainsbury’s when assigning customer deliveries from their store network. As shown in Figure 5.6 and Table 5.2, these customers largely reside within the suburbs, with 4% living in more rural areas. These customers tend to reside far away from Sainsbury’s stores, where delivery drivers travel around 7km between the customer's home and the store that prepared their delivery. The analysis of Sainsbury's delivery network is beyond the scope of this thesis; however, these data can provide insight for those pieces of research. For more context, Figure 5.7 presents an overview of the interactions between online fulfilment supermarkets and online customers. Each yellow dot represents locations where customers live at the OA level. The mini maps show that online orders tend to be filled by the closest store that delivers them. However, some stores, such as Halifax's supermarket and Shorehead, at some point delivered to customers in all 5-districts of West Yorkshire. This could be due to the lack of delivery drivers on a specific day or a fault with one of the more local stores, therefore, another store within the same 'appraisal territory' (a Sainsbury's grouping of proximate stores to share online order fulfilment capacity) is fulfilling the online order. Studying this observation more thoroughly can support those researching Sainsbury's online delivery network to understand better why some stores deliver across multiple districts. Relating back to the LCFS data (Table 4.2), analysis found that these customers, on average as a group, met the expected weekly expenditure 6 times out of the 12 weeks. The high sales value of online transactions means these customers are most likely to meet the expected spending on food and beverages each time they transact online. These customers are most likely purchasing the bulk of their groceries from Sainsbury's, notably those who transact approximately once per week. Those who are purchasing their 'main' baskets from Sainsbury's every 2 weeks or more at the LCFS expected value (Table 4.2) may be purchasing food items from other supermarket chains in-store. Without data evidencing this though, only logical assumptions can be made.

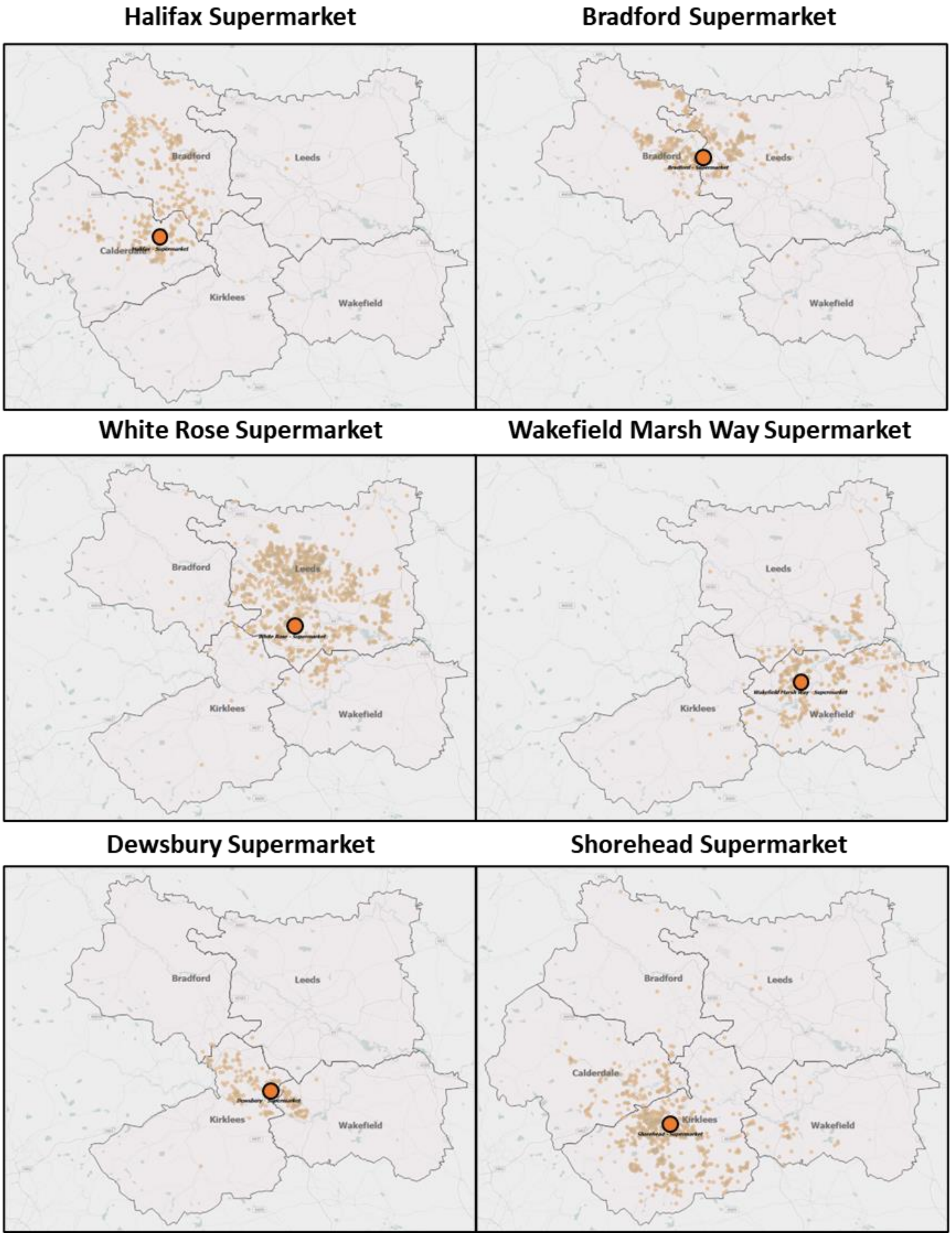



**Figure 5.7** Consumer type group 1 (online customer) interactions between customer OAs and individual stores. Orange dots represent supermarkets, yellow dots are customer home locations.

### 5.5.2 Consumer type group 2: Clicks and Mortar

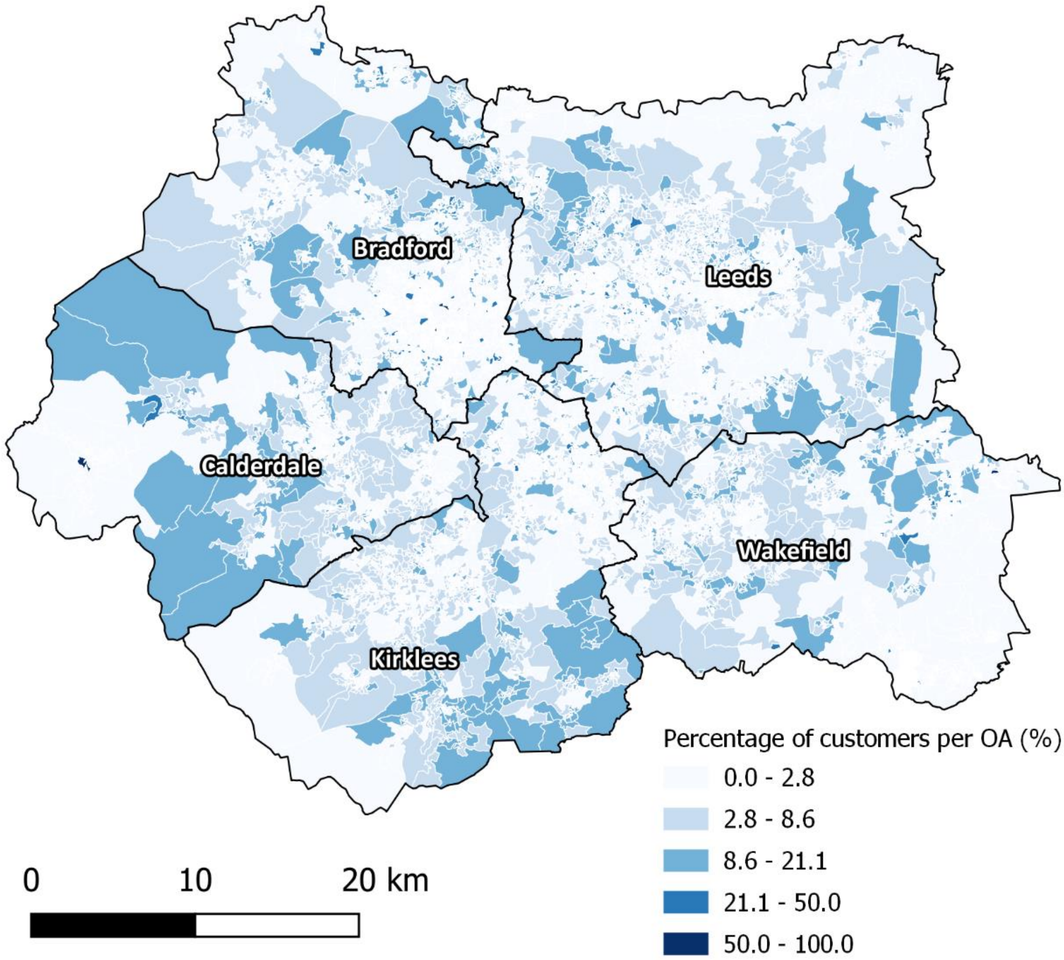



**Figure 5.8** Percent of Consumer Type Group 2 customers by Output Area.

Representing 4% of customers, these customers, on average, transacted 2 – 5 times a week, with 67% of transactions taking place in-store (Figure 5.3). At least one weekly transaction was for a ‘main’ online basket, and the others were split between ‘for now’ and ‘top-up’ baskets in-store. These customers performed most in-store transactions within 4.3km of their home’s OA and did most in-store transactions during a weekday afternoon (Figure 5.2). Customers within this consumer type group are most likely to live in suburban areas or be urban professionals and have families, according to the OAC (Figure 5.2). These customers are residents across all districts in West Yorkshire and are the only group to purchase from Sainsbury's both in-store and online. These customers are likely to be 'loyal' due to their frequent purchasing behaviour at the retailer, whether due to their brand preferences or distance from competitors. On average at the group-level, these customers met the weekly LCFS expected expenditure 4.2 weeks out of 12. These customers purchased their online transactions almost every week but did not follow a clear pattern compared to those in consumer type group 1. These customers are purchasing mixed baskets from Sainsbury's both online and in-store but are most likely spending at other retailers too as they don't always meet the weekly expected expenditure.

### 5.5.3 Consumer type group 3: Convenience Suburbanites

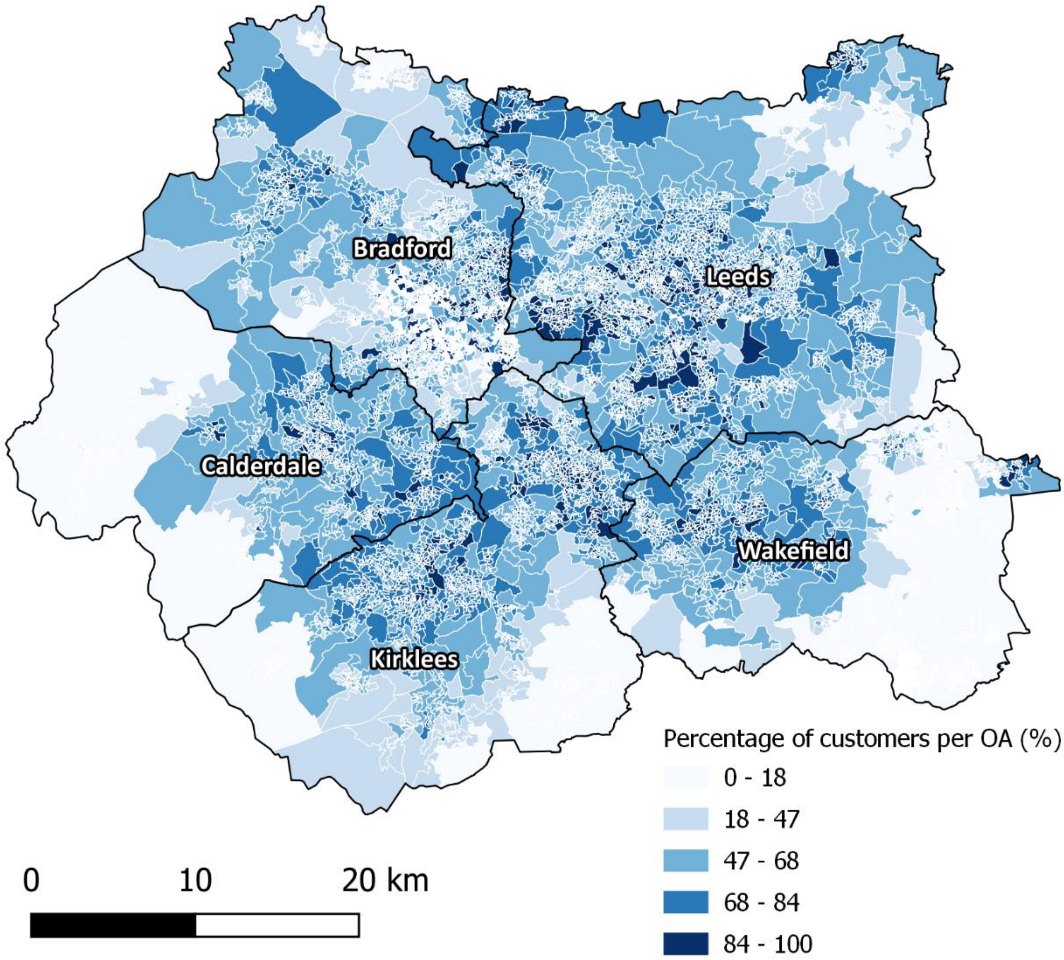



**Figure 5.9** Percent of Consumer Type Group 3 customers by Output Area.

This consumer type group represents the larger number of Sainsbury's customers in the dataset at around 68%. With customers across all districts, these customers cover the widest variety of geodemographic areas, except any areas classified as rural (Table 5.2). Due to the large number of customers in this group, it is unsurprising that a higher proportion of customers live in less affluent areas labelled as 'rented family living' or 'challenged Asian terraces'. This consumer type group is almost a catch-all group that identifies all customers who transact at a Sainsbury's store for smaller baskets only. Around 17-17% of all transactions by these customers took place on a weekday afternoon in-store for ‘top up’ and ‘for now’ basket types, and on average, travelled 2.5km from home to transact. These customers also transacted for similar basket types on weekends and rarely purchased ‘main’ basket transactions. These customers are most likely not ‘loyal’ Sainsbury’s customers and prefer a different retailer for the bulk of their grocery purchases. For these customers, Sainsbury’s stores are more likely a place of transacting due to locality convenience and they would rather not travel further afield to a different store. In terms of the LCFS, this group did not meet the expected LCFS spend during any of the study weeks at Sainsbury’s (Table 4.2). This is not surprising as we infer that these customers are shopping at their local Sainsbury’s stores for food out of convenience as opposed to being their primary retailer for groceries.

### 5.5.4 Consumer type group 4: Small Town Shoppers

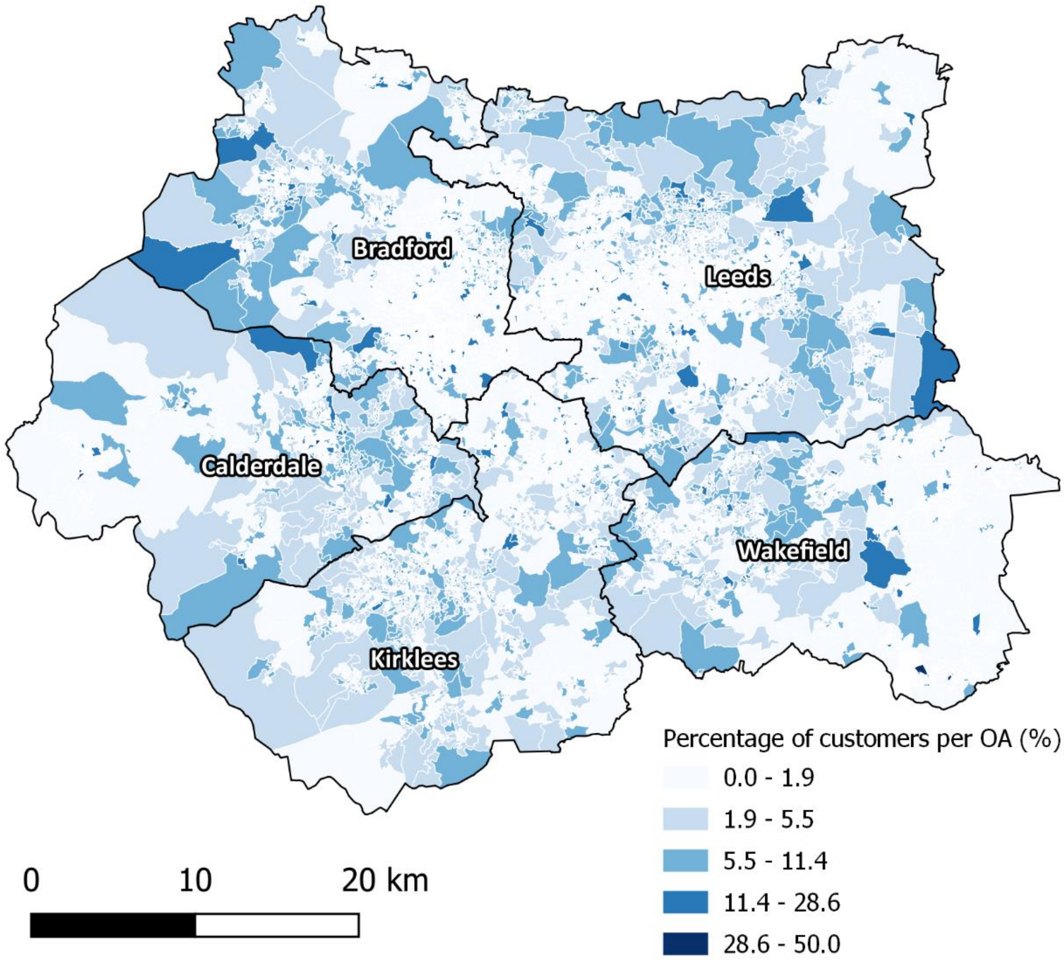



**Figure 5.10** Percent of Consumer Type Group 4 customers by Output Area.

Around 4% of Sainsbury's customers belong to consumer type group 4. These consumers frequently transact at a Sainsbury's store multiple times a week and travel between 2 to 3.5km on average from their homes. Transacting during the morning is rare for these individuals, and they predominantly purchase 'top-up' and 'main' baskets on weekday evenings and weekend afternoons. Around 50% of customers transacted at a store in a residential area, followed by 47% in town centres. These customers transacted within various areas where stores are located, including retail parks. These customers shop locally around home and town centres but travel further for weekend transactions. Customers in this group met the LCFS expected expenditure 4 weeks out of 12 on average. Considering the mixed types of transaction baskets performed by these customers, it was expected that they would meet the weekly expenditure more often. However, many customers in this group do not purchase main baskets on a habitual basis, therefore do not meet the weekly LCFS expected expenditure, bringing down the group's average. Perhaps the customers in this consumer type group could be broken down further into those who purchase 'main' baskets more regularly. Although, these customers make up such a small number of cardholders in the dataset that further segmentation may not bring much value at the macro-scale.

### 5.5.5 Consumer Type Group 5: Supermarket Weekend Warriors

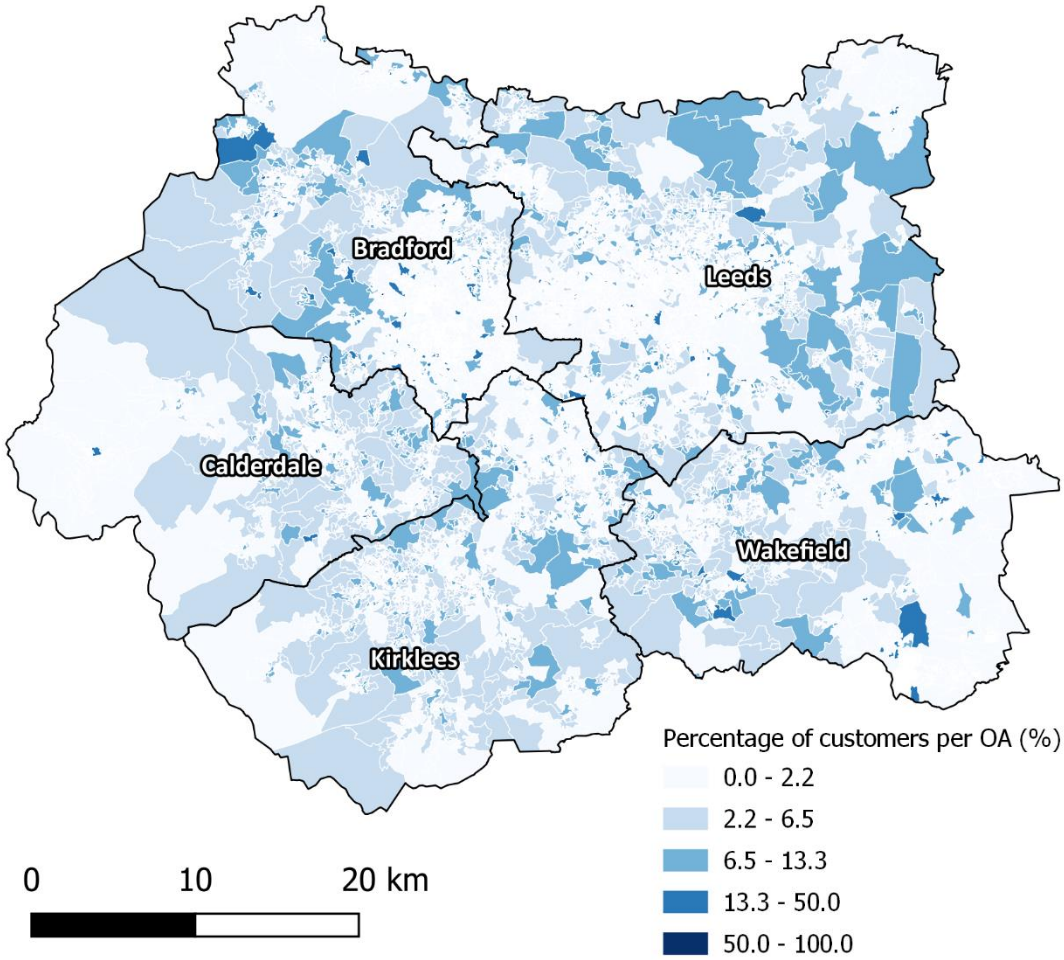



**Figure 5.11** Percent of Consumer Type Group 5 customers by Output Area.

Consumer type group 5 are unique, as these customers tend to favour a particular type of transaction. On average, over the 12 weeks, around 36% of all transactions by these customers took place on a weekend afternoon for a 'main' basket. Customers in this group transacted around twice a week, once for a 'main' purchase and the other times for an afternoon 'top-up' on weekends or weekdays. These customers met the weekly LCFS expected expenditure 5 weeks out of 12, mostly due to 'main' basket purchases. These customers were the most frequent to visit a

Sainsbury's in-store store overall and are most likely shopping at the retailer for their 'main' weekly grocery purchases.

### 5.5.6 Consumer Type Group 6: If I need it, Sainsbury's has it!

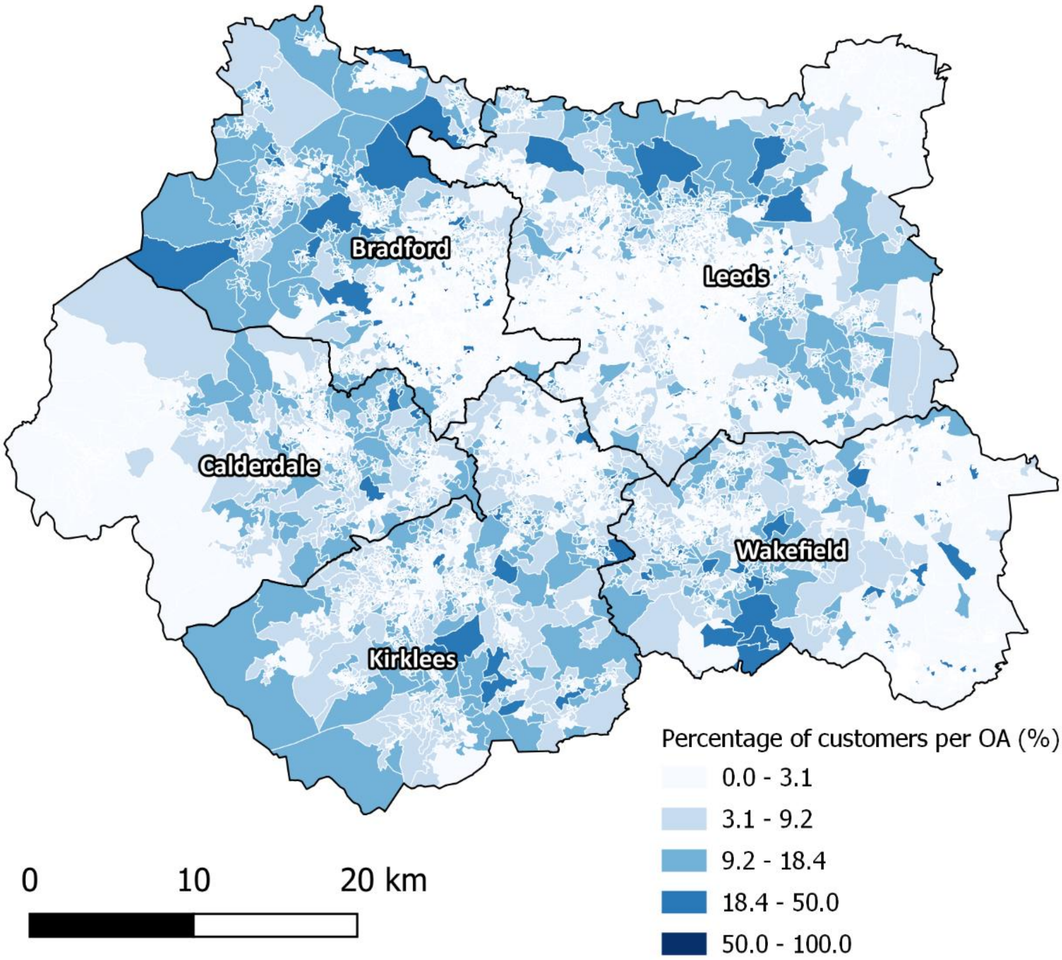



**Figure 5.12** Percent of Consumer Type Group 6 customers by Output Area.

Consumer type group 6 frequently purchase their 'main' baskets from Sainsbury's, especially during a weekday afternoon or morning, but do not transact during the evening or on the weekend. This group has more customers residing in Wakefield than all other groups. It has the highest proportion of customers living in the OAC group semi-detached suburbia and suburban achievers. It is difficult to explain why these customers purchase groceries from Sainsbury's for such specific baskets at certain times. They do not travel further than 3.86km on average and purchase their 'main' food shops from Sainsbury's. Interestingly, these customers do not transact on weekends. Due to their high proportion of 'main' basket purchases, these customers met the LCFS weekly expected expenditure amount for 6 weeks on average. These are high-value customers for Sainsbury's, as their weekly transactions indicate they are purchasing most of their groceries from them.

### 5.5.7 Consumer Type Group 7: Hungry and Far From Home

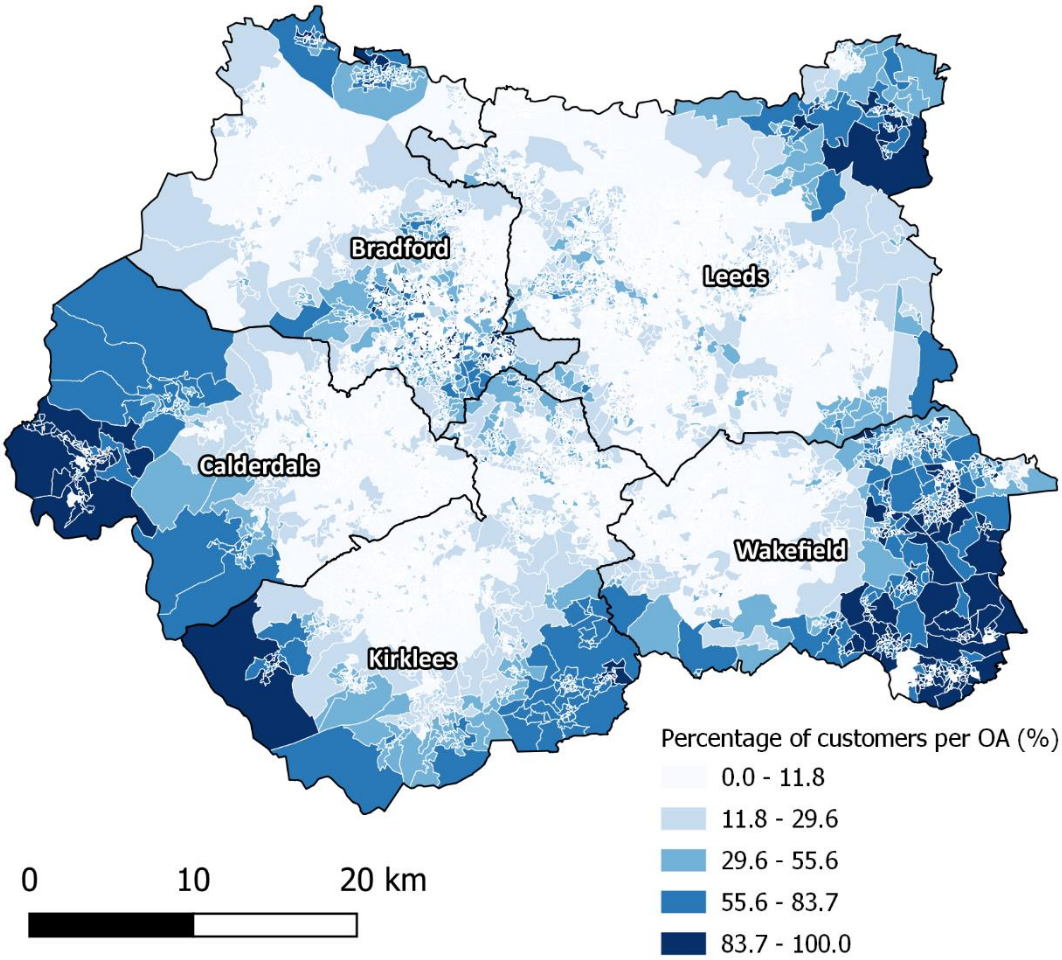



**Figure 5.13** Percent of Consumer Type Group 7 customers by Output Area.

The 7th and final consumer type group is an incredibly distinct group of Sainsbury's customers, representing 13% of all customers in the dataset. Around 66% of the transactions these customers perform are classified as ‘for now’ transactions, indicating the purchase of ‘on the go’ meals such as meal deals, snacks, and lower-value purchases. This is followed by ‘top-up’ transactions at 27%, including food purchases that will likely be eaten later or the next day. This group’s average distance between home and stores is the furthest, at around 7.3km to 10.5km (Figure 5.4). As this group is unique, the stores they visited were analysed and compared to other consumer type groups. One key store was identified, which experienced 58% of its transactions from consumer type group 7, and that was the Leeds Station Local store highlighted in Figure 4.8. Considering that these customers are predominantly located along the outer edges of the districts of West Yorkshire (Figure 5.2) and are most likely workers who purchase their lunches within the town and city centres, it is logical to assume that this consumer type group represents the commuting group of Sainsbury’s customers. Whether these customers work in Leeds city centre is unknown. However, as they are transacting at a train station, these transactions are opportunistic and likely part of a longer journey the customer is making. These customers are an integral find, backing the literature of multi-purpose shopping transactions and workplace food purchasing (Berry et al., 2016). These customers did not meet the weekly LCFS once during the 3 months, largely due to their low value yet frequent purchases. Customers within this segment are most likely transacting at other retailers for their groceries, and purchase from Sainsbury’s due to the store’s location to their destination.

## 5.6 Chapter 5 summary

This chapter has guided the reader through the clustering process performed in this thesis to derive consumer type groups. The k-means clustering algorithm successfully identified seven unique consumer behaviour typologies regarding the consumer’s temporal and spatial choices. These customers vary regarding how often, where, and when they transact their shopping baskets and the stores they transact at. The addition of geodemographics provides an idea of the types of Sainsbury's customers that reside within West Yorkshire based on the population they live within and how they differ based on shopping behaviours. This analysis presents how individualised behaviours can be aggregated to form groups of grocery shoppers who vary in terms of where they live, when they purchase their groceries, what channels they use, when they transact, and these relationships to basket type. From the county-level perspective, 41.2% of customers of this retailer in West Yorkshire fall within the 'urban professionals and families' and 'semi-detached suburbia' groups within the ONS' geodemographic Output Area Classification (Table 5.2). This indicates that this retailer mainly attracts those who reside within suburban areas of the UK, who are most likely employed and have families, following similar research output in Clark et al. (2021). Other demographic groups do transact at this retailer, but less often. The segments identified in this study vary demographically, implying that demographical features such as affluence, accessibility, employment, and rurality play a key role in consumer behaviour depending on the brand being considered (Office for National Statistics, 2020b; Hood et al., 2020).

Further work investigated each shopping indicator, comparing consumer type groups to the county level. It was found that customers in West Yorkshire shopped at the retailer relatively locally at 2 to 4km, visited between 1 to 2 stores, and transacted every 8.4 days. Out of all groups, customers in consumer type group 7 (Hungry and Far From Home) had travelled the furthest between store and home (by 10km+) compared to the county average and visited a wider variety of stores. Consumer type group 5 customers visited the most frequently at the county level. These customers mainly visit local supermarkets and are spread evenly throughout the county, residing primarily in suburban family areas (Table 5.2). The ability to seek out and identify these unique consumer segments provides retailers with a deeper insight into the purchasing actions of their customers, linking transaction frequency, location, distance, and basket type with temporality.

The findings regarding the online and multi-channel shoppers, identified as consumer type groups 1 ‘Weekly Clickers’ and 2 ‘Clicks and Mortar’, support the confirmation of the hypothesis that those who live more rurally are likely to use online for grocery purchase as discussed in Clarke et al. (2015). These customers who live away from towns and suburbs are most likely to reside on the outskirts of district boundaries and make use of the online services available. Understanding the spatial demand for online helps retailers plan their store network and adjust their stores to provide better access to delivery. Following the findings in Kirby-Hawkins et al. (2018) and Waddington et al. (2018), non-Sainsbury’s customers in these rural areas are likely to be using online grocery services provided by competitor retailers. Modelling the individualised behaviours of online customers can allow grocers to identify their most successful areas of demand and adjust their store network accordingly when competing with others.

Customers in consumer type group 7 ‘Hungry and Far From Home’ are a unique group that represent those who transact at Sainsbury’s stores far away from home. Their distinct behaviour is likely linked to workplace location as their transactions are frequent and at stores close to city centres and retail parks. Research in Waddington et al. (2018) identified a specific store type that attracts workplace demand, which could possibly link the customers identified in consumer type group 7 to those stores in cluster group 2 in their work. Further work could combine the consumer clusters identified here, and the clustered stores presented in Waddington et al. (2018). This would allow for a more sophisticated approach to modelling store allocation in the IBM presented here. Customers within consumer type group 7 could have behavioural rules that ‘pull’ them towards workplace demand stores.

This study has successfully segmented customers into groups based on similarity, but it has limitations. Notably, being the coverage of transaction data, firstly, not all customers that own a loyalty card will swipe for every transaction, notably small baskets (Wright and Sparks, 1999). Secondly, some customers do transact at the grocery retailer who do not own a loyalty card, yet still generate a large amount of revenue for the retailer. The traceability of who is performing the transaction is unknown; therefore, knowledge cannot be derived from their behaviours. The other key drawback of the analysis is using the ONS 2011 Census and Output Area classification, as these data are seven years older than the transaction data. During analysis, there were significant discrepancies between the number of cardholders in various output areas and the number of households. This indicates that some output areas have experienced significant population increases over the years, as van Dijk et al. (2021) found. Finally, a key caveat of the dataset is the way in which online transactions are recorded. Despite the loyalty card data being rich, providing insight into spatiality and temporality of consumer behaviour, there is limited insight into the true behaviours and choices of online shoppers. If Sainsbury’s captured the time the online groceries were to be delivered, they would have a much richer insight into the lives of their customers, identifying whether they are likely to work from home, or not work at all, or be able to spot patterns in delivery times. Considering the unprecedented uptake in online grocery shopping during the Covid-19 pandemic (Pantano et al., 2020; Meister et al., 2023), understanding these behaviours may have helped better manage their responses.

This research has explored a rarely accessed dataset, allowing for further developments in areas of grocery retail geography to understand consumer behaviours. The consumer type groups identified are used to design the IBM developed in this thesis. However, some nuanced behaviours, such as journey-based transactions that occur at locations such as Leeds train station, are anticipated to be challenging when modelling store choice due to limited insight regarding where customers relocate at different time points of the day.

It is duly recognised that while distinctive customer segments have been meticulously identified based on their transactional behaviours at Sainsbury's stores, a notable gap persists concerning their behaviours outside of Sainsbury's. This absence of external behavioural insights is imperative for a comprehensive understanding of consumer store and channel preferences, particularly considering competition. Access to data providing insights into customer activities at competitor retailers would substantially increase the robustness of the model designed in this study, extending its applicability beyond Sainsbury's. Nonetheless, the study collaborator required an initial modelling framework predicated on their observed customer transactions within Sainsbury's store network. This approach facilitates a deeper understanding of how *their* customers behave and their relationships with *their* store network. Incorporating other brands would further enhance this research, which is elaborated upon in Chapter 7.

This chapter has successfully achieved aim 3 by segmenting the observed customers in the transaction dataset into consumer type groups based on their key behaviours regarding store location, channel, and basket purchase choices. This novel research provides insight into the spatial and temporal store choices of consumers and can be used to support retailers to better understand who their customers are, and how they behave.

# Chapter 6 Building an individual-based model of consumer store and channel choice behaviours

The core aim of this chapter is to design an IBM founded upon the observations and real data analysis in Chapter 5, incorporating temporality and spatiality using an IBM methodology. To achieve this aim, this chapter explicitly attains objectives 5 and 6:

*To perform data mining on the customer segments, identifying the probability of making a transaction at any given time based on their linked loyalty card transactions.*

*To design and build a prototype individual-based model of consumer behaviours suitable for simulating their transactions by day type, time of day, channel, and basket type.*

To accomplish the core aim, an individual-based modelling framework is presented and applied for the purpose of simulating the store and channel choice behaviours of customers at Sainsbury's grocery stores across the study area of West Yorkshire. To build the IBM, a hybrid methodological approach was used, incorporating elements from both MSM and ABM; this is further explained in section 6.1.2.

The following sections focus on the steps taken during the model-building process. Section 6.1.2 summarises the methodology and tools used to build the model and data requirements. Section 6.1.3 discusses the model Sturley et al. (2018) created and how their work has guided the design and enhancement of the model built in this thesis. Section 6.1.4 discusses the process of mining the behavioural rules, i.e., probabilities of customers making transactions relating to day type, time of day, basket size, and channel choice derived from the different customer types identified in section 5.5. This section also provides an in-depth overview of the decision tree developed to model the temporal probabilities of consumer transactions. Section 6.2 moves onto the process of incorporating consumer store choice behaviour, focusing primarily on transaction spatiality. This part of the chapter takes the reader through three store choice selection methods attempted in the model building process, with each adding more complexity, following the KISS principle. Finally, Section 6.3.6 presents an example scenario test of how the model could be used by retailers in the future once a sound approach modelling consumer store choice is found.

As with all studies that include the development of an IBM, or any model for that matter, specific criteria and model requirements should be outlined to clearly describe what the model aims to achieve, how it will achieve those aims, and identify what it will and will not desire to attain (Hammond, 2015). Such model design, formulation, and communication are uniquely challenging, and choosing how complex to design the model is notoriously challenging (Grimm and Railsback, 2012). Therefore, before building the individual-based consumer behaviour model, the Overview, Design Concepts and Details (ODD) protocol was followed, as described by Grimm et al. (2010)(Grimm and Railsback, 2012). The ODD protocol is a standard format for communicating and formulating ABMs, following a structure that focuses on model purpose, design, and initialisation.

Prior to delving into the ODD protocol, which describes the model from a modellers perspective, here is a reminder about what is currently missing in the tools used in location analytics, and how the model proposed here offers a solution. As discussed in section 3.2, the top-down methodologies used in location analytics, especially SIMs, provide a sound technique for predicting store sales. These models perform by estimating the demand around a store point, which can account for temporal differences, distance decays, and competition (Heppenstall et al., 2006; Gauri et al., 2008; Newing et al., 2015; Waddington et al., 2019). SIMs in grocery retail contexts are useful for predicting how a new store could perform in a new location, by modelling flows between surrounding geographical areas and the store itself. However, if a location analyst wanted to know *who* in their customer base is shopping at their store, or *how* they will respond to supply changes, these methods do not provide this type of insight (Sturley et al., 2018; Wilkinson, 2023). Analysis of past customers can be performed to identify the who, as presented in Chapter 5, but does not allow for future scenario testing.

Top-down models assume that all customers that generate the predicted store sales behave homogenously, with no unique preferences regarding store location and channel choice. We know that customers have preferences in terms of their grocery purchasing behaviours, whether it is brand choice, willingness to travel, or other aspects (Chapter 2). Customers store choice is inherently linked with their shopping mission (basket type), whether they are transacting at a store out of convenience, product variety, or is a habitual purchase (Wood and Browne, 2007). Therefore, being able to simulate these customer preferences from the individual-level provides researchers and retailers with a powerful tool to better understand their customer base, and scenario test both supply and demand changes. Using an IBM either alone or in conjunction with a SIM model gives retailers and upper hand in knowing their current and potential market, gaining competitive advantage over other retailers. Section  discusses the ways in which an IBM can be used, and the type of scenario tests that can be performed.

### 6.1.1 Overview, Design Concepts and Details Protocol

Despite using an ABM-inspired methodology, the model does not incorporate some of the unique features of ABM, such as agent *adaptiveness* or *complete stochasticity*. While beneficial in some studies, these key features would add more complexity to this model but not more insight and have purposefully not been included. By keeping the model's design and functionality relatively simplistic, the model is effortlessly followed step-by-step and can be easily validated. The model's design could be further developed to include agent adaptivity and stochasticity; however, these developments require rigorous research to be representative of reality.

*Purpose.* The model simulates the spatiotemporal behaviours of Sainsbury's consumers at Sainsbury's stores within OAs in West Yorkshire. The model captures individual transactions undertaken by customer agents who belong to varying consumer type groups as created in section 5.5. The model uses empirical data to calculate the transactional probabilities for any customer within a consumer type group, including individual probabilities regarding day type, time of day, shopping channel, and basket type. Based on empirical data, the model then assigns a store where the customer's transaction will occur.

*Entities, state variables, and scales*. The model entities include static agents representing individual customers, placed at OA PWC coordinates, and store agents representing Sainsbury's stores, placed in actual store location coordinates. Individual customers are loaded into the model based on the loyalty card dataset provided by Sainsbury's and have been assigned to an OAs PWC (after being aggregated from the postcode level to preserve customer anonymity). In some OAs with few individual customers, these customers have been relocated in the model using the OAC data previously linked to the customer type groups. Customers are characterised by their consumer type group, which can be one of 7 groups as outlined in section 5.5. Both their location and identity variables characterise stores; this includes their location type (city centres, town centres, residential areas, or retail parks), store class (supermarket or convenience store), and online provision (true or false).

The geographical space within the model uses the British coordinate system. Instead of having agents *be* mobile, the proposed model uses static locations for various reasons, further discussed in section 6.3. Workplace data was incorporated in section 6.3.5, in which customers are assigned both a residential location and work location and move between the two at time points. Their chosen journey paths through the environment are not captured due to modelling complexities of our transactional data, we do not know where our customers work, if they are employed, or the likely areas they are relocating to at different time periods. Section 3.3 further discusses the types of data that may be sourced and factored into the model, but complexity has been reduced to static agents for this first-time prototype model. Instead of moving agents, distances have been incorporated into the model and are used to assign which store a customer is likely to visit based on their transaction characteristics.

The length of a time step is on a day-type level, broken into weekdays or weekends, and occurs seven times per cycle to represent one whole week. The model can run infinitely but has been limited to 365 time steps, i.e., one year. Alternatively, the model can be based on an emergent outcome; for example, once the consumer has spent a set amount of money, then terminate the customer's behaviour. However, this is unreasonable for the nature of this model, as consumers will constantly be transacting; if there is not a store within their distance preferences, they will shop elsewhere.

*Process overview and scheduling.* The following actions are executed in this order, once per time step, for each customer:

- The model will read the probability of the customer making a transaction, which relies upon the day type and time of day calculations. If the customer is not transacting during that time step, the model moves on to the next step.

- For customers making a transaction, the model calculates the likelihood of the transaction occurring in-store or online. At this point, a transaction must occur.
- Once the transaction channel is determined, the basket type of the transaction is specified using the fitness proportionate selection method. The selection of basket type is selected by both the channel used and the time resolution of the transaction.

At this point in the time step, modelled customers have decided whether they will or will not make a transaction, and if they do, the transaction's time of day, channel, and basket type has been selected. The following actions are all regarding store choice.

- The model reads the chosen scenario for the transaction taking place for each customer. The scenario includes day type, time of day, channel, and basket type. Depending on these variables, the model calculates the likely distance the customer will travel for that specific transaction resolution.
- The model calculates the geodesic distance between the customer and all stores in the model. Stores that fall within the likely distance range are identified and ranked. The closest store within that distance range is chosen depending on store choice variables which relate predominantly to the basket type taking place.
- The customer's transaction is recorded for the customer and store and, expenditure is calculated.
- Outputs for the model are updated, and the next time set occurs.

*Design concepts.* The key outcomes of the model are consumer transaction behaviours and store-level sales. These outputs are a result of a decision-tree structure which is entirely data-driven. The stores visited by those customers are based on a combination of transaction variables and differ not only from consumer type group to consumer type group but also from customer to customer within the same groups, as stochasticity and geography is incorporated into the model.

*Initialisation.* A user-chosen number of households are initialised; for this study, ~210,000 individual customers are initialised, reflecting the loyalty card dataset. These customers are placed in the customer's coordinates derived from the loyalty card dataset, with some changes discussed above. Each customer is given a consumer type group to guide their behavioural rules. The locations of the customers in the model and which consumer type group they belong to are entirely customisable. The number of stores and their features are also customisable, allowing for user-testing and scenario-testing development.

*Input data*. The model uses a variety of input data:

- Consumer coordinates and group type statistics, assigning each simulated customer to a consumer type group with a home.
- Store attributes, including store coordinates, fascia, location type and sales area.
- Customer transaction probabilities are loaded into the model to support the decision tree process and weighted feature selection. These probabilities are unique to each consumer type group.
- A dataset containing distance bins containing the probabilities of customers transacting at a store within a specific distance, which is unique for each consumer type group.
- A table that assigns workplaces for customers for specific transaction types that occur on weekdays during the mornings and afternoons using 2011 Census data (dataset WF01BEW (ONS, 2011)).

### 6.1.2 Establishing a methodological framework

The core concept of the prototype IBM in this thesis is to generate transactional behaviours of grocery consumers, reflecting those observed transactions in the loyalty card-linked dataset, specifically around transaction day type, time of day, channel, and shopping basket type. To create such a model, a focus on modelling temporality is required, as each transaction is tied to a specific state of time. A second focus is required on modelling spatiality, which is interlinked with transaction temporality (section 2.2.1). Building these models from scratch is a challenging process and requires prioritisation of actions; if the base of the model does not function, then the latter work will not perform (section 3.2). Therefore, a methodological framework was created for this IBM, splitting the model into two steps, first modelling temporality, secondly modelling spatiality.

The seven unique consumer type groups identified in Chapter 5 provides us with a baseline population to simulate, in which their individual behaviours have been categorised and grouped based on similar observations. Using these groups, the likelihood of these individuals transacting can be simulated by using the group's average behaviours of transacting. The key indicators of consumer behaviour had been established in section 2.3, and these are used to guide the temporal aspect of model building. The most logical approach to model the probabilities of a consumer transacting at a given time is to use a decision tree structure. A core example of a decision tree approach in an IBM is that developed in Bell and Mgbemena (2018), in which they proposed a framework referred to as the CADET framework. The CADET framework is a data driven approach used for modelling agents and stands for the "Customer Agent DEcision Tree". The CADET design method allows for an ABM to ran using the foundations of a decision tree, similar to that used in Jordan et al. (2014)

to model residential mobility. In the context of temporal grocery transacting behaviour, the decision tree framework is used in this study to decide what type of transaction will take place by the simulated customers. Section 6.1.4 provides more detail into the methodological approach taken to simulate the temporal aspect of consumer behaviour, including the use of weighted roulette selection to incorporate behavioural stochasticity.

Due to the model being the first of its kind in the application of retail location planning, a manual implementation of the decision tree was used as opposed to machine learning (ML). The manual method was preferable to be able to identify exactly what was occurring within the simulation, allowing for continual validation against the observed data as more complexity was added. In future models, ML may be used to implement consumer temporal behaviour; however, rigorous analysis of the model output would be required and can use the output from this thesis as a validation dataset. Each 'level' of the tree is predetermined via the identification of shopping indicators in section 2.3. The tree structure allows for a simple and straightforward method to represent individual decisions within an ABM using an 'if-then' structure (DeAngelis and Diaz, 2019). Of course, human behaviours are not always 'if-then' dependent, we have considerations from all facets of life when it comes to decision making. However, using the loyalty card data, inferences can be made which logic follows an 'if-then' ideology. For example, for those in consumer type group 7 'Hungry and Far From Home', then the decision tree will *most likely* decide for a customer in that group to transaction a weekday during the morning or afternoon.

Section 6.3 introduces the approaches explored for modelling spatiality, and the challenges faced when doing so. The complex interactions between consumer transaction scenarios and store choice are difficult to incorporate in such model at the individual level, as customer locations outside of the home are unknown. Despite this, three spatial modelling approaches are presented using the ABM 'Keep it Simple, Stupid' design principle in which complexity is added at each step of model development (Edmonds and Moss, 2005). The spatial developments are used to explore to what extent loyalty card data can help the assignment of store choice from the individual level. As opposed to using more complex algorithms to 'pull' in customers to specific stores from the offset, the model development framework begins with a simplistic Tobler's first law approach in which customers will always transact at their closest store. Of course, this limits the ability to simulate the most complex behaviours such as a workplace transactions or multi-purpose shopping trip transactions. However, due to time and data accessibility constraints during this study, these more complex modelling developments must be implemented in future work; further recommendations are presented in Chapter 7.

#### 6.1.2.1 Modelling methodologies: microsimulation and agent-based modelling

The novel design of this model uses elements from both MSM and ABM, as both provide unique modelling functions appropriate for this study (Harland and Birkin, 2015). As discussed in Richiardi (2014), using these two modelling methodologies in conjunction provides a powerful approach to modelling individual behaviours, especially when not all features from one method is required.

Initially, the model developed in this thesis was to be built solely using agent-based methods in purpose-built multi-agent software. The user-friendly simulation platform 'NetLogo' (Tisue and Wilensky, 2004) was considered due to its extensive use in multiple disciplines, and ability to integrate GIS data (Walker and Johnson, 2019). However, due to the complexity of the model and its nature in handling probabilities within a decision tree, it was found that NetLogo was not an appropriate software as it is unable to utilise nested probabilities. Additionally, despite NetLogo incorporating GIS data, it struggles to capture non-arbitrary distances (Scott and Koehler, 2011), an essential part of the proposed model. The drawbacks from NetLogo and the application of this type of IBM are highlighted in section 6.1.3, in which a notable flaw was identified in Sturley et al. (2018).

Whilst designing the model, several elements of ABM modelling were found to not be required for the model in this thesis. ABMs are profound for their ability to incorporate interactions by individual agents within an artificial environment, which can adapt over time (Bonabeau, 2002; Crooks and Heppenstall, 2012). However, this study does not involve individual agent's (customers) interactions, as the focus is on their interactions between stores as opposed to one another. Additionally, ABMs utilise an element of mobility, as agents can move across the virtual environment, either randomly or following rules.

The first time this project was discussed, we had this visual of a fully-fledged ABM in which Sainsbury's customers traversed across a virtual West Yorkshire, moving between places at different time points, and making transactions along the way. After consulting the data, literature, and designing the model, it quickly became evident that such a model is beyond the scope of this thesis. As this is the first attempt at building such a model, it would be either impossible or highly unlikely that a fully validated and calibrated ABM could be produced. That is not to say it can never be done, it just is beyond the scope of this study.

The key challenges for creating such a sophisticated model include the lack of data regarding *who* the customers are within the loyalty card data? (are they employed? how many family members do they have? do they own cars?), *where* do the customers go throughout the week? (are they employed? do they work from home? do they have children?), and *what* other retailers do they purchase their groceries from? (are they loyalty to one brand? do they only use Sainsbury's for opportune transactions?). Additional data is required about the observed customers to be able to fully simulate their movements across space and time, and to do so accurately. These types of issues in data-driven ABMs are notorious, especially when simulating a variety of complex interactions in attempt to simulate reality (Malleson et al., 2022; Wilkinson, 2023).

As the temporal aspect of the model does not incorporate all features within ABM, a hybrid ABM and MSM approach was used. Agents in the model do not move from point A to point B following paths across space, instead they are relocated at time steps to reflect different types of grocery demand (for example, workday hours they are moved to workplaces, and return to their homes for evenings and weekend). As the model requires agents to have spatial awareness, an ABM approach was used in which agents are aware of their surroundings and the stores available to them. At different time steps of the model, and when different transactions scenarios are chosen using the decision tree (presented in section 6.1.4), customer agents follow different behavioural rules to better represent the observed customers in the loyalty card dataset (using the mined behaviours identified for each consumer type group in section 6.2).

The model uses a unique approach in a novel application and was built entirely from scratch. This was written using the Python programming language; Python is a high-level general-purpose programming language that supports many programming paradigms and is found to be a flexible tool for this study (Nagpal and Gabrani, 2019). Due to Python's versatility, aspects of decision trees, microsimulation, and agent-based modelling can be incorporated into the model, all which were used to build the model.

Prior to discussing the model's design and application in-depth (section 6.1.4), the following section provides a critical analysis of the most recent study that attempts to simulate the store and channel choice of grocery consumers via a proof-of-concept model. The model logic in Sturley et al. (2018) was used as a guide for this thesis, as requested by Sainsbury's, and significant drawbacks of their method and software choice are highlighted. The findings from critically assessing their work helped enhance the model presented in this thesis.

### 6.1.3 Expanding on previous work (Sturley et al., 2018)

The study by Sturley et al. (2018) inspired this thesis, in which Sainsbury's Supermarkets wished to understand the potential of alternative modelling techniques, such as IBM, to incorporate nuanced customer behaviours. Sturley et al. (2018) undertook the research challenge of building a proof-of-concept ABM, incorporating pseudo-transaction data. Their findings, at a conceptual level, found it possible to design and build such a model, despite some significant challenges. The following section discusses what Sturley et al. (2018) achieved, and how it has been used to support the development of the model presented within this thesis.

The study by Sturley et al. (2018) can be divided into three essential parts: first, analysing a transaction dataset to create customer type groups. Second, deriving probabilistic behavioural rules from those customer type groups, and third, building an agent-based model in NetLogo using those behavioural rules in an abstract environment loosely representing the study area of Leeds. Their study uses pseudo-transactional data based on Sainsbury's Nectar card data, in which customers were assigned homes within the Leeds local government area. The key aim of their study was to build a proof-of-concept model that incorporates the captured consumer behaviours observed within the loyalty card dataset. Their model focuses on recreating the behaviours regarding transaction frequency, time of day visits, and distance travelled.

The first part of their study entails mining similar consumer behaviours from the dataset. The indicators of consumer behaviour regarding store choice were chosen, which included: shopping frequency, number of stores visited, the percentage of transactions that took place at a convenience store, the percentage of high-value transactions, how many transactions took place on a weekday or evening, and the distance to the customer's most frequently visited store (Sturley et al., 2018, p.33). Each synthetic customer's shopping indicators were calculated to create an overall summary. They were then run through a k-means clustering algorithm in which a seven-cluster solution was identified as the optimal number of clusters. Each of the seven consumer type groups varied in shopping preferences, with some preferring supermarkets over convenience stores, some travelling further than others and at certain times of the day and some who spent more money at Sainsbury's than others.

The second part of Sturley et al. (2018) attempts to summarise the behaviours of those seven consumer type groups for model implementation. Consumer type group characteristics were analysed and summarised to create behavioural rules for customers in the model, as shown in Table 6.1. What determines these characteristics as 'low', 'average', 'high', 'short' and so forth is not fully explained in this study but is used to signify differences in consumer behaviour between the consumer type groups. The probabilities of transacting for each different consumer type group were calculated based on day type (weekday or weekend) and day part (morning, afternoon, or evening) to determine the chance of customers making a transaction at any given time using the summaries in Table 6.1.

**Table 6.1** Summary of consumer type groups characteristics used in Sturley et al., (2018, p.36)

| Consumer type | Store type | Frequency | Time | Distance | Car ownership | Spend |
|---|---|---|---|---|---|---|
| 1 | Supermarket/ Online | Low | Evening | Average | High | High |
| 2 | Supermarket | Low | Weekday daytime | Short | Low | High |
| 3 | Convenience | Average | Weekday evening | Short | Average | Low |
| 4 | Supermarket/ Convenience | Very high | Weekday | Very short | Average | Low |
| 5 | Supermarket/ Convenience | High | No preference | Average | High | Medium |
| 6 | Supermarket | Low | Weekend daytime | Average | Above Average | Medium |
| 7 | Supermarket/ Convenience | Low | Weekday evening | Long | Average | Low |

The probabilities used to assign agent behaviour in their model are shown in Table 6.2. The probabilities implemented are incredibly simplistic, with the chances being one number or another, resulting in a primitive model. The choice of using such simplistic (and most likely not representative) probabilities indicates a fundamental limitation in either the quality of the data used, the method and software chosen, or a time constraint on developing the model. The use of these specific probabilities is discussed further in the next section.

**Table 6.2** Probabilities of an agent making a shopping trip by consumer type, derived and implemented in Sturley et al. (2018, p.37)

| | | Consumer type | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Day | Day Part | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Weekdays | Morning | 0 | 50 | 20 | 90 | 50 | 0 | 0 |
| Weekdays | Afternoon | 0 | 50 | 20 | 90 | 50 | 0 | 0 |
| Weekdays | Evening | 75 | 0 | 75 | 90 | 50 | 0 | 40 |
| Weekend | Morning | 0 | 0 | 20 | 0 | 50 | 50 | 0 |
| Weekend | Afternoon | 0 | 0 | 20 | 0 | 50 | 50 | 0 |
| Weekend | Evening | 75 | 0 | 20 | 0 | 50 | 0 | 0 |

The third and final part of their study involves the development of the prototype agent-based model. Their model is based on a 21-tick cycle representing a whole 7-day week, each tick representing a day and day part. For example, ticks 0, 3, 6, 9, and 12 represent ‘weekday mornings’, whereas ticks 17 and 20 represent ‘weekend evenings’. As each tick in the model runs, every agent (simulated customer) has a random number generated for them to determine whether they will make a transaction. If the customer has been chosen to make a transaction, then that customer needs to be assigned a store. Due to each consumer type group having different shopping characteristics (Table 6.1), their behaviours have been coded to differ. To implement these differing distance preferences, the researchers used a simple logic for customers likely to travel “long” distances; they have been hard coded to transact at any store, whereas those who travel “short” distances will only transact at their closest store.

Each time a customer visits a store, the distance travelled is recorded, allowing the user to read the average distance customers of a specific consumer type travel. The researchers had data about the actual distances travelled by those in the transaction dataset, but these were not directly applied to the model. This is likely a limitation due to their software choice, as distance in NetLogo is entirely arbitrary, and it is difficult to identify and create a representative geographical scale (Scott and Koehler, 2011). The researchers could have incorporated a road network system into their model using NetLogo (Chen and Crooks, 2021); however, this is highly computationally intensive. For further complexity, customers who spend a “high” amount of money will only transact at supermarkets. However, those who spend a “low” amount will transact at supermarkets or convenience stores. Despite not using direct statistics from their data and using more abstract measures such as “low” or “high”, the model captures the shopping indicators that were previously identified (Sturley et al., 2018, p.33).

The model's artificial environment is designed to represent an abstract interpretation of part of northern Leeds, as shown in Figure 6.1. A vital feature of the transaction dataset used in this study was the residential address provided at the OA level. The OAC was analysed using these data to determine which customer groups are likely to reside within a particular OAC Supergroup (Sturley et al., (2018, p.38, figure 2a.). The researchers do not fully explain their process of determining which OAC supergroup their consumer type groups fall within, as each group is likely to contain customers belonging to various OAC groups. Additionally, there is likely to be a skew of customers to specific OAC groups across all seven consumer type groups due to the expected target demographic of Sainsbury's customers, the location of Sainsbury's stores and the general populous of north Leeds as found in this thesis (section 4.1). Regardless, using a colour system, the NetLogo model populates the environment with customers of distinct types based on the OAC demographic groups and inserts the different stores.

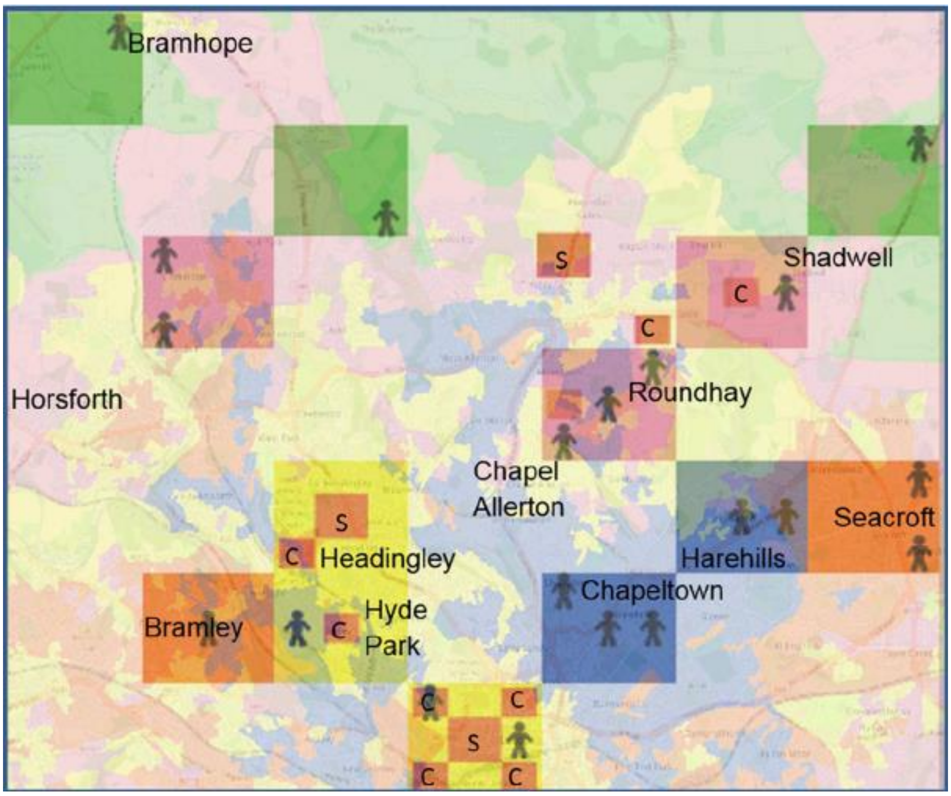



**Figure 6.1** NetLogo abstract environment based on the northern area of the city of Leeds created and used in Sturley et al., (2018, p.37). Squares labelled as ‘C’ are convenience stores, and squares labelled as ‘S’ are supermarkets.

Table 6.3 is an excerpt of code used to determine the behaviours of customers belonging to consumer type 1 using the numbers in Table 6.2.

**Table 6.3** Code extract in Sturley et al. (2018) NetLogo file.

| | |
|---|---|
| 1 | ask consumer-as [ ; (represents consumer type 1 group) |
| 2 | set probability 0; Set the initial probability of going to a store |
| 3 | if member? ticks [2 5 8 11 14 17 20] [ |
| 4 | set probability 75; Update the probability based on the time counter. |
| 5 | ] |
| 6 | if random 100 < probability [; Compare the probability to a randomly generate number. |
| 7 | shop; Tell the consumer to go to a store (call the ‘shop’ procedure) |
| 8 | set destination one-of supermarkets; Reset the consumers destination. |
| 9 | ] |
| 10 | ] |

With this logic (Table 6.3), line 2 instructs the model to have 0 chance of any customer in this consumer type group making a transaction at any time point or tick. However, line 3 states that if the model is at tick 2, 5, 8, 11, 14, 17, or 20, which represents any 'evening' time point, then the probability of a customer in that consumer type group is set to 75. Line 6 asks the model to generate a random number between 1 and 100. If the random number is less than 75, then the model triggers a transaction for that customer. Therefore, a customer of consumer type 1 has a 75% chance of making a transaction daily in the model during the evening. At any other time, there is a 0% chance of transacting. If the customer makes a transaction, the model has been coded to ensure that the transaction occurs only at a supermarket, where distance is not considered. Due to this consumer type group only making transactions at supermarkets, the researchers coded for all supermarket transactions to be worth £100 or £120 if the consumer type group contained 'high value' customers. If a transaction occurred at a convenience store, the transaction would be worth £20.

After running the model on a 21-tick cycle 100 times for one customer belonging to consumer type 1, it was found that this customer, on average, transacted six times a week out of a possible seven, which does not correlate with a 'low' frequency shopper as stated in Table 6.1. The model was also tested on a 21-tick cycle 100 times for one customer belonging to consumer type 4, who has a 90% chance of making a transaction on any day part on a weekday. The results of this test found that, on average, this customer transacted fourteen times a week and spent an average total of ~£400 per week. Considering this customer is aimed to be a 'low' spend consumer (Table 6.1), the model is fundamentally flawed due to how probabilities have been coded. Instead of breaking down the probability of making a transaction to day part (or each tick), the researchers applied the day's chance of transacting down to each tick, therefore overly modelling the actual chances of transacting. To combat the probability flaw identified in Sturley et al. (2018), this thesis aims to identify and implement a more accurate method of converting the observed behaviours of customers into probabilities that are correct and representative of the data.

The overall logic of the study by Sturley et al. (2018) is sensible, providing a clear and concise structure of analysing a transaction dataset, generating consumer type groups, and subsequently using these groups to build an agent-based model. The study demonstrates how agent-based modelling can generate new data based on an empirical dataset, mimicking the behaviours of ‘real’ customers. The researchers discuss how elements of spatial interaction modelling can be implemented, such as store attractiveness, to support store choice behaviours and the incorporation of demographics. However, a more sophisticated method of deriving the probabilities of consumers making a transaction in the model needs to be addressed before attempting to incorporate these elements.

Therefore, this thesis follows a similar route to building the individual-level model. First, the loyalty card-linked transaction dataset is analysed to identify distinct groups of customers who exhibited similar shopping behaviours. Next, customers are clustered into their unique groups, notably separating those who shop via different channels, such as those who only shop in-store, only online or multi-channel. The probability of them transacting at different time points is calculated, incorporating which channel they will likely use and what type of basket they will purchase. By calculating the probabilities of a customer transacting based on a day and time of day basis, the model evades the flaw found in Sturley et al. (2018) which did not consider the day part. The study then proceeds onto the model-building process, differing in the tools and methodologies used by Sturley et al. (2018). Instead, this thesis incorporates geographical distances and uses a decision tree as a base. Other studies, such as Garcia-Magarino et al. (2018) used NetLogo to build a decision tree; however, it works on a true/false basis. Realistically, the model in this thesis requires more than true/false choice; therefore, it utilises Python to build the decision tree.

### 6.1.4 Decision trees of transactional behaviour

The key objective of this thesis is to utilise the loyalty card-linked transaction data provided by Sainsbury's to design an individual-level model. A decision tree was created as a foundational base for the agent-based model, incorporating data directly from the transaction dataset. The analysis in Chapter 5 was further explored, calculating the probability of a customer of any consumer type group making a transaction under different scenarios (section 6.2). The following section discusses the process of identifying and extracting key behaviours of customers in the dataset to create a set of realistic and robust behavioural rules for the model.

One of the most crucial elements of the model is to include temporality, as customer shopping behaviours vary over space, time, and needs. Therefore, the model performs on an iterative time step basis, in which each time step represents a 24-hour day. The decision tree is made up of nodes, branches and leaves. All nodes represent key factors of consumer behaviour and are connected through branches which are decisions made by the customer, for example, transact = yes or no. 'Leaves' result in the termination of further decisions, such as at the end of the tree where the final decision is made. All branches in the tree lead to either a node or leaf.

Depending on the time step, the customers either behave as if it is a weekday (time steps 1 to 5), or weekend (time steps 6 to 7). To determine whether the customer will transact on a given time step, a weighted roulette wheel selection function within the first node is triggered based on the probabilities of a customer transacting at the day-level. The roulette wheel selection, commonly known as *fitness proportionate selection*, is used due to its ability to weight probabilities, best representing the behaviours of the customers in the model (Mitchell, 1998; Bell and Mgbemena, 2018). Essentially, the model asks a simple question: "what is the chance of this customer making a transaction?". Each customer is ‘asked’ this question, and the model reads in the probability based on the customer’s consumer type group. This first node has a ‘yes’ branch connecting to ‘time of day’, and a ‘no’ leaf which terminates transaction options for that customer and moves onto the next time step.

If the customer does transact, then the weighted roulette wheel selection function is used until the full profile of the transaction is recorded. The transaction profile includes the key factors (represented as nodes) of the day type of the transaction (weekday or weekend), day part (morning, afternoon, or evening), channel (online or in-store), and basket type (for now, top-up or main) (Figure 6.2). The probabilities have been calculated for every combination of the factors above and for each consumer type group to fully represent the observed behaviours of those in the loyalty card-linked transaction dataset (section 6.2). The method of creating these probabilities is entirely novel, and to the best of our knowledge, the first time this has been developed in such a way. By breaking down the probabilities to account for transactional behavioural differences in time, space, and purpose, the model developed in this study provides a solution to the key limitation found in Sturley et al. (2018), as discussed in section 6.1.3. The following section outlines how the probabilities for customer behaviour were directly derived from the loyalty card-linked transactional dataset.

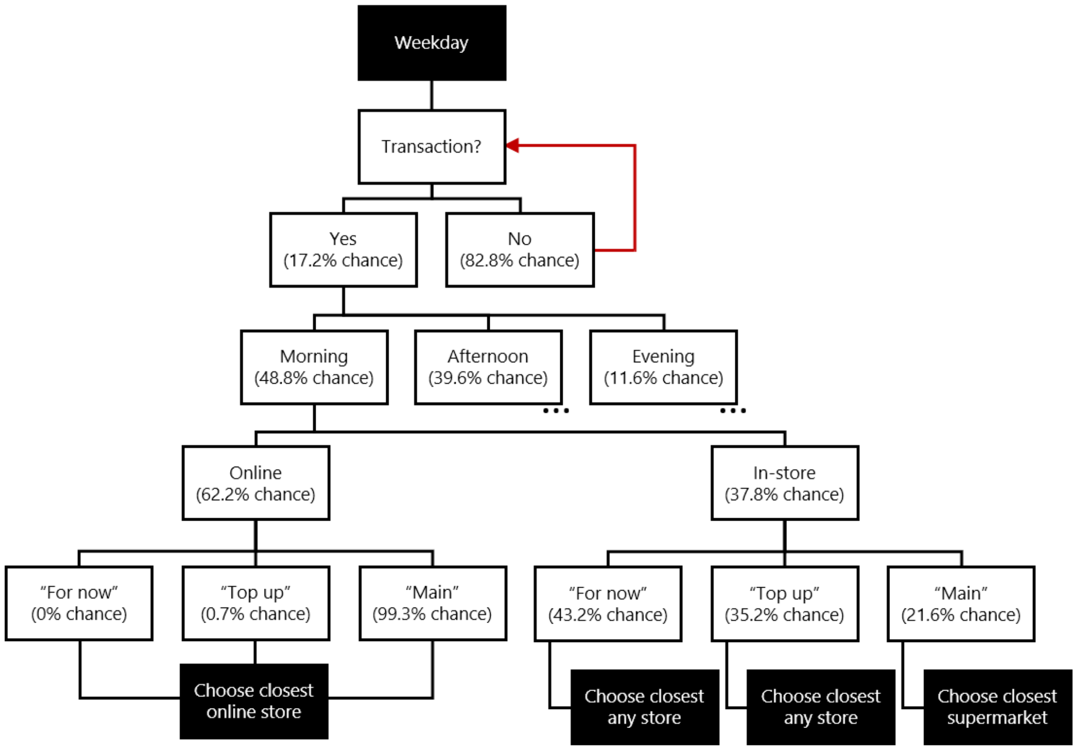



**Figure 6.2** Decision tree process for each iteration for customer type group 2, probabilities are also calculated for afternoon and evening however are not included for demonstrative purposes.

Logic tests were performed before building the complete decision tree code to identify whether the model functions as intended. Random customers of different consumer type groups were manually put through a makeshift model and compared to the input data. Once the logic was found to be logically sound, the decision tree was built in Python. Initially, NetLogo was considered; however, it was found that the software is limited in nesting multiple functions and weighted probabilities. This was likely a prominent issue in Sturley et al. (2018) as that model avoided the complexity of more than two customer choices. Using Python, the weighted roulette wheel selection function was used to apply weighting to the probability values at each decision tree branch. A random number is generated at each node; the number is weighted depending on the probabilities used in the model. Using Figure 6.2 as an example, the model weights the ‘no’ option as heavier than ‘yes’ at the first node (‘transaction?’); therefore,  it is more likely that the customer will not transact on that weekday. If the customer does transact, the model weights the ‘morning’ branch heavier than ‘afternoon’ or ‘evening’, and so forth.

The decision tree provides the agent-based model’s behavioural foundation, reading in the exact behaviours of those customers observed within the dataset at a semi-aggregated level. This data-driven approach has rarely been used in individual-based modelling; often as individual-level data of this calibre are challenging to work with. The novel application of decision trees presented here provides retailers the ability to replicate the genuine behaviours of their customers. By using a weighted roulette wheel selection function, the model allows for variation in unique behaviours of customers who belong to the same consumer type group. For example, a customer might have a higher chance of transacting on a morning based on the weighted probabilities but might transact in the afternoon instead based on the weighted roulette wheel selection function output. This allows for dynamic transaction behaviour within the limits of the consumer type group’s behavioural boundaries.

As the model was built using entirely empirical data, the model’s simulated output reflects the observed behaviours of those in the original dataset. Yet, by using the weighted roulette wheel selection function, the model incorporated an element of stochasticity, allowing for some flexibility in consumer type behaviour. This allows for a direct comparison between model output and input for model validation to ensure the code performs as intended. Additionally, any supply-side changes in the model, such as an introduction of new stores, store relocations, closures or additions of online provision, can all be simulated to see how consumers’ store choices will change and which current stores will be impacted. This novel method of building a foundational base of an agent-based model is yet to be documented in academic literature and provides an alternative to the standard method, making more use of empirical data.

## 6.2 Modelling consumer transaction temporality

Each consumer type group was individually assessed to derive the probability of a customer making a transaction at any time. Each factor was broken down to calculate the average number of transactions under different circumstances. For example, the average number of ‘weekday’ transactions by customers, the average number of ‘for now’ transactions, or ‘online’ transactions. These factors were initially calculated individually and processed to account for the time period of the dataset (12-weeks). To build the decision tree, the probabilities were compounded for each step of a transaction decision. For example, the probabilities for transacting on a ‘weekday’ versus ‘weekend’ were initially calculated, then the probabilities of a ‘weekday’ morning’, ‘weekday afternoon’, and ‘weekend evening’. This process was continued until there were individual probabilities at the most discrete level of a transaction scenario after all decisions had been made, for example, ‘weekday morning instore main’.

To calculate the probabilities, the number of transactions of a specific type had to be broken down into a day level. The following was performed for each consumer type group:

**Step 1.** Calculate each customer’s total number of transactions in the loyalty card-linked dataset.

**Step 2.** Divide the total number of transactions by 12 to determine how many transactions took place on average by individuals.

**Step 3.** Normalise the data by multiplying the count of transactions by the average transactions per week, i.e., apply a weighting factor.

**Step 4.** Divide the normalised value by the total number of transactions overall to identify the proportion of total transactions that took place per person for a week under different circumstances (weekday, morning, online, main, and so forth).

**Step 5.** Calculate the percent chance of a customer making a transaction at each event opportunity in the model based on day type. All transactions were split into weekdays and weekends; if a transaction took place on a weekday, the total transactions were divided by 5, and the total average weekend transactions were divided by 2.

The formula below demonstrates the calculation for deriving probabilities per day for each transaction factor.

$$p = \frac{\left(\frac{(t \cdot \bar{x})}{n}\right)}{dt}$$

Where,

$t$ = number of transactions per type

$\bar{x}$ = average of total transactions per week

$n$ = total number of transactions

$dt$ = day type

**Example of the equation in action**: How can the model predict the behaviour of customer A if they only perform top-up transactions? In the original dataset, customer

A has made 13 transactions over the 12 weeks. Out of those transactions, 7 were top-up baskets. Therefore, on average, customer A purchased 1.083 top-up baskets per week. To calculate the weighted normalised result, multiply the 7 top-up baskets by 1.083, resulting in 7.583 top-up baskets over 12-weeks. To find the percentage of total top-up transactions per week, 7.853 is divided by 13, resulting in 0.583, meaning 58% of transactions are top-up baskets per week. To implement into the model on a time step day basis, 0.5833 is divided by 5 to calculate the chance of transacting top-up baskets on a weekday and divided by 2 for weekend values.

The proportion of transactions occurring under all transaction scenarios was calculated in a single row per customer. The customer's consumer type group was then appended to the rows to calculate the consumer type group's average behaviours. All transactional scenarios were averaged and calculated into proportions for each consumer type group. Next, the percentage chance of transacting for each consumer type group was calculated for any combination of transaction scenarios (see below). For the weighted roulette wheel selection function to perform, the percentage chance of transaction scenarios was standardised to 100%. For example, if the percentage chance of a transaction on a weekday was 50%, and the percentage chance of a transaction on a weekend was 20%, these were standardised to 100% by calculating 50% / (50% + 20%) and 20% / (50% + 20%). This would result in the final probabilities of a 71.4% chance of transacting on a weekday and a 28.6% chance of transacting on a weekend. This removes the ability for the model to roll a *null* action.

Probabilities for all transaction scenarios were calculated such as the following:

% transactions on a weekday

% transactions on a weekend

% transactions online

% transactions instore

% transactions morning

% transactions afternoon

% transactions evening

% transactions classed as for now

% transactions classed as top up

% transactions classed as evening

These were then further broken down into:

% transactions on a weekday online morning for now

% transactions on a weekday online morning top up

% transactions on a weekday online morning main

% transactions on a weekend online morning for now

% transactions on a weekend online morning top up

% transactions on a weekend online morning main

% transactions on a weekday instore morning for now

% transactions on a weekday instore morning top up

% transactions on a weekday instore morning main,

This was repeated for afternoons and evenings.

The output of these calculations is the percentage chance of any customer of a customer type group making a transaction. Figure 6.3 and Figure 6.4 provide an overview of the probabilities derived from the loyalty card dataset, linking each step of the model. Step 1 is simple with a weighed ‘yes’ or ‘no’ roulette wheel decision, with the chance being relatively low to make a transaction; this is due to calculating the probabilities on a day rate, an integral calculation missing in Sturley et al. (2018), as discussed in 6.1.3. Step 2 becomes more complex and is only reached if a transaction was rolled as ‘yes’ in step 1. This continues for step 3, becoming increasingly more specific regarding when the transaction occurs, what channel to use, and for what shopping basket. Finally, the model completes step 4 (Figure 6.4) in which the transaction basket is chosen. The decision tree is triggered for each model time step for every customer, recording the transaction they performed.

| | Day type | Customer type group probabilities 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Step one: Chance of transaction on said tick? (yes/no) | weekday | 8.8% | 17.2% | 12.7% | 21.8% | 17.1% | 25.6% | 21.3% |
| | weekday | 8.8% | 17.2% | 12.7% | 21.8% | 17.1% | 25.6% | 21.3% |
| | weekday | 8.8% | 17.2% | 12.7% | 21.8% | 17.1% | 25.6% | 21.3% |
| | weekday | 8.8% | 17.2% | 12.7% | 21.8% | 17.1% | 25.6% | 21.3% |
| | weekday | 8.8% | 17.2% | 12.7% | 21.8% | 17.1% | 25.6% | 21.3% |
| | weekend | 8.2% | 16.0% | 11.7% | 16.2% | 35.7% | 10.9% | 13.5% |
| | weekend | 8.2% | 16.0% | 11.7% | 16.2% | 35.7% | 10.9% | 13.5% |

| | Day/time | Customer type group probabilities 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Step two: If yes above, when will transaction take place? (Morning, afternoon, or evening) If no above, skip to tick 2 for customer | weekday morning | 95.9% | 48.8% | 26.1% | 10.5% | 18.0% | 48.3% | 31.9% |
| | weekday afternoon | 4.0% | 39.6% | 55.6% | 36.6% | 62.2% | 48.2% | 54.5% |
| | weekday evening | 0.1% | 11.6% | 18.3% | 52.9% | 19.9% | 3.5% | 13.6% |
| | weekend morning | 95.0% | 52.5% | 32.2% | 23.1% | 16.6% | 50.0% | 31.1% |
| | weekend afternoon | 4.9% | 41.1% | 55.5% | 58.7% | 78.0% | 45.8% | 57.1% |
| | weekend evening | 0.1% | 6.4% | 12.2% | 18.2% | 5.3% | 4.2% | 11.8% |

| | Day/time & channel | Customer type group probabilities 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Step three: Depending on day & time of transaction, roll probability for which shopping channel (Online or instore) | weekday morning online | 100.0% | 62.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday morning instore | 0.0% | 37.8% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | weekday afternoon online | 100.0% | 4.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday afternoon instore | 0.0% | 95.3% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | weekday evening online | 100.0% | 0.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday evening instore | 0.0% | 99.7% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | weekend morning online | 100.0% | 62.9% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend morning instore | 0.0% | 37.1% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | weekend afternoon online | 100.0% | 4.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend afternoon instore | 0.0% | 95.3% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |
| | weekend evening online | 100.0% | 0.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend evening instore | 0.0% | 99.7% | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**Figure 6.3** Probability results for steps 1 to 3 of the agent-based model's decision tree for each customer type group.

| | | Customer type group probabilities | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **Day/time & channel & basket type** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **Step four: Depending on day type, time of day and channel rolled, what basket type will be purchased? (Food for now, topup, or main?)** | weekday morning online main | 98.5% | 99.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday morning online now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday morning online topup | 1.5% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday afternoon online main | 99.4% | 99.4% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday afternoon online now | 0.0% | 0.1% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday afternoon online topup | 0.6% | 0.6% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday evening online main | 100.0% | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday evening online now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday evening online topup | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend morning online main | 98.8% | 99.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend morning online now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend morning online topup | 1.2% | 0.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend afternoon online main | 99.0% | 98.6% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend afternoon online now | 0.0% | 0.2% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend afternoon online topup | 1.0% | 1.3% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend evening online main | 83.3% | 100.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend evening online now | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend evening online topup | 16.7% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekday morning instore main | 0.0% | 21.6% | 8.5% | 22.4% | 21.3% | 58.9% | 3.4% |
| | weekday morning instore now | 0.0% | 43.2% | 46.8% | 38.9% | 39.1% | 12.7% | 77.2% |
| | weekday morning instore topup | 0.0% | 35.2% | 44.6% | 38.7% | 39.5% | 28.3% | 19.4% |
| | weekday afternoon instore main | 0.0% | 20.3% | 11.0% | 33.3% | 29.7% | 55.5% | 6.3% |
| | weekday afternoon instore now | 0.0% | 38.8% | 42.9% | 25.4% | 25.6% | 14.7% | 64.9% |
| | weekday afternoon instore topup | 0.0% | 40.9% | 46.1% | 41.4% | 44.7% | 29.9% | 28.8% |
| | weekday evening instore main | 0.0% | 19.2% | 5.6% | 49.7% | 19.8% | 26.9% | 5.0% |
| | weekday evening instore now | 0.0% | 39.3% | 50.5% | 14.3% | 30.1% | 31.5% | 63.2% |
| | weekday evening instore topup | 0.0% | 41.6% | 43.9% | 36.1% | 50.1% | 41.6% | 31.9% |
| | weekend morning instore main | 0.0% | 32.7% | 21.8% | 43.6% | 60.7% | 40.1% | 12.1% |
| | weekend morning instore now | 0.0% | 29.1% | 36.6% | 21.6% | 13.0% | 23.4% | 60.9% |
| | weekend morning instore topup | 0.0% | 38.2% | 41.7% | 34.8% | 26.3% | 36.5% | 27.0% |
| | weekend afternoon instore main | 0.0% | 51.9% | 25.1% | 62.7% | 89.0% | 57.2% | 15.5% |
| | weekend afternoon instore now | 0.0% | 48.1% | 74.9% | 37.3% | 11.0% | 42.8% | 84.5% |
| | weekend afternoon instore topup | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% |
| | weekend evening instore main | 0.0% | 17.2% | 8.0% | 35.0% | 45.0% | 21.2% | 6.8% |
| | weekend evening instore now | 0.0% | 42.6% | 51.2% | 21.1% | 20.6% | 33.8% | 61.3% |
| | weekend evening instore topup | 0.0% | 40.2% | 40.8% | 43.9% | 34.4% | 45.0% | 31.9% |

**Figure 6.4** Probability results for steps 1 to 3 of the agent-based model's decision tree for each customer type group.

### 6.2.1 Model validation

This section focuses on the validation of simulating temporal consumer behaviours. The model aims to create a simulated output that reflects the loyalty card dataset (hereby referred to as the *observed* dataset) whilst incorporating behavioural stochasticity. This provides the model with both a degree of predictability and variation, representative of human nature, notably in consumer purchasing behaviour.

To assess the model's performance and ability to represent the observed transactional dataset, the model was executed 30 times, producing 12 weeks' worth of transactional data each time. These model runs were used to assess model stability and validate output against the observed transactional dataset provided by the study collaborator. To assess agreement between the observed data and simulation data, the following were calculated: central tendency (mean), accuracy (relative bias (RB)), precision (coefficient of variation (CV)), and variance (standard deviation and 95% confidence intervals).

Initially, the overall behaviour of every consumer type group was compared to the observed dataset (Table 6.4). The most pertinent factor to consider when looking at the model output data was the total number of transactions per consumer type group and comparing these against the observed dataset. The total number of transactions over model runs 1 to 30 were calculated, and the mean was determined per consumer type group. The average number of transactions per consumer type group was similar to those in the observed dataset, with a high degree of precision as indicated by the low CV values of ≤ 0.57%. The RB, or accuracy, is a measure of the closeness of the simulated results to the expected observed data values. The accuracy of the mean transaction numbers for each consumer type group was high, with RB results within a range of -0.09% to +0.03%. The variance of the number of transactions was measured using the standard deviation of the total number of transactions over the 30 model runs. The standard deviation for each consumer type group shows the spread of the data from the mean number of transactions. For example, consumer type group 1 has a standard deviation of 168.03, meaning that the 1 standard deviation above the mean is 29,434 + 168.03 transactions. This infers a level of stochasticity and variance within consumer type group 1 across the 30 model runs. The standard deviation is proportional to the mean number of transactions of a consumer type group which is expected as the probabilities are static but use a weighted roulette selection function. The analysis demonstrates that the frequency of transactions for each consumer type group is simulated with a high degree of accuracy and precision and is successful in integrating a degree of randomness in consumer behaviour.

**Table 6.4** Total transaction comparison between original data and model output over 30 runs by consumer type group.

| Consumer Type Group | Observed Data | Model runs 1 to 30 | | | RB (%) |
|---|---|---|---|---|---|
| | Total transactions | Mean Transactions | Standard Deviation | CV (%) | |
| **1** | 29,462 | 29,434 | 168.03 | 0.57 | -0.09 |
| **2** | 120,999 | 121,019 | 308.92 | 0.26 | 0.02 |
| **3** | 1,528,663 | 1,528,896 | 1288.04 | 0.08 | 0.02 |
| **4** | 129,775 | 129,784 | 313.84 | 0.24 | 0.01 |
| **5** | 140,563 | 140,528 | 395.97 | 0.28 | -0.02 |
| **6** | 250,760 | 250,845 | 446.95 | 0.18 | 0.03 |
| **7** | 458,731 | 458,661 | 669.71 | 0.15 | -0.02 |

The model validation has thus far demonstrated that the number of transactions that the model produced is representative of consumer type group behaviour; this section aims to demonstrate the model's ability to capture transaction scenario behaviours accurately and precisely. The output data from model runs 1 to 30 were compared to the observed data transactions in terms of proportional probabilities of transaction behaviour for each transaction scenario.

#### 6.2.1.1 Consumer type group 1

Consumer type group 1 consists of 4,063 customers who solely transacted using the online channel. Their most frequent transaction scenarios were 'weekday morning online main' and 'weekend morning online main', contributing 48.98% and 45.20% of transactions, respectively (Table 6.5). Less frequent transaction scenarios include 'weekday afternoon online main' (2.05%), 'weekend afternoon online main' (2.35%) and 'weekday morning online top-up' (0.76%). The remaining scenarios contributed ≤ 0.69% collectively. 'Weekday morning online for now' had a < 0.00% probability of occurring but is not equal to zero (0.0024%).

As shown in Table 6.5, the mean transactions (%) produced by the model follow the same pattern as the observed data regarding proportional probabilities. For example, the 'weekday morning online main' and 'weekend morning online main' mean transactions (%) were 48.98% and 45.21%, respectively. To ensure the accuracy of the mean transactions (%) results, a 95% confidence interval assessment was completed using the individual proportional transaction data from each model run. The confidence intervals were calculated to identify the percentage of transactions that contain the actual population mean. This means that the 'weekday morning online main' percentage transaction range of 48.98% ± 0.001059 has a 95% chance of containing the true population mean.

**Table 6.5** Consumer type group 1 model output summary.

| Consumer Type Group 1 (4,063 customers) | | | | |
|---|---|---|---|---|
| **Transaction Scenario** | **Observed Data Transactions (%)** | **Model runs 1 to 30** | | |
| | | **Mean Transactions (%)** | **CV (%)** | **Confidence Interval (95%)** |
| Weekday Morning Online For Now | 0.00% | 0.00% | 86.54 | 0.000010 |
| Weekday Morning Online Top-up | 0.76% | 0.75% | 4.44 | 0.000121 |
| Weekday Morning Online Main | 48.98% | 48.98% | 0.59 | 0.001059 |
| Weekday Afternoon Online Top-up | 0.01% | 0.01% | 56.92 | 0.000022 |
| Weekday Afternoon Online Main | 2.05% | 2.05% | 3.99 | 0.000298 |
| Weekday Evening Online Main | 0.05% | 0.05% | 28.04 | 0.000051 |
| Weekend Morning Online Top-up | 0.54% | 0.53% | 9.90 | 0.000190 |
| Weekend Morning Online Main | 45.20% | 45.21% | 0.72 | 0.001179 |
| Weekend Afternoon Online Top-up | 0.02% | 0.02% | 41.90 | 0.000037 |
| Weekend Afternoon Online Main | 2.35% | 2.36% | 4.60 | 0.000395 |
| Weekend Evening Online Top-up | 0.01% | 0.01% | 97.25 | 0.000020 |
| Weekend Evening Online Main | 0.03% | 0.03% | 45.54 | 0.000048 |

To measure the consistency of the model runs, the CV (%) was calculated for the percentage of transactions for each transaction scenario (Table 6.5). Several transaction scenarios demonstrated a high level of precision with CV values ≤ 9.90%: ‘weekday morning online top-up’ (4.44%), ‘weekday morning online main’ (0.59%), ‘weekday afternoon online main’ (3.99%), ‘weekend morning online top-up’ (9.90%), ‘weekend morning online main’ (0.72%), ‘weekend afternoon online main’ (4.60%). The remaining transaction scenarios had high CV values ranging from 28.04% (‘weekday evening online main’) to 97.25% (‘weekday evening online top-up’), which indicates a low precision level for these transaction scenarios. To visualise this variability, a Bland-Altman plot showing the RB (accuracy) of each transaction scenario across the 30 model runs was produced (Figure 6.5). Various methodologies are used to validate individual-based models, and the Bland-Altman method (Bland and Altman, 1999) was found to be a robust analytical method in this type of simulative study (Timmins and Edwards, 2016). In Figure 6.5, each dot represents the result of an individual model run. For each transaction scenario, there are 30 dots representing the relative bias for each model run. As mentioned previously, RB measures the closeness of the simulated results to the expected observed data values.

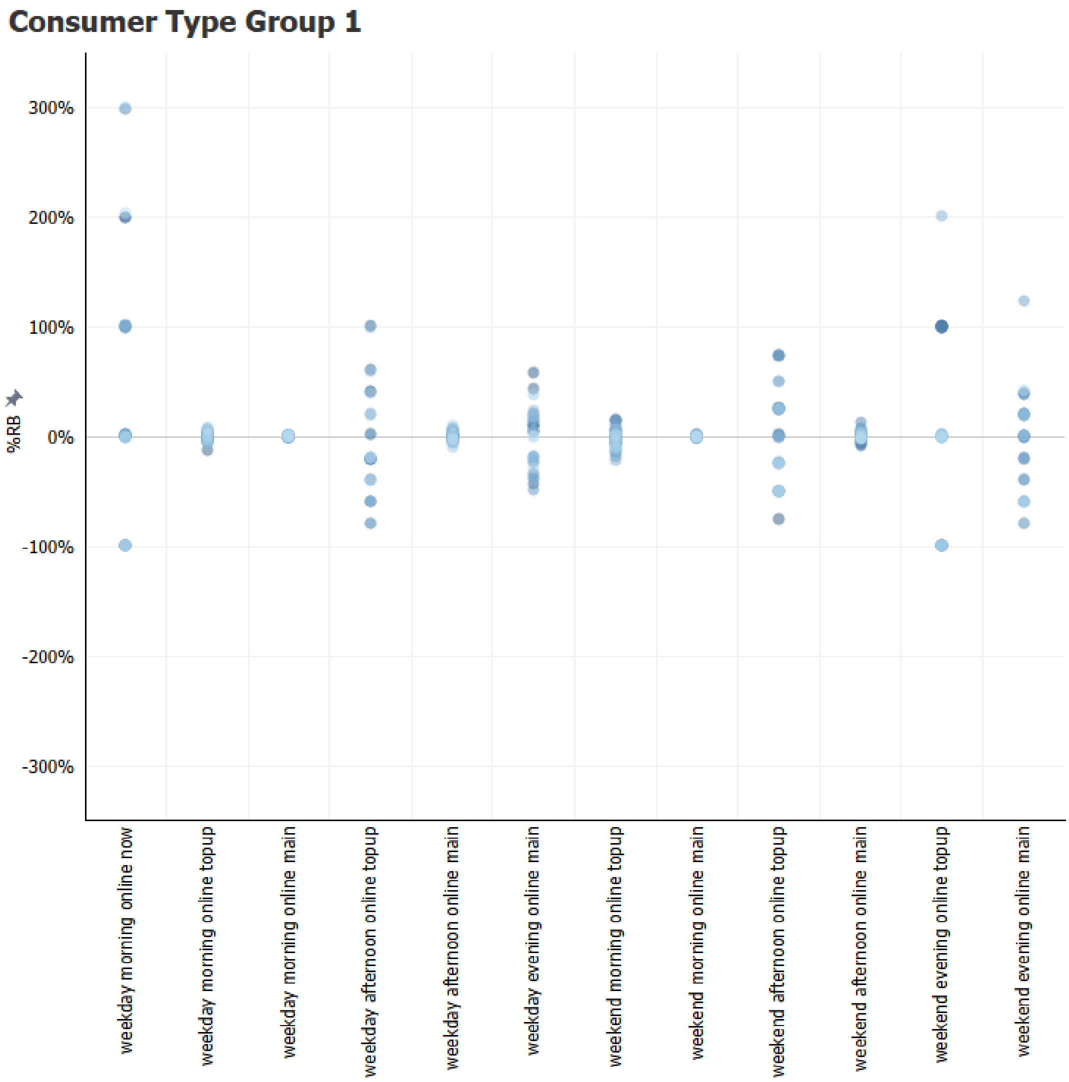



**Figure 6.5** Bland-Altman plot for Consumer Type Group 1. %RB for each transaction scenario across 30 model runs.

The 'weekday morning online now' and 'weekend evening online top-up' transactions scenarios have the highest variability, as demonstrated by the wider range of relative bias results. These results are likely due to the low number of transactions overall that occurred for these transaction scenario types in the observed transaction dataset. For example, 'weekday morning online now' contributed 0.0024% of all consumer type group 1's transactions. Although there was a low chance of this type of transaction occurring in the model, there were model iterations whereby 0.0097% of all transactions belonged to this transaction scenario. Similarly, some model runs had 0.0000% of all transactions originating from this transaction scenario. This is due to the inherent variability of the roulette wheel selection function and the low probability of this type of transaction occurring. The lower the probability of a transaction scenario occurring, the higher the variability. This variability was observed mainly in the 'top-up' and 'now' transactions scenarios in this consumer type group, as these customers solely shop online, and online shops are typically 'main' shops. However, although there is considerable variation and lower precision for these transaction scenarios from run to run, overall, the mean transactions (%) proportions are representative of consumer type 1 behaviours.

#### 6.2.1.2 Consumer type group 2

Consumer type group 2 is unique amongst the seven consumer type groups as these customers transact using both online and in-store channels. To analyse the model output of consumer type group 2, the simulated data were separated into online (Table 6.6) and in-store (Table 6.7) transaction scenarios. In total, this group contains 8,537 customers who predominantly shop in-store (~70% of the time) and occasionally perform 'weekday morning online main' (15.66%) and 'weekend morning online main' (15.76%) transactions. Their in-store behaviours are spread across a wider variety of transaction scenarios.

The proportional probabilities for the online mean transactions (%) closely resemble the observed data, and the 95% confidence intervals demonstrate a close range in which the true population mean falls within (Table 6.6). Overall, the model performed consistently for the online transaction scenarios across the 30 runs as, demonstrated by the low CV values ≤ 2.44% except for ‘weekday morning online for now’ (43.92%). Similar to consumer type group 1, ‘weekday morning online for now’ contributed 0.0012% of all transactions, with the model producing a maximum of 0.0041% and minimum of 0.0000% proportion of all transactions across the 30 model runs.

**Table 6.6** Consumer type group 2 model output summary (online).

| Consumer Type Group 2 (8,537 customers) (Online) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Online For Now | 0.00% | 0.00% | 43.92 | 0.000003 |
| Weekday Morning Online Top-up | 0.11% | 0.11% | 0.97 | 0.000031 |
| Weekday Morning Online Main | 15.66% | 15.63% | 1.05 | 0.000357 |
| Weekday Afternoon Online For Now | 0.00% | 0.00% | 1.71 | 0.000003 |
| Weekday Afternoon Online Top-up | 0.01% | 0.01% | 0.86 | 0.000006 |
| Weekday Afternoon Online Main | 0.95% | 0.94% | 1.11 | 0.000082 |
| Weekday Evening Online Main | 0.02% | 0.02% | 1.34 | 0.000012 |
| Weekend Morning Online For Now | 0.00% | 0.00% | 1.56 | 0.000010 |
| Weekend Morning Online Top-up | 0.11% | 0.11% | 1.43 | 0.000045 |
| Weekend Morning Online Main | 15.76% | 15.76% | 2.44 | 0.000366 |
| Weekend Afternoon Online For Now | 0.00% | 0.00% | 2.04 | 0.000004 |
| Weekend Afternoon Online Top-up | 0.01% | 0.01% | 1.89 | 0.000017 |
| Weekend Afternoon Online Main | 0.91% | 0.91% | 1.67 | 0.000137 |
| Weekend Evening Online Top-up | 0.00% | 0.01% | 1.55 | 0.000018 |
| Weekend Evening Online Main | 0.01% | 4.14% | 1.30 | 0.000147 |

For in-store transactions, the model could precisely and consistently produce simulated data representing the observed data. This is shown by the mean transactions (%) following the same pattern as the observed data transactions (%) and the low CV values of ≤ 5.93% across all transaction scenarios (Table 6.7).

**Table 6.7** Consumer type group 2 model output summary (in-store).

| Consumer Type Group 2 (8,537 customers) (In-store) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 4.13% | 3.36% | 1.17 | 0.000129 |
| Weekday Morning Instore Top-up | 3.37% | 2.07% | 3.52 | 0.000129 |
| Weekday Morning Instore Main | 2.07% | 7.63% | 3.64 | 0.000239 |
| Weekday Afternoon Instore For Now | 7.61% | 8.00% | 5.93 | 0.000323 |
| Weekday Afternoon Instore Top-up | 8.02% | 3.97% | 0.00 | 0.000194 |
| Weekday Afternoon Instore Main | 3.97% | 2.37% | 0.00 | 0.000135 |
| Weekday Evening Instore For Now | 2.37% | 2.51% | 0.00 | 0.000131 |
| Weekday Evening Instore Top-up | 2.50% | 1.16% | 0.00 | 0.000103 |
| Weekday Evening Instore Main | 1.16% | 2.73% | 0.00 | 0.000203 |
| Weekend Morning Instore For Now | 2.73% | 3.59% | 0.00 | 0.000247 |
| Weekend Morning Instore Top-up | 3.57% | 3.06% | 0.00 | 0.000186 |
| Weekend Morning Instore Main | 3.06% | 5.26% | 0.00 | 0.000296 |
| Weekend Afternoon Instore For Now | 5.24% | 7.92% | 0.00 | 0.000375 |
| Weekend Afternoon Instore Top-up | 7.93% | 5.66% | 0.00 | 0.000240 |
| Weekend Afternoon Instore Main | 5.66% | 1.30% | 0.00 | 0.000167 |
| Weekend Evening Instore For Now | 1.31% | 1.24% | 0.00 | 0.000164 |
| Weekend Evening Instore Top-up | 1.24% | 0.52% | 0.00 | 0.000113 |
| Weekend Evening Instore Main | 0.53% | 0.00% | 0.00 | 0.000000 |

A Bland-Altman plot was used to visualise the relative bias across the 30 model runs for both online and in-store transaction scenarios (Figure 6.6). All in-store transaction scenarios had a low RB ranging from -8.66% to +7.61%, demonstrating consistently accurate results for each transaction scenario over all 30 model runs. The online transactions scenarios have a greater degree of variability, with RB results as high as 501.49% for ‘weekday afternoon online now’ and as low as -100.00% for ‘weekday morning online now’. As iterated previously, these transaction scenarios contributed 0.0006% and 0.0012% of all consumer type group 2 transactions. Therefore, the variability observed with these types of transaction scenarios is expected as small fluctuations in the number of transactions will influence the RB. Overall, the variability does not impact the proportional mean transactions (%) results, demonstrating the model's ability to accurately reflect the observed data over multiple runs.

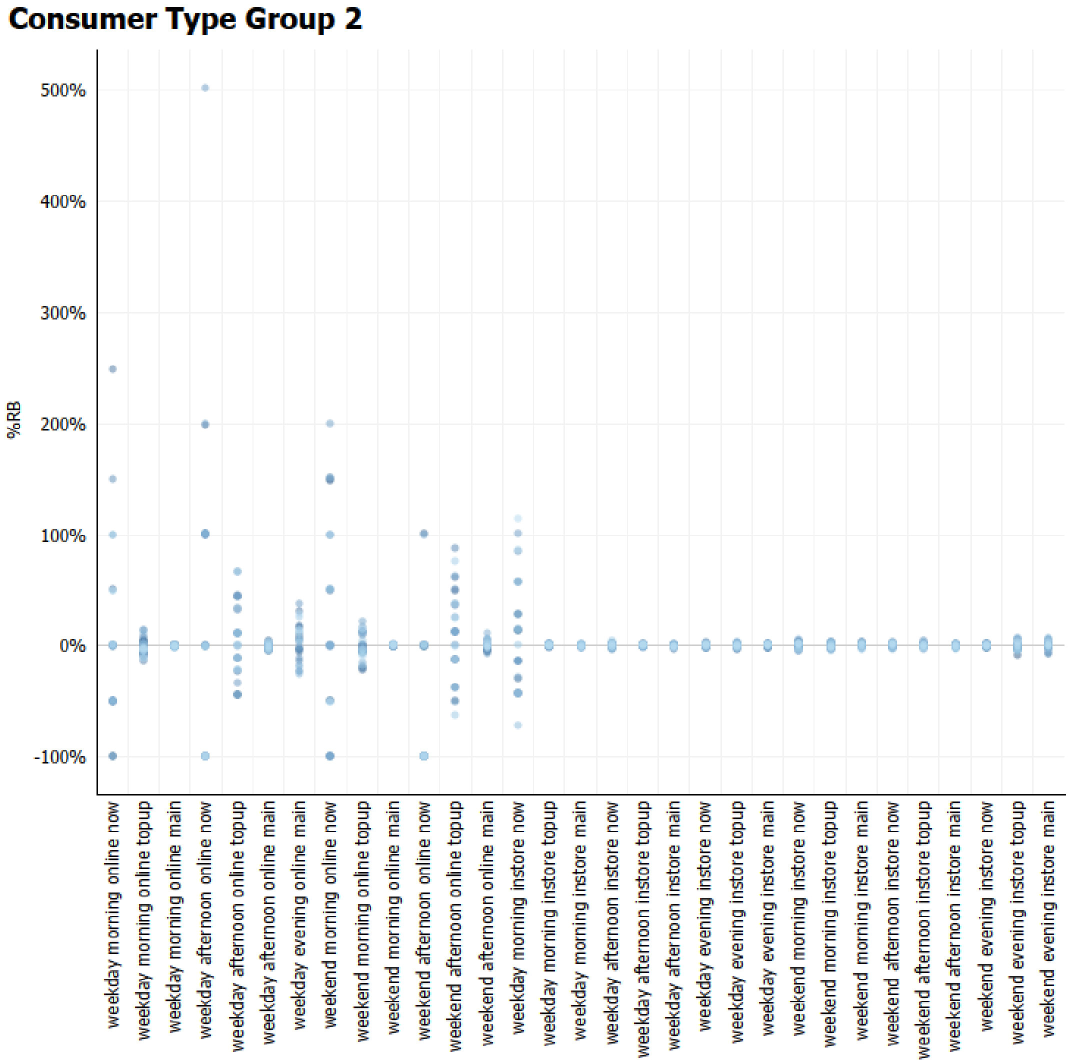



**Figure 6.6** Bland-Altman plot for Consumer Type Group 2. %RB for each transaction scenario across 30 model runs.

#### 6.2.1.3 Consumer type group 3

Consumer type group 3 has the highest number of customers (146,403) who shop online in-store. This consumer type group has a slight preference for 'afternoon' shops with 'weekday afternoon instore for now' (12.45%), 'weekday afternoon instore top-up' (13.38%), 'weekend afternoon instore for now' (10.20%) and 'weekend afternoon instore top-up' (12.92%) collectively constituting 48.95% of all transaction behaviours (Table 6.8). The mean transactions (%) output from model runs 1 to 30 closely match the observed dataset transaction scenario proportionality, with the above scenarios contributing 48.96% of all transactions in the model runs. As this consumer type group has more customers, the model output has more agents performing transactions, providing a more extensive dataset from each model run. The more data points produced in the model output, the higher the level of confidence in the results in terms of accuracy and precision. This is represented by the CV values of all transaction scenarios achieving ≤0.57%, thus demonstrating a high degree of precision and consistency across the 30 model runs.

**Table 6.8** Consumer type group 3 model output summary.

| Consumer Type Group 3 (146,403 customers) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 6.38% | 6.39% | 0.30 | 0.000071 |
| Weekday Morning Instore Top-up | 6.08% | 6.09% | 0.34 | 0.000076 |
| Weekday Morning Instore Main | 1.16% | 1.17% | 0.51 | 0.000022 |
| Weekday Afternoon Instore For Now | 12.45% | 12.46% | 0.18 | 0.000082 |
| Weekday Afternoon Instore Top-up | 13.38% | 13.38% | 0.18 | 0.000088 |
| Weekday Afternoon Instore Main | 3.20% | 3.20% | 0.38 | 0.000044 |
| Weekday Evening Instore For Now | 4.83% | 4.83% | 0.33 | 0.000058 |
| Weekday Evening Instore Top-up | 4.19% | 4.19% | 0.28 | 0.000043 |
| Weekday Evening Instore Main | 0.54% | 0.54% | 0.80 | 0.000016 |
| Weekend Morning Instore For Now | 5.63% | 5.63% | 0.44 | 0.000089 |
| Weekend Morning Instore Top-up | 6.41% | 6.40% | 0.36 | 0.000084 |
| Weekend Morning Instore Main | 3.35% | 3.35% | 0.53 | 0.000064 |
| Weekend Afternoon Instore For Now | 10.20% | 10.20% | 0.29 | 0.000107 |
| Weekend Afternoon Instore Top-up | 12.92% | 12.92% | 0.28 | 0.000131 |
| Weekend Afternoon Instore Main | 3.41% | 3.41% | 0.53 | 0.000065 |
| Weekend Evening Instore For Now | 3.00% | 3.00% | 0.57 | 0.000062 |
| Weekend Evening Instore Top-up | 2.38% | 2.39% | 0.54 | 0.000047 |
| Weekend Evening Instore Main | 0.47% | 0.47% | 1.42 | 0.000024 |

This consistency is further reflected in the Bland-Altman RB plot (Figure 6.7), whereby all transaction scenarios had RBs ranging from -2.73% to +2.66%, which is accounted for by a single transaction scenario, ‘weekend evening instore main’. The remaining transaction scenarios have an RB range of -2.08% to +1.65%. This consumer type group did not contain any extremely low probability transaction scenarios similar to those discussed for consumer type groups 1 and 2, which likely accounts for the high precision and accuracy of the consumer type group 3 model output data.

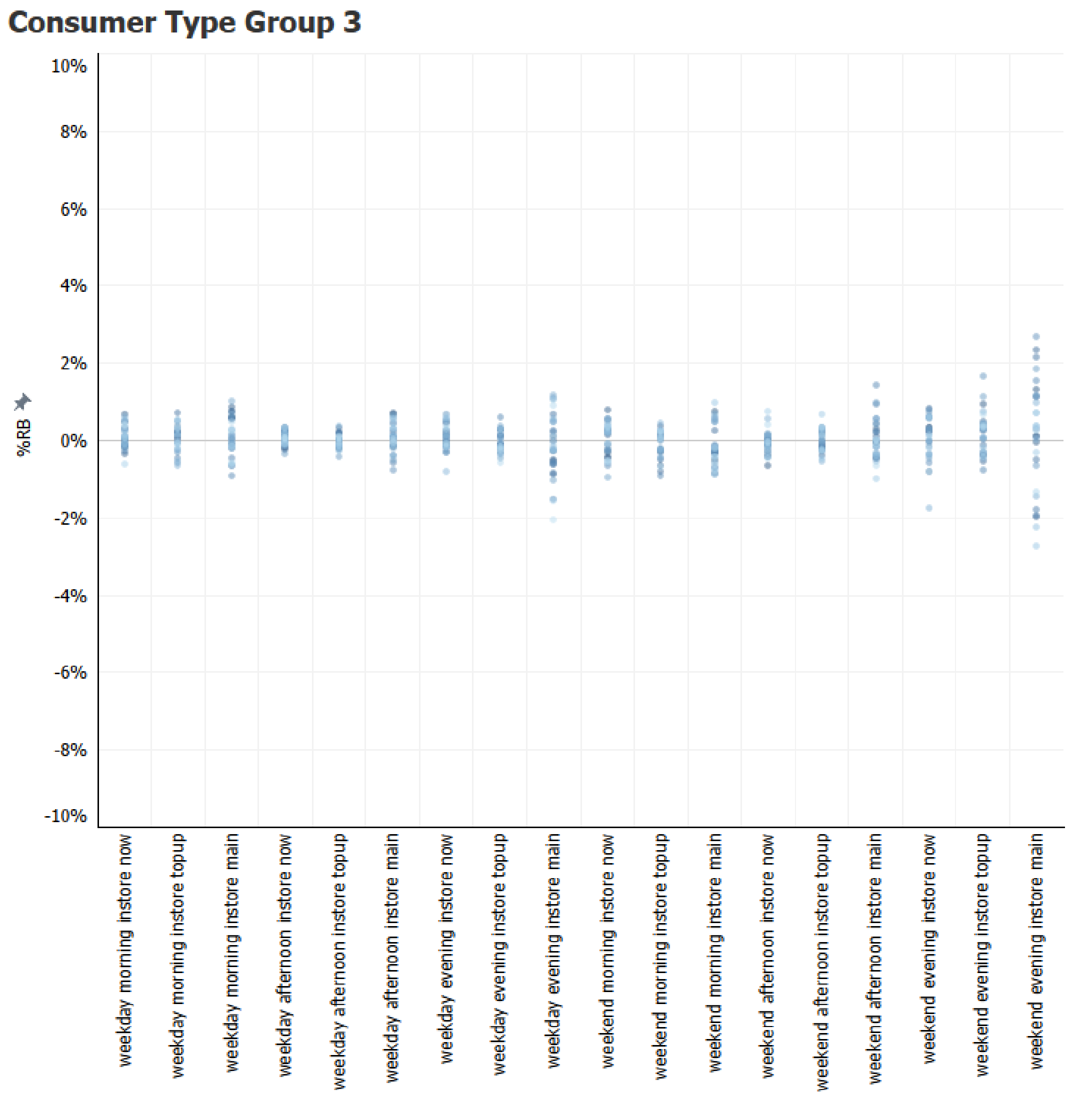



**Figure 6.7** Bland-Altman plot for Consumer Type Group 3. %RB for each transaction scenario across 30 model runs.

### 6.2.1.4 Consumer type group 4

Consumer type group 4 contains a smaller number of in-store customers (7,641) who have a slight preference for 'evening' and 'top-up' transactions as demonstrated by 'weekday evening instore top-up' (10.95%), 'weekday evening instore main' (15.07%) and 'weekday afternoon instore top-up' (11.34%) scenarios accounting for 37.36% of all transactions (Table 6.9). Similarly to consumer type group 3, this consumer type group does not include any extremely low probability transaction scenario types; therefore, high variability is not expected for any transaction scenario type.

The model runs 1-30 once again produced very similar mean transactions (%) to the observed data set, with ‘weekday evening instore top-up’ (10.94%), ‘weekday evening instore main’ (15.10%) and ‘weekday afternoon instore top-up’ (11.37%) scenarios making up 37.41% of all transactions. The CV values for each transaction scenario were low, with the highest CV value calculated for the ‘weekend evening instore for now’ (3.16%) transaction scenario (Table 6.9).

**Table 6.9** Consumer type group 4 model output summary.

| Consumer Type Group 4 (7,641 customers) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 2.34% | 2.34% | 1.65 | 0.000141 |
| Weekday Morning Instore Top-up | 2.33% | 2.33% | 1.21 | 0.000103 |
| Weekday Morning Instore Main | 1.35% | 1.34% | 2.12 | 0.000103 |
| Weekday Afternoon Instore For Now | 5.32% | 5.33% | 0.84 | 0.000164 |
| Weekday Afternoon Instore Top-up | 8.68% | 8.65% | 0.74 | 0.000234 |
| Weekday Afternoon Instore Main | 6.99% | 6.98% | 0.96 | 0.000244 |
| Weekday Evening Instore For Now | 4.33% | 4.31% | 1.33 | 0.000209 |
| Weekday Evening Instore Top-up | 10.95% | 10.94% | 0.82 | 0.000326 |
| Weekday Evening Instore Main | 15.07% | 15.10% | 0.64 | 0.000353 |
| Weekend Morning Instore For Now | 2.13% | 2.12% | 2.77 | 0.000214 |
| Weekend Morning Instore Top-up | 3.42% | 3.42% | 1.97 | 0.000246 |
| Weekend Morning Instore Main | 4.29% | 4.29% | 1.71 | 0.000267 |
| Weekend Afternoon Instore For Now | 5.11% | 5.09% | 1.31 | 0.000243 |
| Weekend Afternoon Instore Top-up | 11.34% | 11.37% | 0.95 | 0.000394 |
| Weekend Afternoon Instore Main | 8.58% | 8.61% | 1.10 | 0.000344 |
| Weekend Evening Instore For Now | 1.64% | 1.62% | 3.16 | 0.000187 |
| Weekend Evening Instore Top-up | 3.41% | 3.43% | 1.80 | 0.000225 |
| Weekend Evening Instore Main | 2.72% | 2.71% | 2.42 | 0.000238 |

The Bland-Altman RB plot for consumer type group 4 (Figure 6.8) demonstrates a high level of accuracy for each transaction scenario over model runs 1-30, with the RB ranging from -7.02% to +7.24%. The slight variance captured within the model is likely influenced by the smaller number of customers who belong to this consumer type group. The less agents there are in a consumer type group, the more variable the results will be, as statistical power is proportional to the number of results. Nevertheless, the results demonstrate that the model can capture some variance in consumer type 4 behaviours whilst retaining the ability to reproduce the observed dataset transactional behaviours.

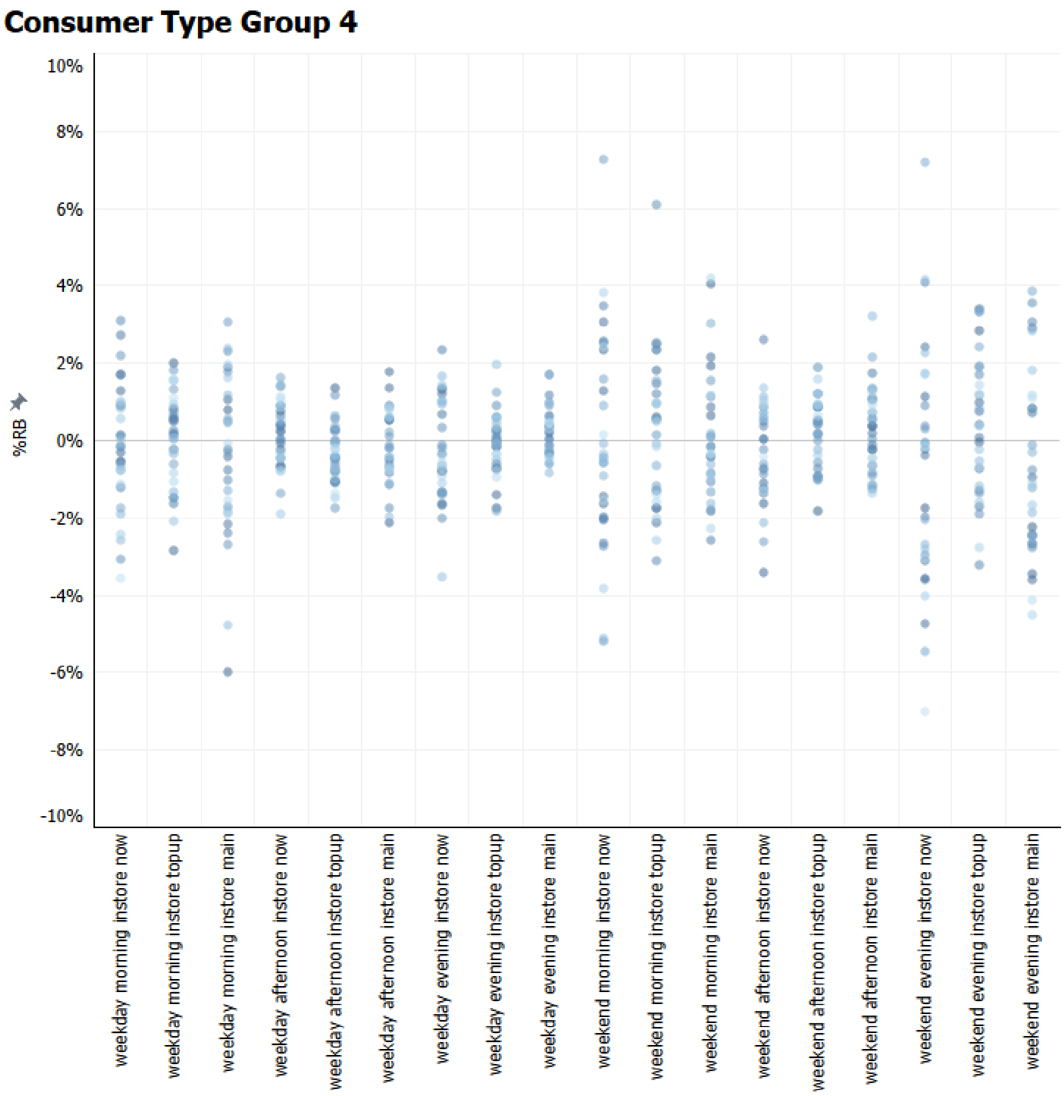



**Figure 6.8** Bland-Altman plot for Consumer Type Group 4. %RB for each transaction scenario across 30 model runs.

### 6.2.1.5 Consumer type group 5

Similar in size to consumer type group 4, consumer type group 5 contains 7,479 customers. This consumer type group has a propensity for completing 'weekday afternoon instore main' transactions, making up 35.95% of all transaction scenarios in the observed data (Table 6.10). Once again, the model was able to accurately simulate consumer transaction behaviours as shown by the mean transactions (%) results from model runs 1 to 30. The proportional transactions (%) from each model run consistently simulated the observed data, as shown by the high degree of precision across all transaction scenarios (CV ≤ 3.18%) (Table 6.10).

**Table 6.10** Consumer type group 5 model output summary.

| Consumer Type Group 5 (7,479 customers) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 2.28% | 2.28% | 1.03 | 0.000085 |
| Weekday Morning Instore Top-up | 2.30% | 2.30% | 1.42 | 0.000119 |
| Weekday Morning Instore Main | 1.24% | 1.23% | 1.37 | 0.000061 |
| Weekday Afternoon Instore For Now | 5.14% | 5.15% | 0.80 | 0.000150 |
| Weekday Afternoon Instore Top-up | 8.99% | 8.99% | 0.60 | 0.000197 |
| Weekday Afternoon Instore Main | 5.98% | 5.97% | 0.92 | 0.000200 |
| Weekday Evening Instore For Now | 1.94% | 1.94% | 1.48 | 0.000104 |
| Weekday Evening Instore Top-up | 3.22% | 3.22% | 1.22 | 0.000143 |
| Weekday Evening Instore Main | 1.27% | 1.26% | 1.39 | 0.000064 |
| Weekend Morning Instore For Now | 1.47% | 1.48% | 3.00 | 0.000161 |
| Weekend Morning Instore Top-up | 2.95% | 2.93% | 2.33 | 0.000248 |
| Weekend Morning Instore Main | 6.83% | 6.81% | 1.23 | 0.000306 |
| Weekend Afternoon Instore For Now | 4.45% | 4.45% | 1.26 | 0.000205 |
| Weekend Afternoon Instore Top-up | 12.38% | 12.39% | 0.78 | 0.000353 |
| Weekend Afternoon Instore Main | 35.95% | 35.98% | 0.31 | 0.000401 |
| Weekend Evening Instore For Now | 0.74% | 0.73% | 3.18 | 0.000085 |
| Weekend Evening Instore Top-up | 1.24% | 1.24% | 2.97 | 0.000134 |
| Weekend Evening Instore Main | 1.62% | 1.63% | 2.26 | 0.000134 |

The RB range for consumer type group 5 was -9.24% to +10.13% (Figure 6.9). These values originated from the ‘weekend evening instore top-up’ transaction scenario, which also had one of the lowest proportions of transactions for this consumer type group (1.24%). The lowest probability transactions scenario, ‘weekend evening in-store for now’ (0.74%), also had higher RB than other transaction scenarios and ranged from -8.4% to +8.9%. As discussed previously, lower accuracy and higher RB values are expected for transaction scenarios with a low probability chance of occurring as small fluctuations in ‘agent’ behaviour based on the weighted roulette selection function will result in lower precision.

The higher RB values obtained for consumer type group 5 resulted from the lower proportional spread of probabilities over the transaction scenarios. Unlike consumer type group 4 where there were multiple transaction scenarios making up ~10-15% of all transactions each, consumer type group 5 had one transaction scenario comprising 35.95% of all transactions. This resulted in the majority of transaction scenarios accounting for ~2% of all transactions (Table 6.10). However, this did not hinder the model’s ability to simulate the observed data accurately and precisely.

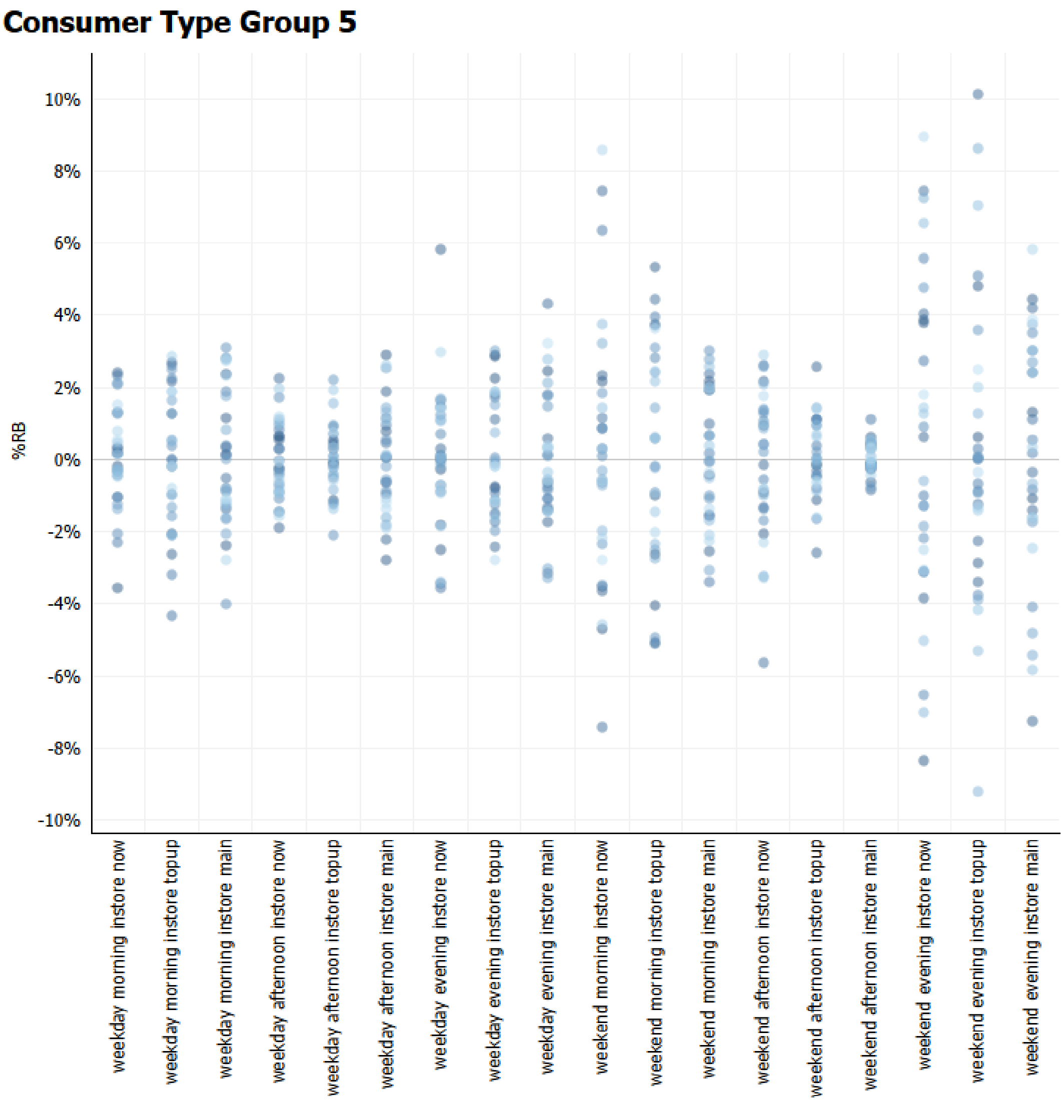



**Figure 6.9** Bland-Altman plot for Consumer Type Group 5. %RB for each transaction scenario across 30 model runs.

### 6.2.1.6 Consumer type group 6

Slightly larger than the previous two consumer type groups, consumer type group 6 contains 13,969 in-store customers. This consumer type group performed 'mean' shops as the main transaction scenario as depicted by the 'weekday morning instore main' (19.98%) and 'weekday afternoon instore main' (18.67%) transaction scenarios totalling almost 40% of all transactions (38.65%). Other notable transaction scenarios were 'weekday morning instore top-up' (9.6%) and 'weekday afternoon instore top-up' (10.11%), which contributed another 19.71% of all transactions (Table 6.11).

The mean transactions (%) for all transaction scenarios accurately reflected the observed data, with the proportions of the above four scenarios comprising (on average) 58.47% of all transactions compared to 58.45% in the observed data. The precision followed a pattern of proportionality, with the lowest probability transaction scenarios resulting in the highest CV values; 'weekend evening instore for now' (proportion: 0.42%; CV 3.46%), 'weekend evening instore top-up' (proportion: 0.56%; CV 4.32%) and 'weekend evening instore main' (proportion: 0.26%; CV 5.90%). The remaining transaction scenarios had CV values ≤2.35%.

**Table 6.11** Consumer type group 6 model output summary.

| Consumer Type Group 6 (13,969 customers) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 4.32% | 4.33% | 0.79 | 0.000124 |
| Weekday Morning Instore Top-up | 9.60% | 9.60% | 0.56 | 0.000195 |
| Weekday Morning Instore Main | 19.98% | 19.98% | 0.45 | 0.000327 |
| Weekday Afternoon Instore For Now | 4.96% | 4.94% | 0.80 | 0.000145 |
| Weekday Afternoon Instore Top-up | 10.11% | 10.11% | 0.59 | 0.000218 |
| Weekday Afternoon Instore Main | 18.76% | 18.78% | 0.37 | 0.000250 |
| Weekday Evening Instore For Now | 0.77% | 0.77% | 1.85 | 0.000052 |
| Weekday Evening Instore Top-up | 1.02% | 1.02% | 1.63 | 0.000060 |
| Weekday Evening Instore Main | 0.66% | 0.66% | 2.35 | 0.000056 |
| Weekend Morning Instore For Now | 3.49% | 3.49% | 1.56 | 0.000199 |
| Weekend Morning Instore Top-up | 5.43% | 5.42% | 1.14 | 0.000225 |
| Weekend Morning Instore Main | 5.98% | 5.98% | 0.96 | 0.000209 |
| Weekend Afternoon Instore For Now | 3.35% | 3.34% | 1.18 | 0.000144 |
| Weekend Afternoon Instore Top-up | 5.85% | 5.85% | 1.20 | 0.000255 |
| Weekend Afternoon Instore Main | 4.47% | 4.48% | 1.40 | 0.000227 |
| Weekend Evening Instore For Now | 0.42% | 0.42% | 3.26 | 0.000050 |
| Weekend Evening Instore Top-up | 0.56% | 0.56% | 4.32 | 0.000089 |
| Weekend Evening Instore Main | 0.26% | 0.27% | 5.90 | 0.000057 |

The pattern of proportionality was also reflected in the RB variance (Figure 6.10) with the ‘weekend evening instore’-’for now’, -’top-up’ and –’main’ with RB ranges of -5.2% to +5.7%, -7.8% to +9.2% and -10.7% to 15.9%, respectively. Although slightly elevated, these RB and CV results are not sufficient to render the model incapable of simulating consumer type 5 transaction behaviour. In fact, it is an advantage of the model to have the capacity to incorporate variance in consumer behaviour whilst still accurately simulating real customer behaviours.

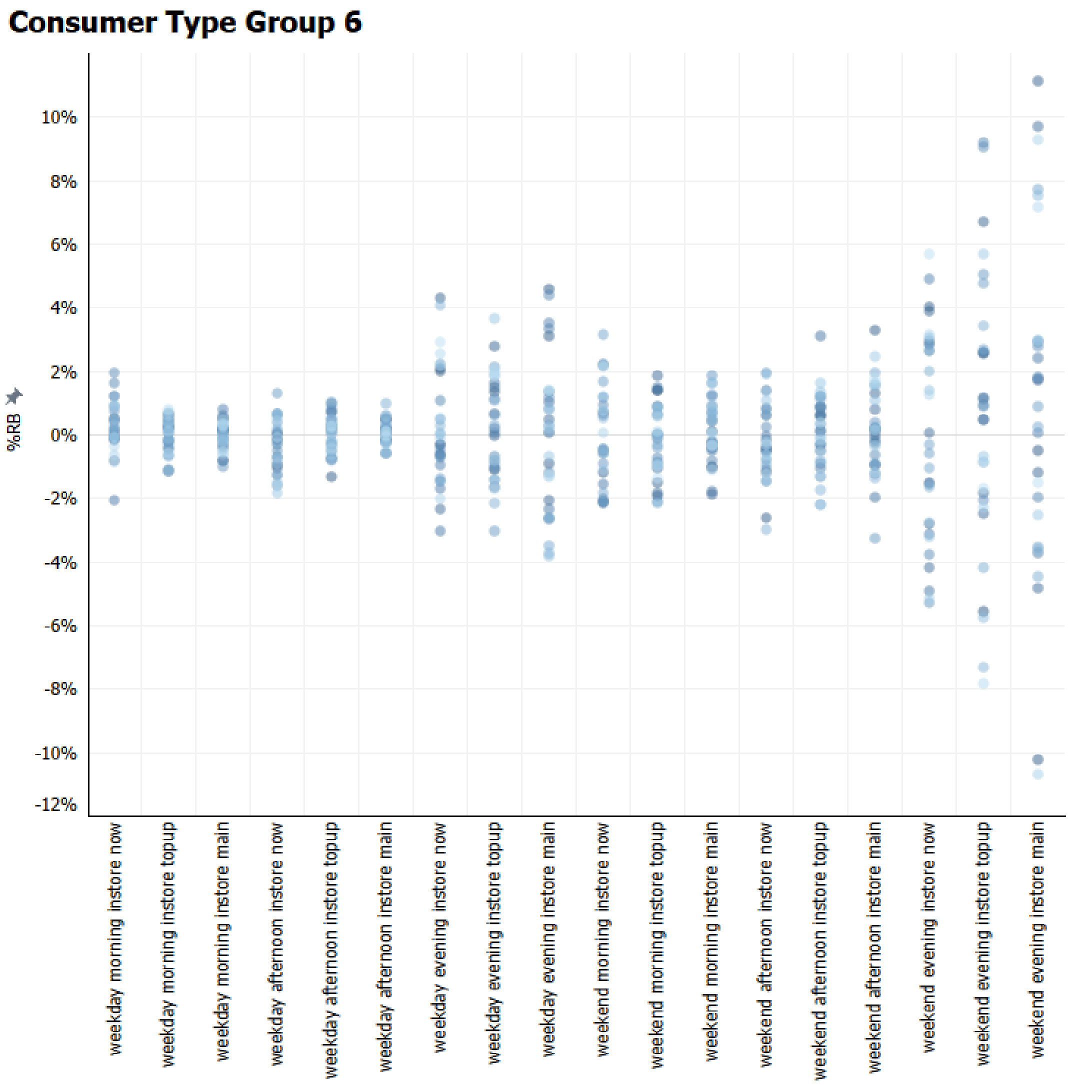



**Figure 6.10** Bland-Altman plot for Consumer Type Group 6. %RB for each transaction scenario across 30 model runs.

### 6.2.1.7 Consumer type group 7

Consumer type group 7 comprises the second highest number of customers (28,618) who proportionally perform higher numbers of 'for now' and 'top-up' transactions in comparison to 'main' shops. In this consumer type group, only 7.16% of all transactions are 'main shops'. The highest proportion of transaction scenarios are 'weekday morning instore for now' (15.10%), 'weekday afternoon instore for now' (21.69%), 'weekday afternoon instore top-up' (9.61%) and 'weekend afternoon instore for now' (12.07%) which constitute more than 50% of all transactions (58.47%) (Table 6.12).

The model runs 1 – 30 achieved a high degree of precision in the transaction (%) results with only 'weekend evening instore main' having a CV value > 2.00% (3.84%). This suggested that the model was not only able to simulate an overall mean transaction (%) that was representative of the customers in consumer type group 7 but that it was able to do so consistently and precisely.

**Table 6.12** Consumer type group 7 model output summary.

| Consumer Type Group 7 (28,619 customers) | | | | |
|---|---|---|---|---|
| Transaction Scenario | Observed Data Transactions (%) | Model runs 1 to 30 | | |
| | | Mean Transactions (%) | CV (%) | Confidence Interval (95%) |
| Weekday Morning Instore For Now | 15.10% | 15.10% | 0.30 | 0.000163 |
| Weekday Morning Instore Top-up | 3.80% | 3.80% | 0.85 | 0.000117 |
| Weekday Morning Instore Main | 0.66% | 0.65% | 1.35 | 0.000032 |
| Weekday Afternoon Instore For Now | 21.69% | 21.70% | 0.26 | 0.000206 |
| Weekday Afternoon Instore Top-up | 9.61% | 9.60% | 0.43 | 0.000149 |
| Weekday Afternoon Instore Main | 2.11% | 2.10% | 0.94 | 0.000072 |
| Weekday Evening Instore For Now | 5.27% | 5.27% | 0.58 | 0.000112 |
| Weekday Evening Instore Top-up | 2.66% | 2.65% | 0.76 | 0.000074 |
| Weekday Evening Instore Main | 0.41% | 0.41% | 1.55 | 0.000023 |
| Weekend Morning Instore For Now | 7.34% | 7.36% | 0.85 | 0.000226 |
| Weekend Morning Instore Top-up | 3.25% | 3.26% | 1.34 | 0.000159 |
| Weekend Morning Instore Main | 1.45% | 1.45% | 1.24 | 0.000065 |
| Weekend Afternoon Instore For Now | 12.07% | 12.06% | 0.52 | 0.000228 |
| Weekend Afternoon Instore Top-up | 7.80% | 7.80% | 0.59 | 0.000169 |
| Weekend Afternoon Instore Main | 2.22% | 2.21% | 1.19 | 0.000095 |
| Weekend Evening Instore For Now | 2.80% | 2.80% | 1.05 | 0.000107 |
| Weekend Evening Instore Top-up | 1.46% | 1.45% | 1.46 | 0.000077 |
| Weekend Evening Instore Main | 0.31% | 0.31% | 3.84 | 0.000044 |

The higher CV value for ‘weekend evening instore main’ is attributable to the low probability chance (0.31%) of this transaction scenario occurring. The maximum transaction (%) result for this transaction scenario in the model runs was 0.33%, and the lowest was 0.29%. As these numbers are extremely low, small fluctuations in transaction numbers will reduce the precision. However, even though the proportionality of this transaction scenario varied and resulted in the highest CV value, the CV was still very low, indicating a high degree of precision and consistency.

This variability can be observed in the RB Bland-Altman plot (Figure 6.11), where it is apparent that most variance originates from the ‘weekend evening instore main’ transaction scenario (RB range: -7.3% to +6.1%). The other transaction scenarios had an RB range of -4.0 to +3.0%. As with all consumer type groups, the model has demonstrated the capability of integrating inherent variability in consumer type behaviour whilst conserving the overall proportionality of each transaction scenario.

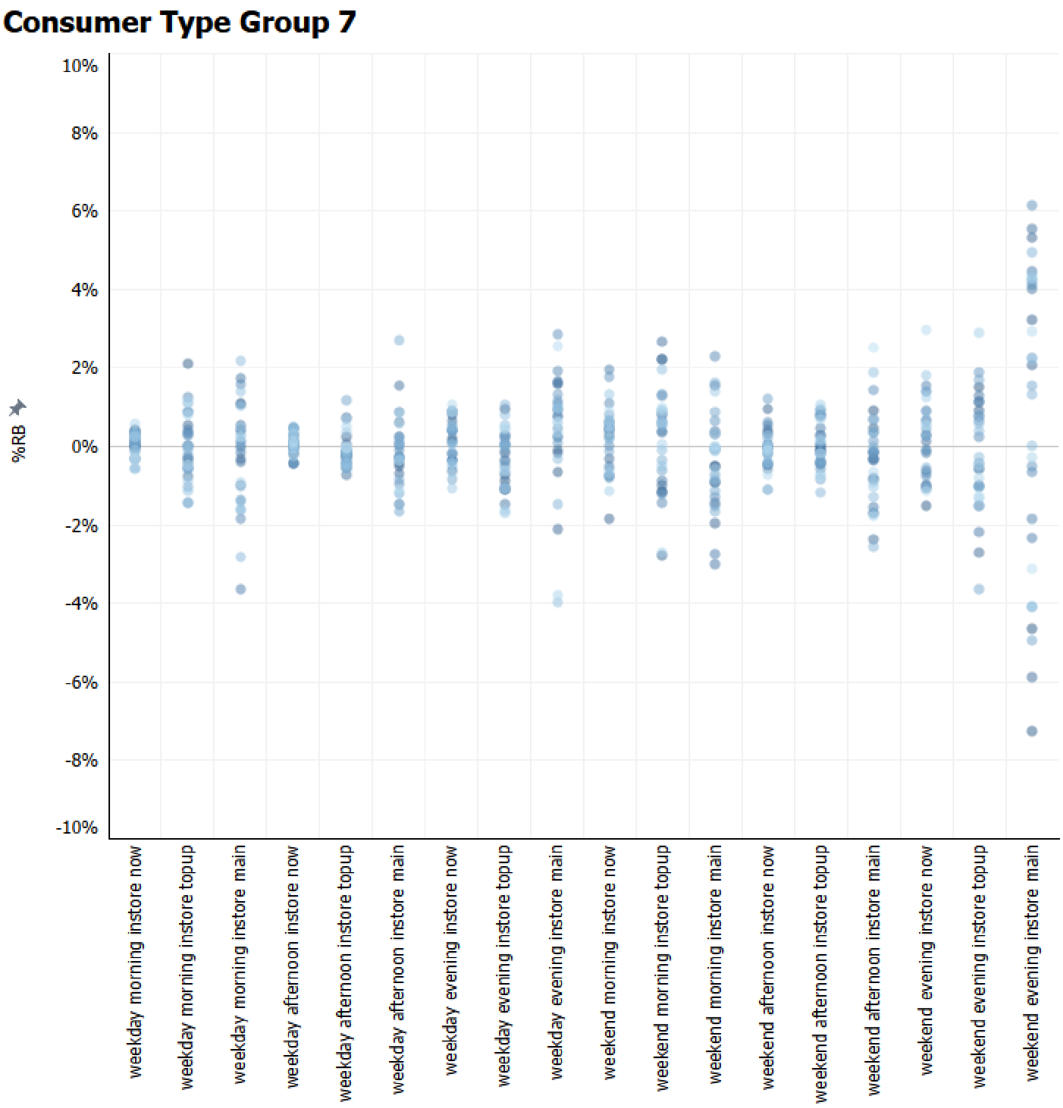



**Figure 6.11** Bland-Altman plot for Consumer Type Group 7. %RB for each transaction scenario across 30 model runs.

## 6.2.2 Model validation summary

The data analysis presented in this section has been heavily driven by the nature of the dataset provided by Sainsbury's and the modelling approach used. The validation approach was specifically chosen to compare the observed data to the simulated output, ensuring that the model performed was representative of the Sainsbury's customers. If other individual-based data was available besides the Sainsbury's dataset, more detailed and varied validation techniques could be performed to further validate this model. However, as these data are inaccessible, the model validation is restricted to assessing the accuracy and precision of the model outputs compared to the observed dataset.

By using RB (%), the analysis has shown that over 30 model runs, the model can provide reproducible data accurately and consistently for consumer type groups 3 to 7. Model runs for consumer type group 1 and 2 produced the most variable results, largely due to the spread of transaction proportionality for online transaction scenarios. For example, 94.19% of consumer type group 1's transactions originated from two transaction scenarios ('Weekend Morning Online Main' and 'Weekday Morning Online Main'). This meant that the other transaction scenarios had a low chance of occurring; thus, any fluctuations in transaction frequency resulted in high variability across model runs. A similar situation was observed with consumer type group, 2 whereby 31.42% of transactions were from the 'Weekend Morning Online Main' and 'Weekday Morning Online Main' transaction scenarios. The variability in the model output was also detected when employing a precision assessment (CV (%)) to assess the closeness of the results across model runs 1 to 30. The online transaction scenarios had the lowest precision for consumer type groups 1 and 2, and this was accounted for by the transaction proportionality as discussed above. Consumer type groups 3 to 7 demonstrated a high level of precision across all transaction scenarios.

Therefore, the model validation has confirmed the applicability and robustness of the IBM and decision tree approach in simulating in-store consumer type behaviours. However, although variability was present in the online transaction scenarios, the model output average transactions for each transaction scenario for all consumer type groups were accurate and precise. Therefore, the variability observed is more an artefact of agents making transaction decisions rather than the model failing to simulate consumer behaviours.

## 6.3 Modelling consumer transactions spatially

After validation of the model in section 6.2, the model confidently simulates the temporal aspect of consumer behaviour, including the relationships between day type, time of day, channel, and basket type. Next, the model needed to consider which store the customer should transact at based on the transaction scenario being made. The following sections begin with a discussion on the importance of modelling the spatial aspect of consumer behaviour, including spatial-based demand and the challenges of modelling these behaviours at the individual level (6.3.1). Next, section 6.3.2 demonstrates how spatiality was initially incorporated into the model using naïve behavioural rules, following Tobler's first law of geography. After analysing the results of naïve behavioural rules, the study considered a second, consumer-based approach in which the actual distances of observed customers were implemented into the model (section 6.3.3); this second method utilises distance bins implemented as an additional, secondary decision tree, in which distance bins were created that relate to transactional temporality, mission, and basket type. Section 6.3.5 then discusses further steps to be taken to incorporate various layers of grocery demand, beginning with work-based demand using census data, whereby customers are assigned both a home location and workplace location. Finally, section 6.4 closes this modelling chapter, discussing the complexities of modelling the behaviours of individual customers across space.

### 6.3.1 The importance of modelling spatiality

Understanding the complexities of consumer purchasing behaviour can be challenging. Multiple factors are considered when making purchasing decisions, ranging from personal preferences, brand loyalty, product pricing, channel choices, store location accessibility, and these factors are all interlinked with temporality. This thesis focuses on the temporal and spatial aspects of consumer behaviour in the context of grocery retail, including channel choice and the relationships to basket types. This study identified key indicators of consumer behaviour specifically for grocery transacting in section 2.3, and many of these indicators are linked to spatiality.

Those in academia and industry researching and working within the field of retail location analytics are aware of the importance of location (Chapter 3) and have continued to develop the tools and methods used to consider all factors of spatial demand. At face value, purchasing groceries is a task that is often to be done near home, online, or at least close to home. The habitual weekly shop is a task that most people do, whether it is performed online, in-store, or using click and collect (section 2.2). However, consumer behaviours are much more versatile, with convenience playing a more important role in our lives (section 2.2). Therefore, consumer grocery demand goes beyond the base of residence-based demand and into all facets of life.

As demonstrated by Bell (2015) and (Martin et al. (2015), the population fluctuates throughout the 24-hour day, with distinct areas becoming densely populated, notably during work hours, thus creating an entirely new demand for groceries outside of residential demand (Berry et al., 2016). Others have identified additional notable areas of demand that fluctuate temporally, including school-based and university demand (Waddington et al., 2019), leisure and tourism demand (Newing et al., 2015; Newing et al., 2018), and convenience demand in general (Wood and Browne, 2007;

Hood et al., 2016). As individuals relocate over the day, grocery shopping becomes an opportunity for various reasons, whether it is purchasing food as part of a multi-purpose trip, buying lunch on a journey, or visiting a place of interest and picking up a bottle of water for the day.

Waddington et al. (2018) demonstrate a sound attempt at incorporating these different sources of demand into a top-down SIM, in which stores were clustered, and demand catchments were created using demand layers, focusing on residential, workplace, school, and leisure-based demand. However, recreating these demand layers at the individual level is incredibly complex, especially in this study, as further information is needed about the customers being modelled. As the study focuses solely on modelling the known behaviours of those identified in the loyalty card-linked transaction dataset and their interactions with Sainsbury's stores, their general grocery behaviours outside Sainsbury's may vary. It is not known whether these customers work from home, have school children, are in university, or when they purchase non-locally as part of a multi-purpose trip. Inferences can be made, but before modelling what we do not know, we must determine how to model what is known.

The loyalty card-linked transaction dataset by the study collaborator provides two key data; customers' home OA and the stores they transacted at. The following store choice methodologies assign each observed customer a home location in the models using OA PWC and loads in store information directly from the Sainsbury's store table. For comparability between model output and observed transactional data, the same observed customers are allocated a home in the model. However, it must be noted that the observed Sainsbury's data calculates home-to-store distances using their in-house GIS that uses road network data. In contrast, the model uses geodesic distances, which will cause disparities between the two to an extent. As a first attempt at incorporating spatiality into the IBM, section 6.3.2 presents the first step in adding spatiality into the temporal model using simplistic behavioural rules.

## 6.3.2 Store choice method 1: Tobler's Law

For the first implementation of spatiality into the temporal model, only residential demand is captured, keeping in line with the model's philosophy of "keep it simple, stupid". This method is based on Tobler's first law of geography (Tobler's law), a naïve approach to modelling demand. Tobler's law states that "everything is related to everything else, but near things are more related than distance things" (Tobler, 1987); therefore, customers are assigned their closest store for all transactions, depending upon the chosen channel and basket type. Figure 6.12 provides a section of pseudo-code to portray the logic of the model. After a customer has been entered through the decision tree process for a transaction scenario, resulting in a time, channel, and basket, the model decides which store the transaction should take place. Essentially, any 'for now' and 'in-store' transaction scenarios will take place at the customer's closest store, regardless of whether the store is a convenience store or supermarket. If a transaction is 'top-up' or 'main' and 'in-store', then the transaction will be assigned to the customer's closest supermarket only. If a transaction scenario has the channel 'online', the transaction will occur at the customer's closest supermarket that fulfils online orders.

**Figure 6.12** Pseudo-code of store choice rules for store choice method 1

```
if transaction basket type = 'for now' AND channel = 'in-store'
        then transact at ANY closest store
    else if transaction basket type = 'top-up' or 'main' AND channel = 'in-store'
        then transact at closest store type = supermarket
    else if transaction channel = 'online'
        then transact at closest store type= supermarket AND online fulfilment = TRUE
endif;
```

Once the customer's transaction scenario is determined, the model calculates the geodesic distances in kilometres between the customer's assigned home location and every Sainsbury's store in the model using coordinates. During the model's initialisation stage, a distance matrix is created between every OA location in West Yorkshire and all stores within the model environment. For each customer, all stores are ranked in terms of closest distance and are split by convenience stores, supermarkets, and supermarkets that provide online ordering. Sainsbury's stores in the model are either classified as 'supermarkets' or 'convenience' stores.

Sainsbury's supermarkets and other supermarket stores tend to have a wider variety of foods available in these stores, catering for all basket types. Convenience stores, however, have a much-limited variety of foods available and are often confined to 'essential' products that are either frequently purchased during the week or ready-to-eat food and drink. With these factors considered, the model assigns a customer to a store based on the transaction's basket type, channel, and the customer's distance to the stores.

The simple logic of this model ensures that all customers are transacting at the store closest to their home location under different scenarios. This implementation of Tobler's law is likely to accurately capture the observed Sainsbury's customer transactions that occur residentially but not transactions that are considered non-residential demand. The reasoning for attributing specific basket types to store choice is related to the provision of different goods at different store types. Convenience stores tend to provide a limited product range compared to supermarkets. As the data presented in section 4.2.5 identified, a strong relationship exists between 'main' and 'top-up' transactions at supermarket stores, which generally occur further away from the customer's home location. Customers who transact online do not have a distance to travel themselves but are still assigned a store from which their order would have originated. Their transactions are assigned to their closest store that provides online delivery (at least one supermarket in each LAD provides online delivery in the case study) (Newing et al., 2022; Urquhart et al., 2022).

Using Tobler's simplistic law, the assignment of *where* the transactions occur is achieved using agent-based modelling techniques, allowing for the simulation of spatiotemporal shopping behaviours at the individual level. The following section provides the results of the model using Tobler's law and naïve behavioural rules.

#### 6.3.2.1 Store choice method 1: Tobler's law results

As explained previously, this first attempt at incorporating spatiality may have limited scope in representing the observed transactional data, as naïve store choice rules were implemented. As the model's logic uses the closest store rule, it is expected that most transactions in the model occur within the same local authority district as the customer or a surrounding district edge. Figure 6.13 presents the cross-district interactions between customer and store locations for in-store transactions only. The analysis finds that, similar to the observed dataset provided by Sainsbury's, customers mainly transact within their district (Figure 4.6). However, cross-district relationships are not identified to the same extent, as customers are not relocating in the model, which would impact location-based demand. The cross-district transactions presented in Figure 6.13 are due to customers residing on the edge of districts or far from a Sainsbury's store within their district. As mentioned in section 4.2.5, Bradford has a limited number of Sainsbury's stores; therefore, customers will have reduced store choices and be assigned a store likely in Calderdale. Predictably, the model does not capture workplace demand as it has not been accounted for or any demand beyond residential-based transacting. Nevertheless, this analysis has shown that the model can simulate individual consumer behaviour, linking channel, basket type, temporality and spatiality to store choice.

**Store choice method 1: Tobler's law - Proportion of in-store transactions between customer and store LADS (%)**

| | Customer LAD | | | | |
|---|---|---|---|---|---|
| **Store LAD** | **Bradford** | **Calderdale** | **Kirklees** | **Leeds** | **Wakefield** |
| **Bradford** | 90.7% | 0.1% | 0.1% | 3.6% | |
| **Calderdale** | 7.7% | 91.3% | 3.1% | | |
| **Kirklees** | 0.1% | 8.6% | 96.4% | 0.3% | 3.4% |
| **Leeds** | 1.5% | | | 94.8% | 1.0% |
| **Wakefield** | | | 0.4% | 1.3% | 95.6% |
| **Grand Total** | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**% of total transactions**
0.1% – 96.4%

**Figure 6.13** Store choice method 1: Tobler's law. Proportion of in-store transactions by customers in each Local Authority District into stores in each Local Authority District (%).

For more context, Figure 6.14 compares each consumer type group's modelled behaviour regarding the average distance (km) travelled and basket type. Despite each transaction being simulated to occur locally, *some* element of distance variation has been simulated. Overall, the model under-predicts the distances the consumers travelled compared to the observed data (Figure 5.4). The minimum average distance travelled in the simulated output is ~1.4km, whereas the observed data is ~2.3km. This could be due to two factors; firstly, the model uses geodesic distances (straight-line distance with a curvature due to the Earth's surface), and the observed data used road network data generated by Sainsbury's in-house GIS. Secondly, as all customers travel locally, the data is likely skewed, bringing the average distance down.

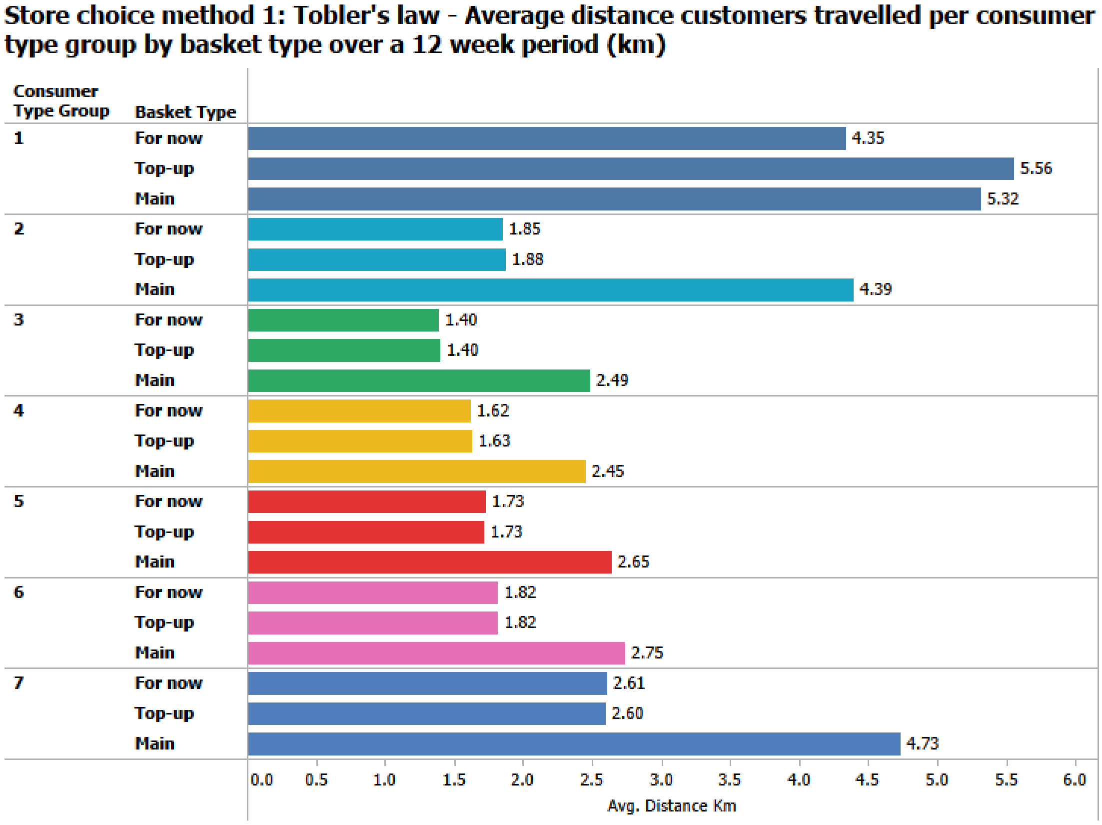



**Figure 6.14** Store choice method 1: Tobler's law. Average distance travelled by basket type for each consumer type group over the 12-week period (km).

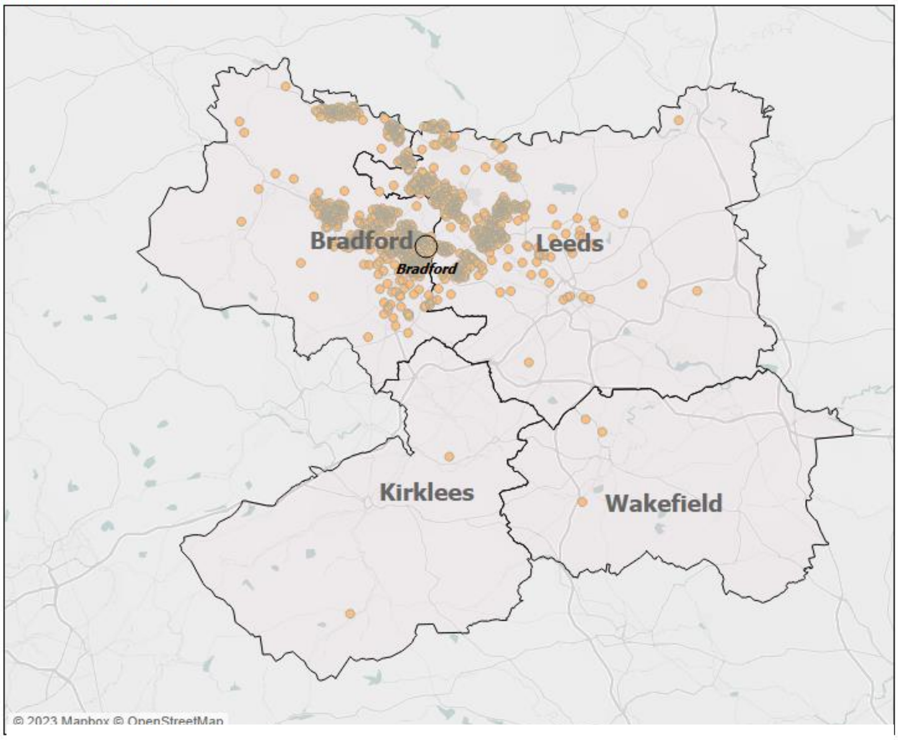
Despite these limitations, as expected, simulated customers travelled much further for 'main' basket types. This is due to the 'main' transactions being assigned to the customer's closest supermarket instead of to any store. Out of the in-store consumer type groups (groups 3 to 7), group 7 was the most notable (Figure 6.14), as they transacted the furthest on average. These customers are generally located in rural areas in the model, reflecting the observed dataset (Figure 5.13). Due to those customers being predominantly rural, their modelled behaviour captured the further distances they would have to travel for 'main' basket type transactions. Although work-based transactions by this consumer type group are not accurately captured (compared to their behavioural summary in section 5.5.7), their residential transactions are.

To visually demonstrate the model's performance, Figure 6.15 presents the difference in customer locations between the observed dataset and the model's output using Tobler's law naïve ruling for all online transactions at the Bradford supermarket. The observed dataset captured customers from 674 OAs across Bradford, Leeds, Kirklees, and Wakefield. The model recorded transactions from 749 OAs within Bradford, Leeds, and Kirklees. The model captured 88.43% of the same OAs as the observed dataset, with the rest being densely located around the Bradford store (Table 6.13). All other stores that fulfil online orders are listed in Table 6.13, along with the OAs that transacted at these stores in the observed and simulated data. On average, the simulated results 'delivered' online orders to 70.58% of the same OAs, with Bradford being the most accurate and White Rose being the least accurate (44.41%). As discussed in section 4.2.5, Shorehead and White Rose supermarkets delivered significantly more online orders than all other stores across a much further distance (Figure 4.5). This was difficult to capture in this store location methodology, as the closest store was always chosen. This analysis is beneficial for Sainsbury's to better understand which customers could be served if their most local store was used to deliver their online orders.

**Table 6.13** Store choice method 1: Tobler's law. Observed vs simulated output; count of OAs captured by online transactions.

| Store Name (Online customers) | Store LAD | Number of OAs transacted from (Observed data) | Number of OAs transacted from (Simulated data) | Shared OAs (%) |
|---|---|---|---|---|
| **Bradford** | Bradford | 674 | 749 | 88.43% |
| **Dewsbury** | Kirklees | 377 | 387 | 72.68% |
| **Halifax** | Calderdale | 774 | 765 | 79.72% |
| **Shorehead** | Kirklees | 1,029 | 676 | 61.22% |
| **Wakefield** | Wakefield | 761 | 869 | 77.00% |
| **White Rose** | Leeds | 1,637 | 953 | 44.41% |

**Bradford supermarket online transactions by customer OA (Observed data)**




**Store choice method 1: Tobler's law - Bradford supermarket online transactions by customer OA (Simulated data)**

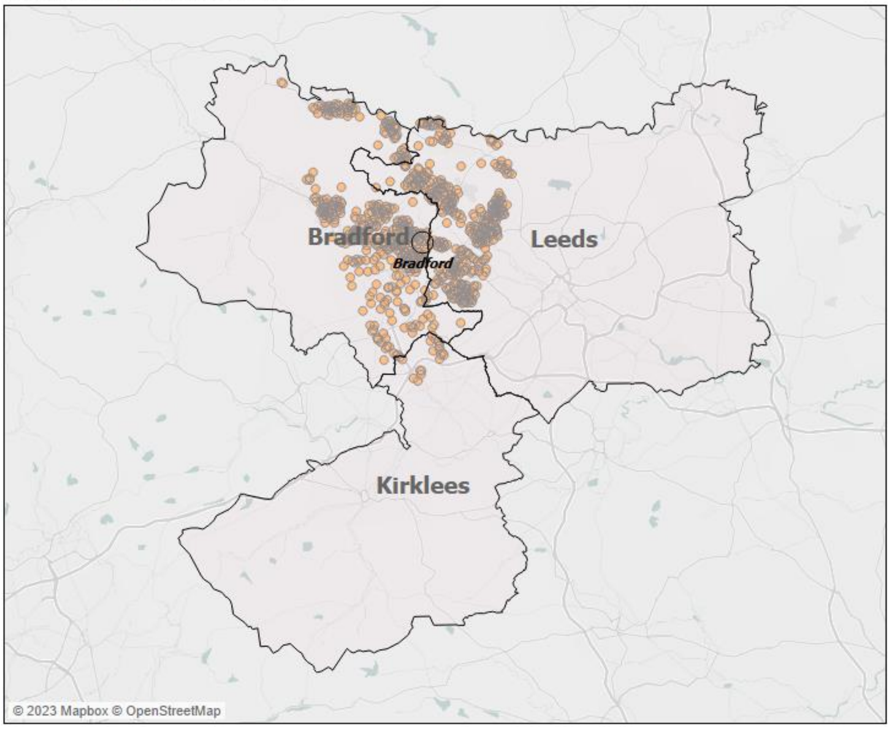
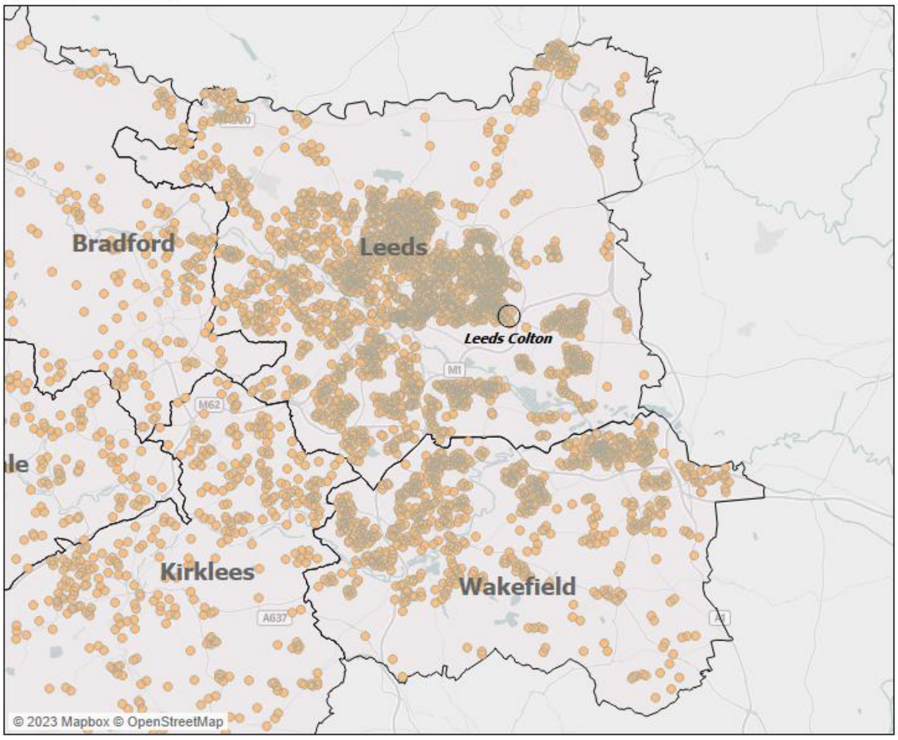



**Figure 6.15** Store choice method 1: Tobler's law. Observed vs simulated output for Bradford supermarket - online transactions by customer OA. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

Whilst Tobler's law-inspired store location methodology is appropriate for modelling online transactions, disparities are more evident when analysing in-store transactions. Tobler's law methodology simplifies the behaviours of simulated customers to an extent in which unpredictable behaviours are not captured, such as multi-trip journeys, workplace purchases, and opportune purchases. Figure 6.16 provides an example of the Leeds Colton supermarket, comparing the OAs of customers who transacted at the supermarket over 12 weeks. The Leeds Colton store was chosen as it is one of the largest stores in the Sainsbury's store network in West Yorkshire, pulling in customers from 3,027 OAs during the 12 weeks in the observed dataset. Leeds Colton is located within the Colton Retail Park, which offers various non-grocery-based retailers. The retail park draws in customers for its selection of shops and Sainsbury's superstore. Customers are likely transacting at the Leeds Colton store for purposeful grocery purchases or out of opportunity as it is close to other retailers they visit.

The model captured local customers of the Leeds Colton store, pulling individuals from OAs in east Leeds and north Wakefield, but has a cut-off point to the east as other Sainsbury's stores are closer to those OAs. The model's logic does not account for Leeds Colton being located in a retail park; therefore, there is no behavioural incentive to draw customers from further afield. The model captured local customers, notably for 'main' transactions. The model identified that customers of Leeds Colton reside in 825 individual OAs, with around 88.64% matching those within the observed dataset. The observed dataset captured individuals from 3,027 OAs, meaning the model matched 24.17% of the OAs. By expanding the consumer's rules to not always consider their closest store for transacting, the model would cover a wider variety of customers. The output of this model has provided further evidence that more sophisticated behavioural rules are required to capture different types of grocery demands beyond residential.

**Leeds Colton supermarket in-store transactions by customer OA (Observed data)**




**Store choice method 1: Tobler's law - Leeds Colton supermarket in-store transactions by customer OA (Simulated data)**

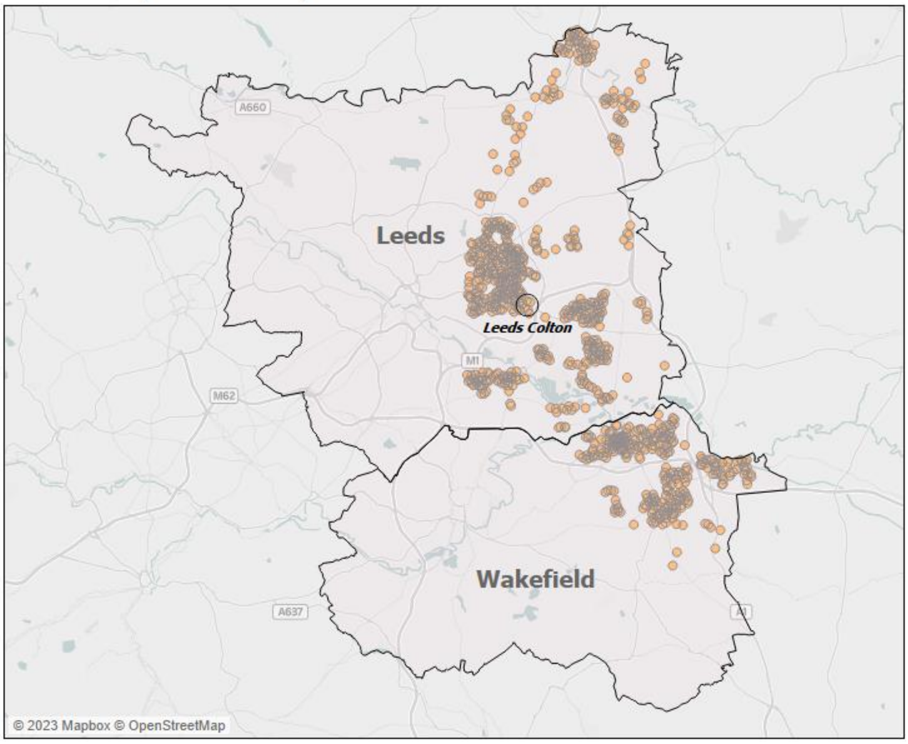
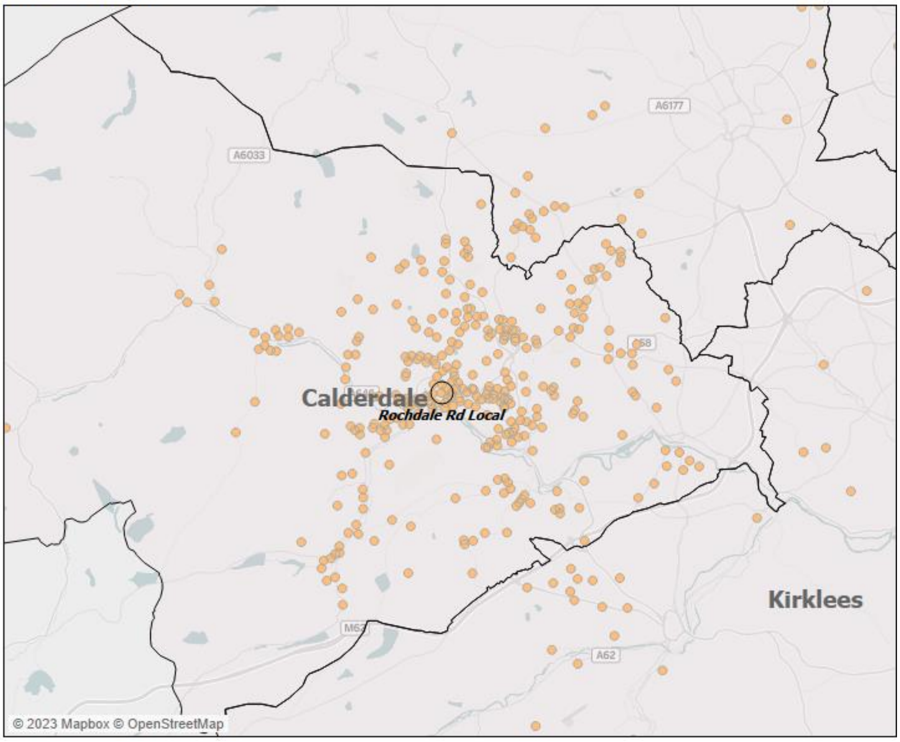



**Figure 6.16** Store choice method 1: Tobler's law. Observed vs simulated output for Leeds Colton supermarket – in-store transactions by customer OA. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

Figure 6.17 provides an example of a Sainsbury's Local store (Rochdale Rd Local), located in the residential area between Halifax town centre and Sowerby Bridge market town centre. This store is the smallest Sainsbury's in West Yorkshire, with a sales area of around 891 $ft^2$, and is mainly visited for 'for now' and 'top-up' basket type transactions. This store captured transactions from around 300 different OAs where their customers resided. Generally, the customer's home addresses are located within a close vicinity of the store, with few customers visiting from elsewhere. The Tobler's law-based rules in the model performed well, with 84.11% of OAs matching between the model and the observed dataset. The transactions from the unmatched OAs in the surrounding districts occurred less often in the observed dataset, with 1 or 2 transactions occurring over the 12 weeks. These rarely occurring transactions are challenging to model at the individual level, notably when the store is located in a residential area, yet customers travel from afar. Outside of local and convenience-based demand, it is difficult to determine why customers would travel from further away and transact at these stores.

To accurately model these unique, low frequency-transactional behaviours, other types of non-residential demand should be incorporated or an added element of store choice randomness. Adding spatiality into a data-driven agent-based model is challenging, especially when attempting to simulate a particular population when only their transaction behaviour is known and nothing more.

**Rochdale Rd Local convenience in-store transactions by customer OA (Observed data)**




**Store choice method 1: Tobler's law - Rochdale Rd Local convenience in-store transactions by customer OA (Simulated data)**

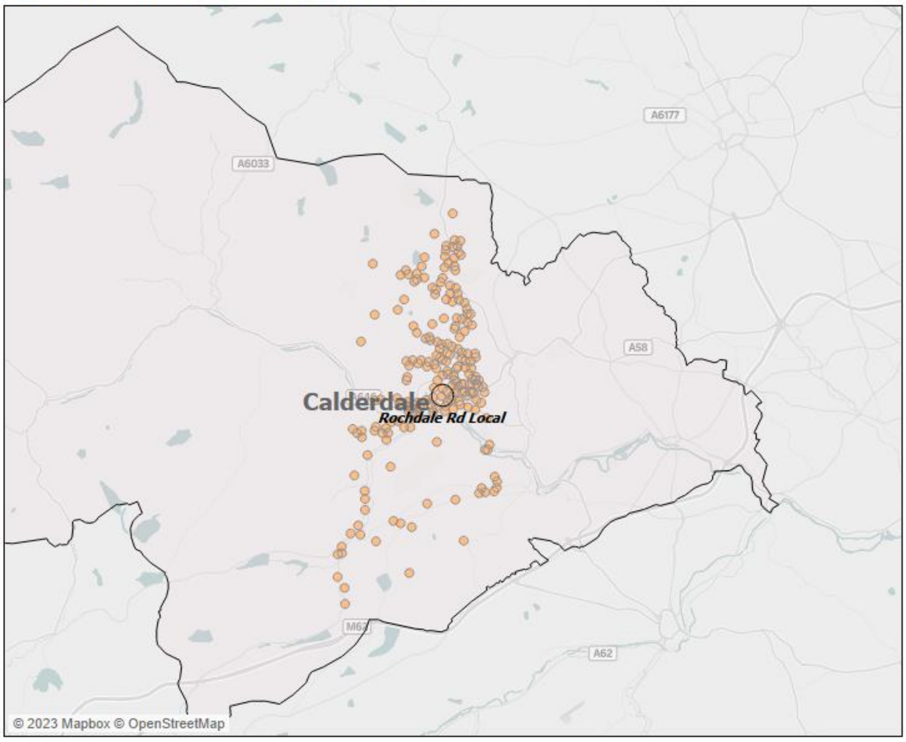



**Figure 6.17** Store choice method 1: Tobler's law. Observed vs simulated output for Rochdale Rd convenience store- in-store transactions by customer OA. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

As this is the first attempt at building an individual-based model of temporal and spatial consumer behaviour using observed data, the study followed the KISS approach. Therefore, the model begins with the simplest behavioural rules, adding complexity to agent behaviour at each iteration of model development. As discussed, this first method of adding store choice to the model followed Tobler's first law, in which the closest store for all transactions would take place, depending on the basket type. The idea behind this logic was to capture residential-based demand in the model by ensuring the customer agents always transacted locally.

The model results show that using Tobler's law, behavioural rules for store choice can capture residential-based demand, notably for stores located in residential areas. Although imperfect, the model captured around 50% to 90% of the same OAs as the observed dataset for each residential store, indicating that the model can capture customer demand locally. Stores in city and town centres shared a much lower proportion of OAs, at an average of 26%. Due to the low number of customers residing in these areas, and an influx of customers from elsewhere, capturing these behaviours using Tobler's law behavioural rules is nearly impossible without relocating customers at different time points in the model.

Linking the customers in the dataset based on their consumer type group is challenging, especially when personal or demographic information is unknown. Therefore, before attempting to use other data sources, the following section discusses how the transaction dataset can incorporate more complex behavioural rules into the model whilst maintaining the ability to capture residential-based demand observed in the Tobler's law-inspired model.

### 6.3.3 Store choice method 2: Distance bands

The next step of the model-building process included a second decision tree-like structure in which customers' distances were determined. In this second store choice method, complexity was added by calculating the distance a customer is likely to travel depending on the transaction scenario being performed. As we do not know where the customer will likely be located at different time points in the model, their known behavioural data from the consumer type group analysis was explored.

For each transaction scenario, i.e., the combination of day type, time of day, basket type and channel, the average distance travelled was derived per consumer type group. The idea behind this method was to assess whether the loyalty card-linked transactional dataset alone can provide enough information to identify the types of areas customers were transacting in, in relation to temporality. Therefore, distance bands were assigned, ranging from 0 - 1 km, 1.1 - 2.5 km, 2.51 - 5 km, 5.1 - 10 km, and ≥ 10.1 km. For each consumer type group, the probability of a customer making a transaction based on every transaction scenario was calculated using the average group-level proportions.

A second decision tree was created using the same weighted roulette wheel selection function discussed in section 6.1.4. Each time a customer performs a transaction, the model reads in a file containing the distance probabilities for each transaction scenario by distance band for each consumer type group. The distance probabilities were standardised to 1, ensuring that a store would always be chosen. For example, customers in consumer type group 7 have a 44.36% chance of transacting at a store between 5.1 km and 10 km away if their transaction is a 'weekend morning in-store main' scenario. For the same scenario, they have a 34.51% chance of transacting between 2.51 km to 5 km away from home.

Using a decision tree structure should ideally reflect the average distances customers in each consumer type group travelled, linking back to the observed dataset from our study collaborator. During the development of this model iteration, several challenges were faced. Firstly, not all customers had a store within the distance band they were assigned to transact within, therefore could not perform the transaction. To overcome this issue, the model re-standardised the distance probabilities in real-time as the model ran. Therefore, if a store was not within a distance band, the next most probably distance band was chosen. The second challenge was store selection. When a customer has more than one store to choose from in the distance band, how should a customer decide where to go? To combat this issue, two solutions were considered. Solution one entailed assigning the customer to transact at their closest store within the distance band, as their distance parameter had already been set using probabilities. Solution two considered using a random function to assign a customer to any store in a distance band. To keep in line with the KISS approach, the first solution was applied to identify whether distance bands and probabilities alone could replicate the observed data.

#### 6.3.3.1 Store choice method 2: Distance bands results

The model was initialised to run for 84-time steps, producing 12 “weeks” of data to match the observed dataset. Using distance bands, Figure 6.18 presents the cross-district interactions between customers and Sainsbury’s stores. The model results found that for each district, most transactions occurred within the same district the customer resided within, similar to the observed data (Figure 4.6) and the output in store choice method 1 (Figure 6.13). The distance bands method produced a high proportion of cross-district transactions, an improvement from the previous spatial model. The model captured a similar pattern of cross-district transactions compared to the observed dataset. For example, customers who live in Bradford performed 9.0% of transactions in Calderdale, 1.4% in Kirklees, and 11.3% in Leeds (Figure 6.18) compared to the observed dataset (Calderdale: 8.6%, Kirklees: 1.1%, and Leeds: 13.0%) (Figure 4.6). The precision for the proportion of transactions in Wakefield was not calculated, as zero transactions occurred in the model. The precision of these results was assessed with the following calculated CV values for the above Bradford example: 1.3% (Bradford), 3.2% (Calderdale), 17.0% (Kirklees), and 9.9% (Leeds). Overall, the average CV for Bradford was 7.8%. At the aggregate level, the proportion of cross-district transactions is accurate for Bradford but may be inaccurate at the micro-level, as the same OAs may be being assigned to stores in other districts; this is explored further in the following sections.

The average cross-district CVs were calculated for all other customer LADs, with the following results: 53.0% (Calderdale), 28.8% (Kirklees), 31.7% (Leeds), and 17.1% (Wakefield). The higher the CV, the less similar the model output is to the observed dataset. The distance bands methodology worked well for Bradford but needs refinement to find similarities for Calderdale and Leeds. These disparities are due to the residential locations of customers in the model and the logic rules applied in this store choice method. The model assigns customers to a store to transact at based on the chosen distance band; however, it does not inform the customer which direction to choose a store and automatically forces them to choose the closest. Therefore, customers who reside on the border of districts may be pulled into an adjacent district to transact at a store, resulting in the proportions presented in Figure 6.18. The output from the model that used Tobler's first law (section 6.3.2) produced relatively close results for the customers within the Leeds LAD compared to the observed data (Figure 6.13). This is due to the higher proportion of Sainsbury's stores within Leeds, and customers' closest stores were within the same LAD.

**Store choice method 2: Distance bands - Proportion of in-store transactions between customer and store LADs (%)**

| | Customer LAD | | | | |
|---|---|---|---|---|---|
| **Store LAD** | **Bradford** | **Calderdale** | **Kirklees** | **Leeds** | **Wakefield** |
| **Bradford** | 78.3% | 2.7% | 0.4% | 6.1% | |
| **Calderdale** | 9.0% | 70.6% | 7.9% | | |
| **Kirklees** | 1.4% | 26.7% | 87.6% | 1.1% | 2.5% |
| **Leeds** | 11.3% | | 1.7% | 88.5% | 10.1% |
| **Wakefield** | | | 2.4% | 4.4% | 87.5% |
| **Grand Total** | 100.0% | 100.0% | 100.0% | 100.0% | 100.0% |

**% of Total Count of Extract**

0.4% – 88.5%

**Figure 6.18** Store choice method 2: Distance bands. Proportion of in-store transactions by customers in each Local Authority District into stores in each Local Authority District (%).

The average distances that the modelled customers 'travelled' are presented in Figure 6.19, split by consumer type group and basket type. The average distances are slightly higher than the output in the store choice method 1 model (Figure 6.14) and represent the observed data more accurately (Figure 5.4). It is not easy to compare the average distances between the observed data and simulated output, as different measures were used. The data provided by Sainsbury's used their in-house GIS, in which the distances between customer homes and stores used a road network distance in kilometres, whereas the model uses geodesic (straight line on a curved surface) distances. The different distance measures are expected to produce different outputs and are difficult to model without using a common distance methodology. Therefore, an alternative method to distance bands may be more appropriate for modelling store choice unless the exact measures of distance are able to be used.

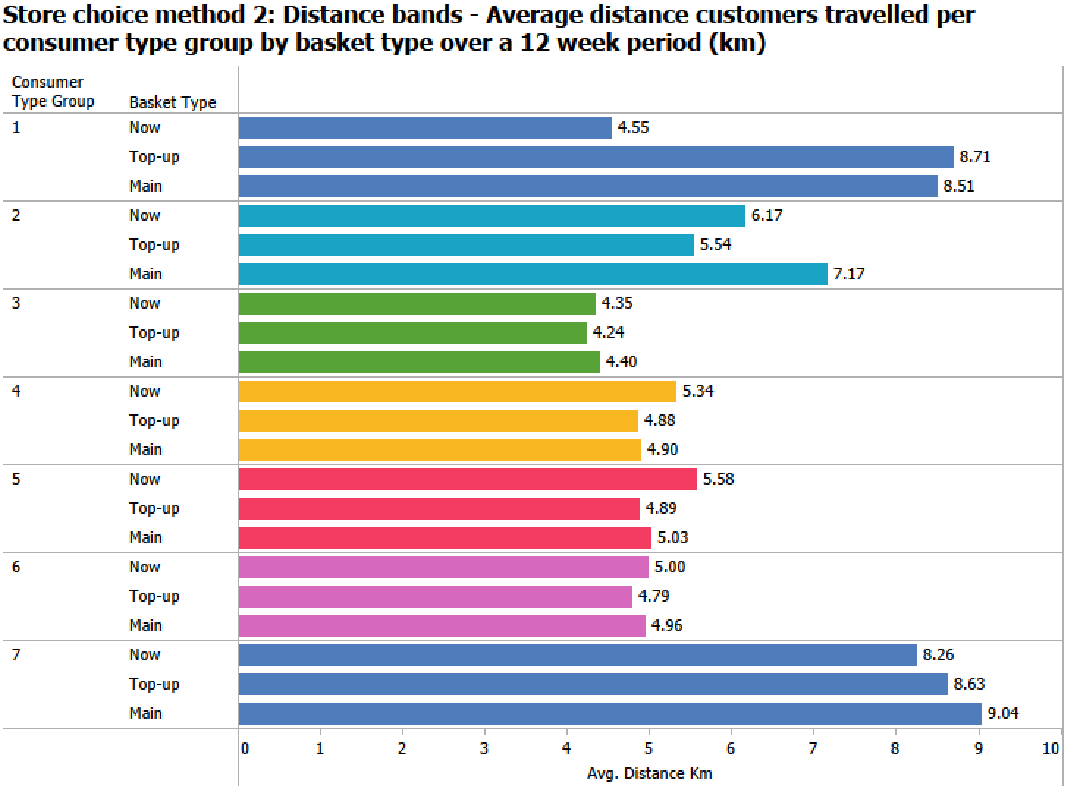
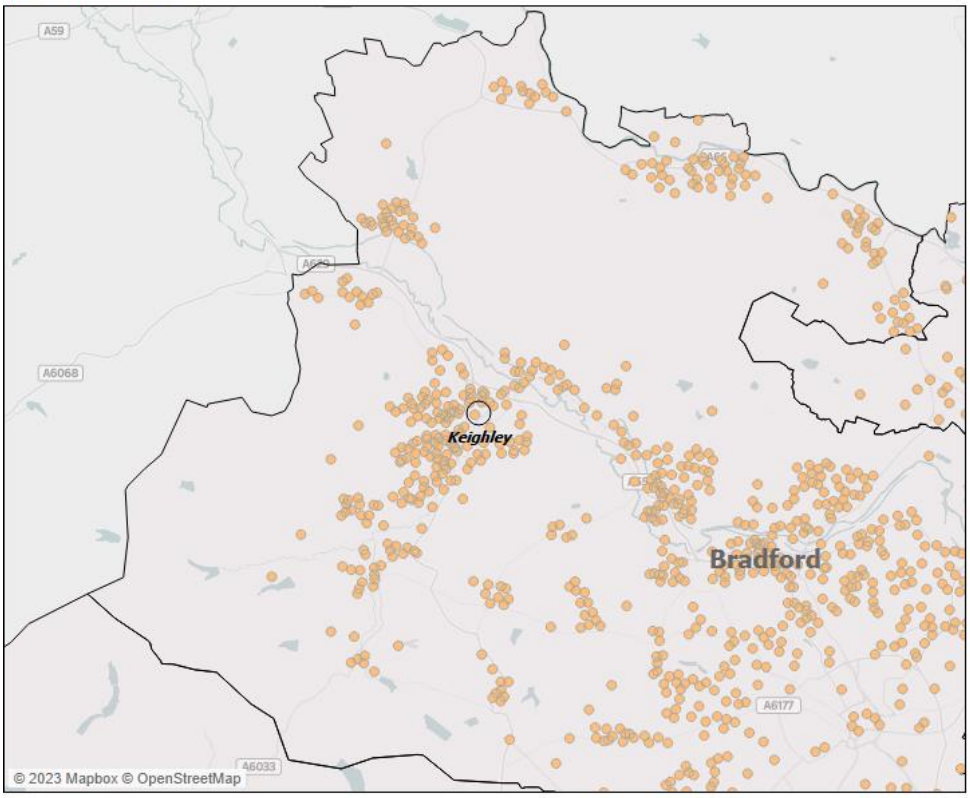



**Figure 6.19** Store choice method 2: Distance bands. Average distance travelled by basket type for each consumer type group over the 12-week period (km).

A series of maps are presented in the following section to visualise the interactions between select Sainsbury's stores and OAs, comparing the observed data to the first store choice method results and the results of using the distance band methodology. Figure 6.20 compares in-store transactions for the Keighley supermarket in Bradford. Customer OAs are highlighted in yellow circles and only OAs are shown in which a transaction happened by a customer in the area, at the Keighley store. The observed dataset found a wide range of customer OAs transacting at this particular store, predominantly customers in Bradford, followed by Leeds (to the east) and Calderdale (to the south) (Figure 6.20, map (a)). The first spatial model using Tobler's law successfully identified the most local OAs of Sainsbury's customers (89.6% of the OAs in the model output matched those in the observed dataset). However, as the model only uses the closest store rules, the model overpredicted the number of local

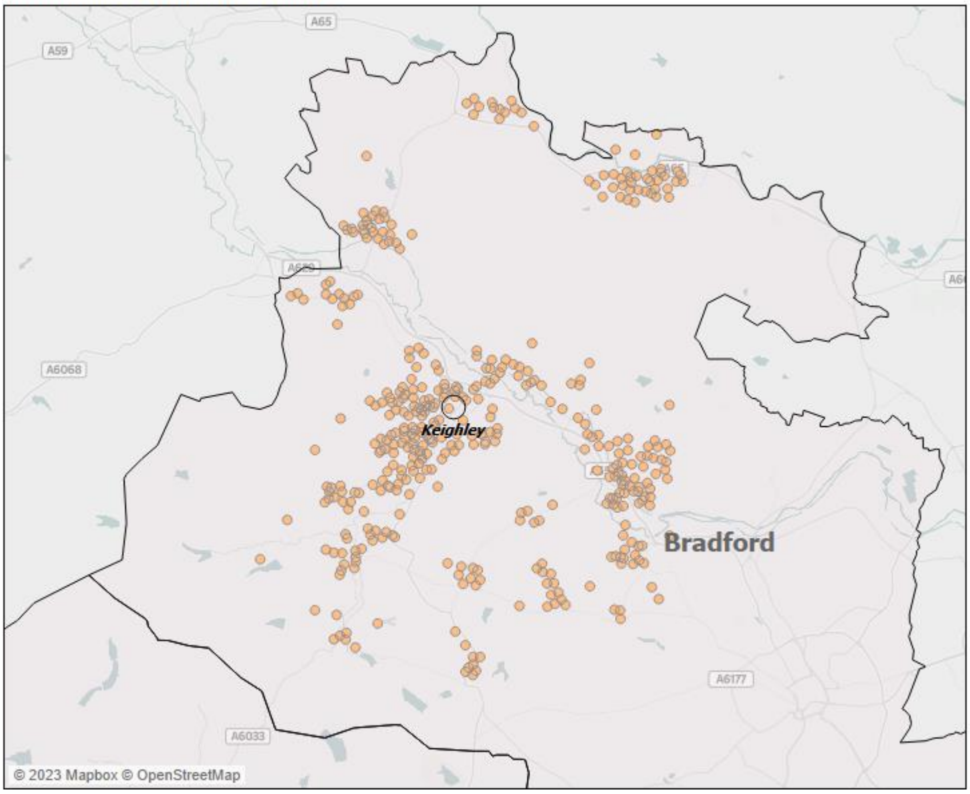
OAs in which customers visited the Keighley store (Figure 6.20, map (b)). As unique store distances are not assigned as a rule to those who often transact far away (such as those in consumer type group 7, ‘Hungry and Far From Home’), customers of that groups are forced to transact locally. This overprediction has been reduced using the distance bands method (Figure 6.20, map (c)). As the maps show, the model using distance bands enabled the Keighley store to draw in customers from further away and from fewer OAs. Striking a middle ground between the two models is a challenge, though, and further complexity is required to fully capture consumer store choice.

**Keighley supermarket in-store transactions by customer OA (Observed data)**

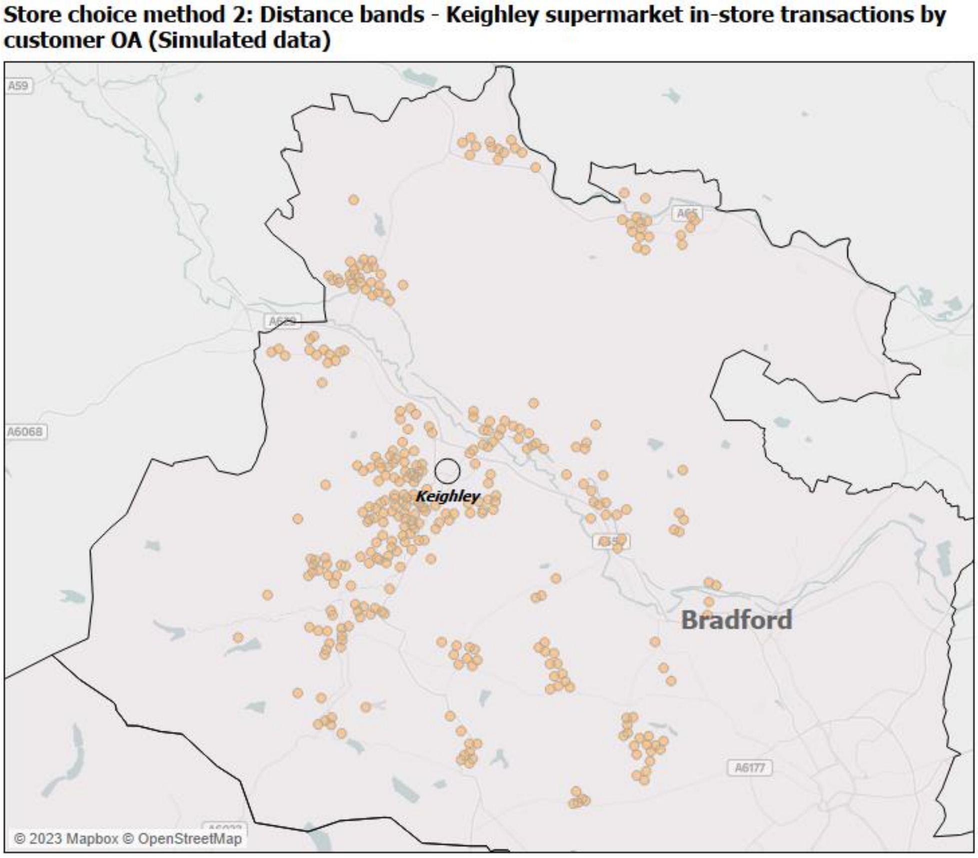



(a)

**Store choice method 1: Tobler's law - Keighley supermarket in-store transactions by customer OA (Simulated data)**




(b)

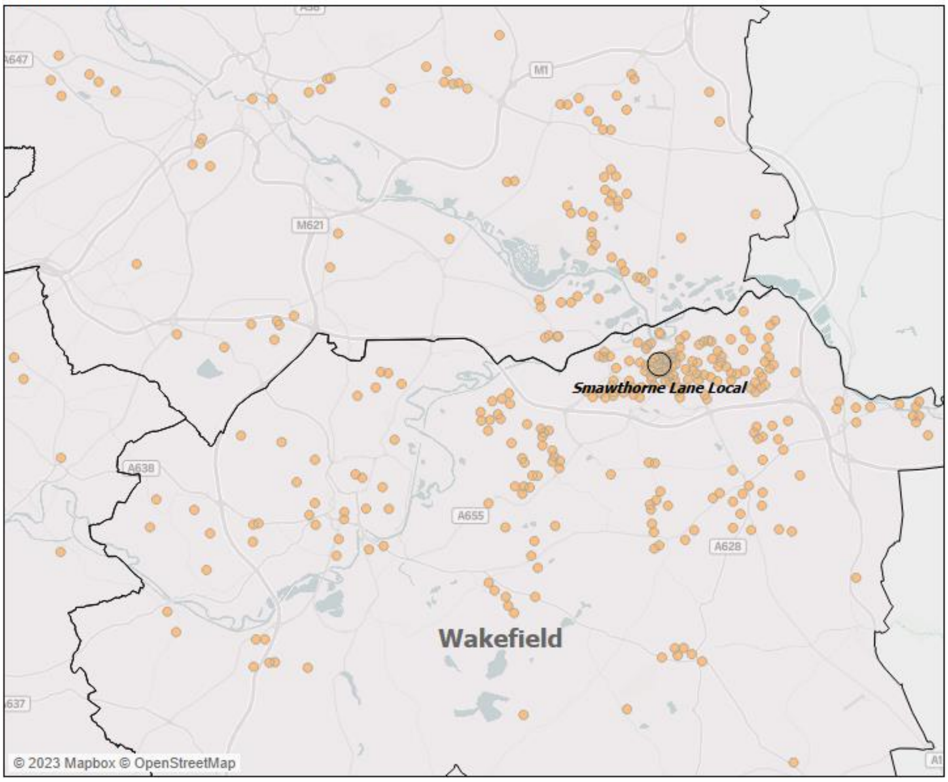



(c)

**Figure 6.20** Store choice method 2: Distance bands. Observed vs simulated method 1 vs simulated method 2 - output for Keighley supermarket store- in-store transactions by customer OA. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

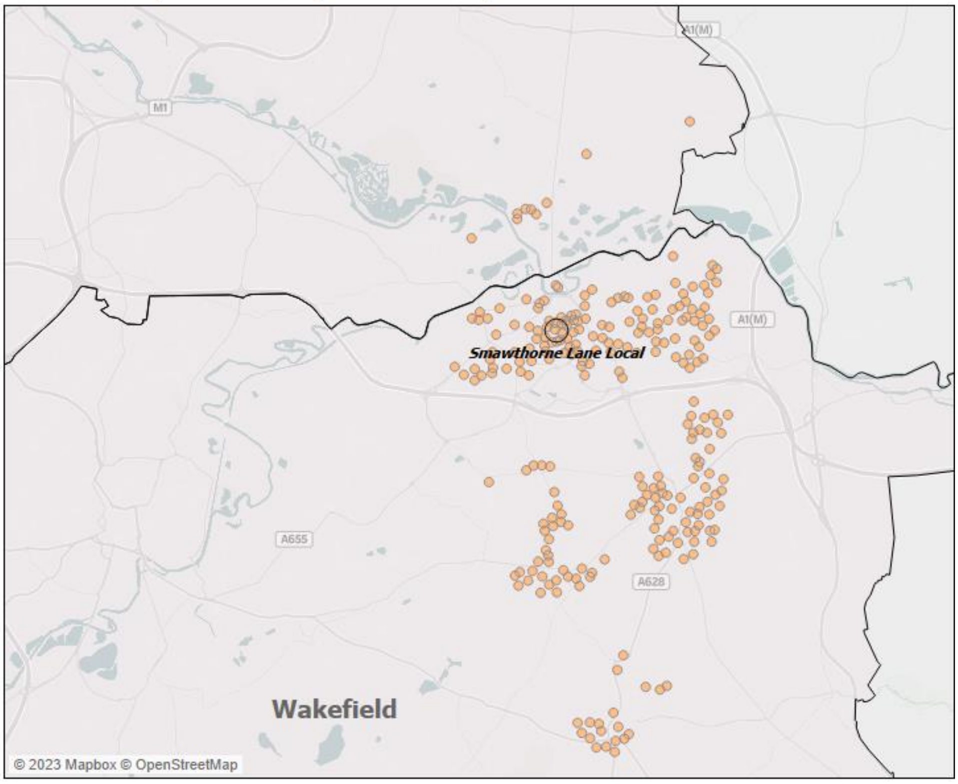
A second example of the model's spatial results is presented in Figure 6.21 for the Smawthorne Local convenience store located in Wakefield. This store offers a 1,250 ft$^2$ sales area, offering a limited range of goods due to its small size and convenience fascia. This local store is intended to serve the residential population around the area, with most customers transacting locally (Figure 6.21, map (a)). Similar to the Keighley store, the model using Tobler's law behavioural rules pulled in mostly local customers for the Smawthorne store, overpredicting some local areas to the south, and missing out customers to the north of the Wakefield boundary (Figure 6.21, map (b)). The second store choice method captured largely the same OAs of customers for the store and customers located in neighbouring districts (Figure 6.21, map (c)). By using the distance bands, the model manages to simulate customers transacting at a wider variety of stores, similar to the observed dataset but still has its inaccuracies despite assigning behavioural rules unique to consumer type groups. The following section provides discusses the benefits and limitations of solely using the loyalty card-linked transaction dataset to incorporate spatiality into the individual-based model, highlighting what alternative rules combinations could be used, along with other data sources.

**Smawthorne Lane Local convenience in-store transactions by customer OA (Observed data)**



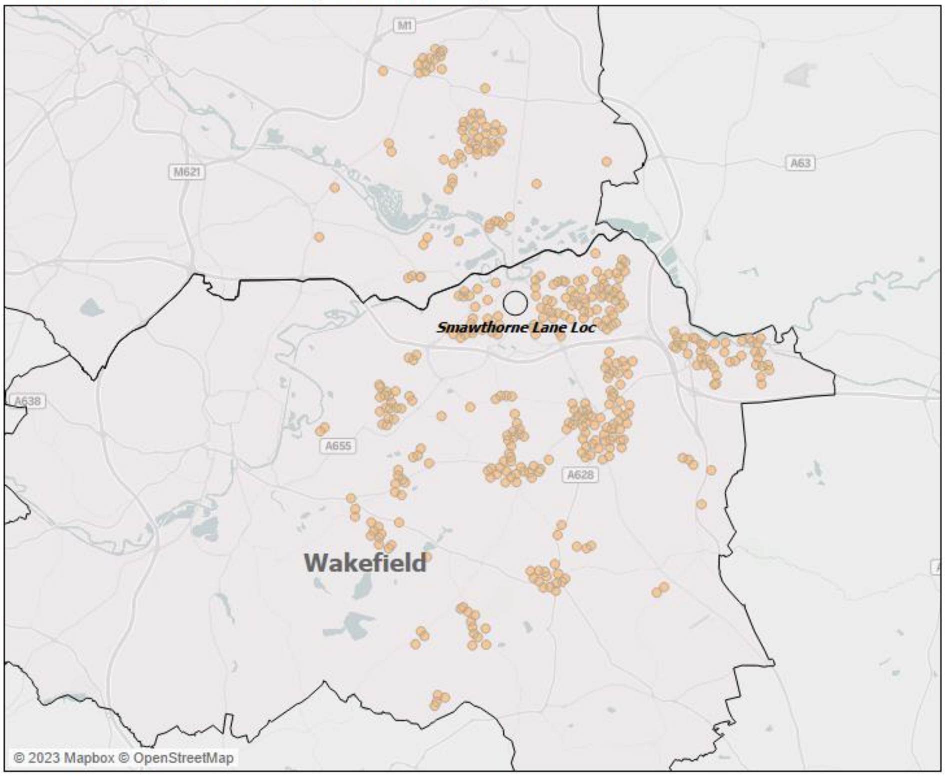

(a)

**Store choice method 1: Tobler's law - Smawthorne Lane Local convenience in-store transactions by customer OA (Simulated data)**




(b)

**Store choice method 2: Distance bands - Smawthorne Lane Local convenience in-store transactions by customer OA (Simulated data)**




(c)

**Figure 6.21** Store choice method 2: Distance bands. Observed vs simulated method 1 vs simulated method 2 - output for Smawthorne Lane Local convenience - in-store transactions by customer OA. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

### 6.3.4 Benefits and limitations of solely using transaction data to model consumer store choice

When modelling individual consumers, it is notoriously challenging to derive general spatial behavioural rules when little information is known about the customers (Wilkinson, 2023). The first two attempts at incorporating store choice into the model followed the KISS approach, utilising the most available data for this study: the loyalty card-linked transaction data. For both attempts, individual customers were placed in the model using the observed dataset to directly compare observed customer store choices and model output.

The first model (section 6.3.2) used the most straightforward behavioural rules by assigning customers to their closest stores. Basket type determined whether the customer would choose their closest supermarket for ‘main’ or ‘top-up’ transactions or any closest store for ‘for now’ transactions. At the local scale, the model performed well for modelling local, residence-based demand, capturing various surrounding OAs for each store. The model also worked reasonably well for assigning transactions to online stores. Overall, the Bradford district performed the best in the model due to the few stores within the district (Table 4.4). The model did not perform well for Leeds, which has the greatest number of stores. This is due to the low number of Sainsbury’s customers that reside in the city centre and the abundance of Sainsbury’s stores within the city centre. The residential grocery demand in city centres is often low and is replaced with workplace demand, as found in Berry et al. (2016) and Waddington et al. (2019). Overall, the first store choice method worked reasonably well for residential-based demand, though not quite capturing a wide range of OAs.

The second model (section 6.3.3) used a more complex store choice methodology, in which distance bands were used. The proportional chances of transacting within a distance band were calculated, unique to each consumer type group and transaction scenario. This attempt followed the same ruling for linking basket types to store types but incorporated more sophisticated rules via a decision tree. This method was tested to see whether we could retain the local, residence-based grocery demand captured using the first modelling method for store choice whilst capturing a wider range of OAs for each store. In essence, the method sought to capture a wider variety of store distances based on the consumer type group preferences. This method of assigning store choice produced model results similar to that produced in the first modelling method and was slightly better at representing the observed data.

Using the transactional dataset solely to assign behavioural rules for store choice provided an easy-to-follow method, as each attempt built up from the previous. Ideally, the dataset alone would provide enough information to create behavioural rules representing the observed consumer behaviours. However, we need more information about these customers to be able to use the data for modelling individual store choices. The dataset can simulate residential-based grocery demand, as customer home locations are known, but other types of grocery demand are unknown. Therefore, as discussed in section 3.3, other data sources are required to make more representative models of consumer store choice behaviour. The loyalty card data provides a solid base for modelling individual consumers' temporality and store choice, but further data regarding customers' location outside the home is required. The prior sections discussing how store choice behavioural rules can be derived from the loyalty card dataset provide insight into the extent that these data can be used and the critical limitations of these data.

The following section provides context and an example of incorporating census data into the IBM, focusing on a single OA in Leeds.

### 6.3.5 Store choice method 3: Place of work

This thesis's third and final store choice methodology utilises the 2011 Census workplace data(Office for National Statistics, 2014). These data provide the flows between each OA in England and Wales, in which one side is the individual's place of residence, and the other is the OA where the individual's place of work is. This dataset was chosen for the model to incorporate workplace grocery demand during usual weekday working hours. Keeping in line with the whole model's philosophy of 'keep it simple, stupid', the model output in this section focuses on one single OA in Otley, Leeds. The particular OA has not been shared due to consumer confidentiality. Due to time constraints in the development of this research (as discussed in the Covid-19 impact statement provided separately), the amount of time required to calculate the flows between thousands of OAs, and then assigning them to Sainsbury's customers, select examples are presented to show the potential of the model.

In this model iteration using workplace data, around 88 unique customers within an Otley OA were loaded into the model. The 88 customers were identified within the loyalty card dataset out of a total area population of ~1,000. This study area example was chosen as Otley is a well-known market town in the north of Leeds, with a variety of local supermarkets, including Waitrose, Asda, and Sainsbury's. The population of this OA are categorised as 'Ageing industrial workers', who have a slighter older age profile and are employed in IT and financial-related industries (Office for National Statistics, 2015). Although the model does not incorporate competitors, this area is of interest in future case studies due to the variety of supermarkets, with Waitrose being classified as a 'luxury' brand, Asda as a ' lower-end mass market' brand, and Sainsbury's as a 'higher end mass market' retailer (Clark et al., 2021).

For this example, the 88 Sainsbury's customers in this OA were assigned a home location, *the OA,* and a workplace location. Using the workplace data from the 2011 Census, the proportion of those Sainsbury's customers that would leave the Otley area for work was calculated. Approximately 77% of the customers worked within Otley, were unemployed, or were retired. The remaining 33% of these individuals *do* leave Otley for work and visit various other OAs across Bradford and Leeds, according to the 2011 Census data. These are inferences, as we do not know exactly which Sainsbury's customers leave Otley, but these assumptions were made. Therefore, the model was initialised with 88 customers from a particular Otley OA, and each customer was assigned a workplace within Otley or elsewhere. Customers were then assigned a consumer type group, depending on the proportion of customers who lived within the OA during the 3-month period of the loyalty card dataset. As expected, most customers belonged to consumer type group 3 (section 5.5.3 for this group's pen portrait).

The behavioural rules of the model were then amended from the previous store choice methods. If a customer were to transact during a weekday morning or weekday afternoon, that transaction would occur from their workplace location. Using the same ruleset as previous store choice methods, the closest store would be chosen, with 'for now' transactions taking place at any store and 'top-up' and 'main' transactions only taking place at a Sainsbury's supermarket store. This methodological approach aims to replicate the demand layers used within spatial interaction models (Newing, 2013), albeit it is not as simple as using the top-down methodology. Essentially, this model iteration attempts to capture individual-level behaviours of consumers who are being relocated during the model to represent workplace demand (Waddington et al., 2019).

Validating this section of the model is challenging, especially when working with assumptions within an agent-based model (Pullum and Cui, 2012). Therefore, the same observational validations used in the prior store method choice sections are used, comparing the mapped model outputs. Figure 6.22 Figure 6.23 are to be analysed together, with each image acting as a progressive step in modelling spatiality in this model. All models were run on a timescale of 12 weeks for comparison to the observed dataset.

#### 6.3.5.1 Store choice method 3: Place of work results

First, Figure 6.22 presents the output of the observed dataset, showing the stores chosen by the customers within the Otley OA. The percentages on the map are the proportions of transactions at the mapped stores. As shown in the observed dataset, Sainsbury's customers in this OA mostly transacted at Otley (95.37%). The remainder of the transactions occurred at various stores across Leeds, Bradford and Halifax. These customers predominantly shopped locally, with few transactions occurring in Leeds city centre and no online transactions. The second map in Figure 6.22 shows the output from the store choice method 1 model, in which customers *always* transact locally. Unsurprisingly, these customers did 100% of their transactions at the Sainsbury's Otley supermarket due to only one store near this OA. Although this model is 4.63% off from the observed dataset, it cannot capture transactions from out of Otley.

Figure 6.23 presents the output from store choice method 2, which uses distance bands. This modelling method captured a wider variety of Sainsbury's stores. However, the proportions of transactions at those stores are *very* dissimilar to those within the observed data (Figure 6.22). The distance bands methodology may be able to capture varied distances, but there needs to be a fundamental improvement in the way it performs. The problem is if the model cannot find a store within a distance band and is constantly standardising the distance bands for the weighted roulette wheel function used for spatiality. This results in customers in the model favouring the Menston Local store on Bradford Road for non 'main' transactions. This model aimed to incorporate store variety based on customer basket types but did not perform optimally.

The second map in Figure 6.23 presents the output for the third store choice method, 'Place of work'. By moving customers out of Otley during workday mornings and afternoons, the model output performed exceptionally well, being the most similar to that presented in the observed dataset (Figure 6.22). The output of this model gives hope that future model developments will be successful in incorporating other datasets, such as the census 2011 workplace data. The model captured various stores as customers moved to Leeds and Bradford, forming work-based grocery demand. Other study areas were analysed, using a few OAs from select areas, and produced similar results. Using workplace data resulted in the most similar model results to the loyalty card dataset and should be considered for future models. There are obvious drawbacks to using the census data in this model, however, especially with the discrepancies between the periods of the datasets; this is discussed further in Chapter 7.

The next section of this thesis provides an example of how the model can be used for future scenario testing for retailers. As this is a prototype model, the scenario test output is not fully explored but is for demonstrative purposes.

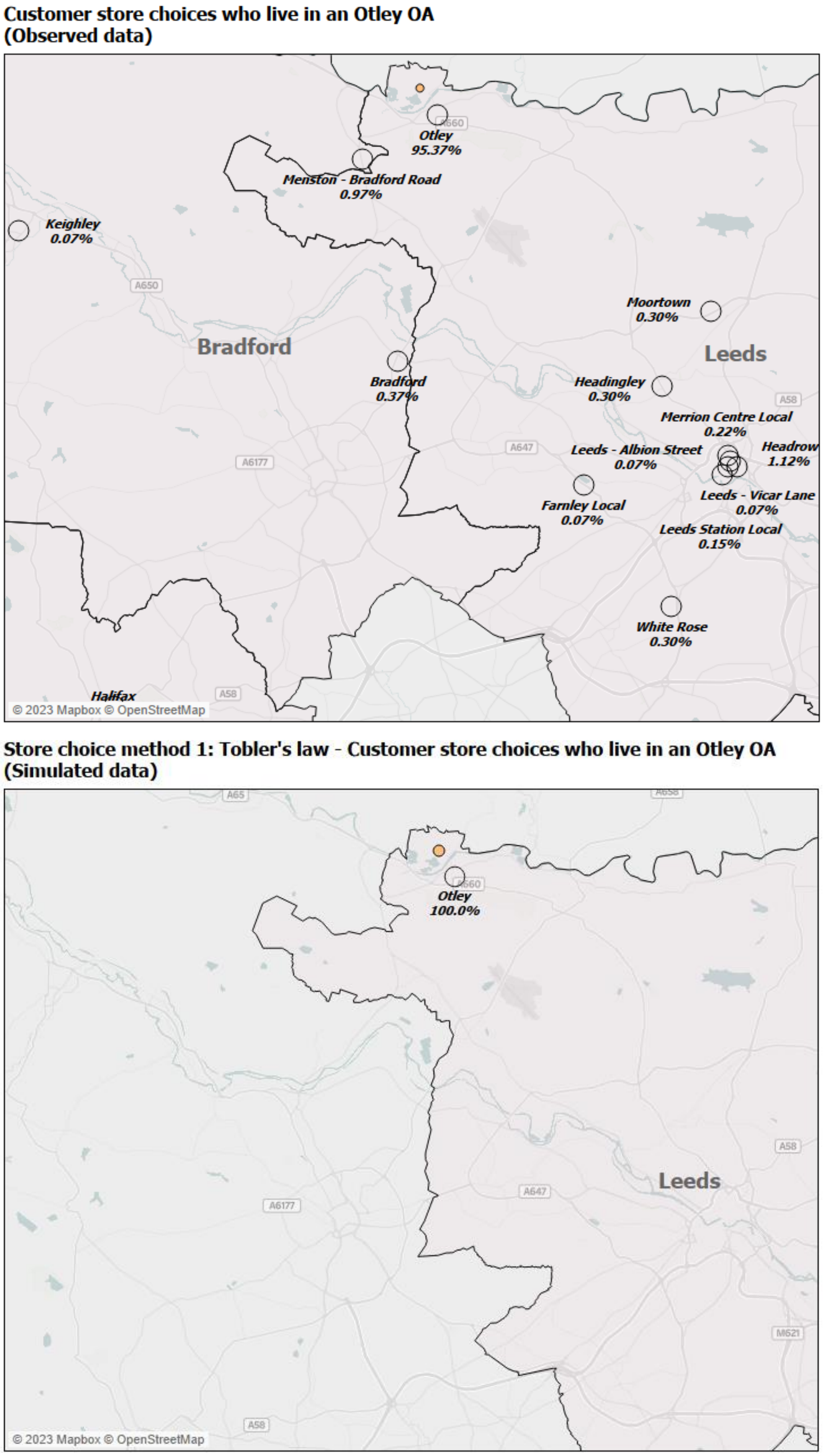



**Figure 6.22** Store choice method 3 comparison: Tobler's law - Observed vs simulated method 1 – stores visited by those in a single Otley OA. Percentages are the proportion of transactions that took place at each store in total. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

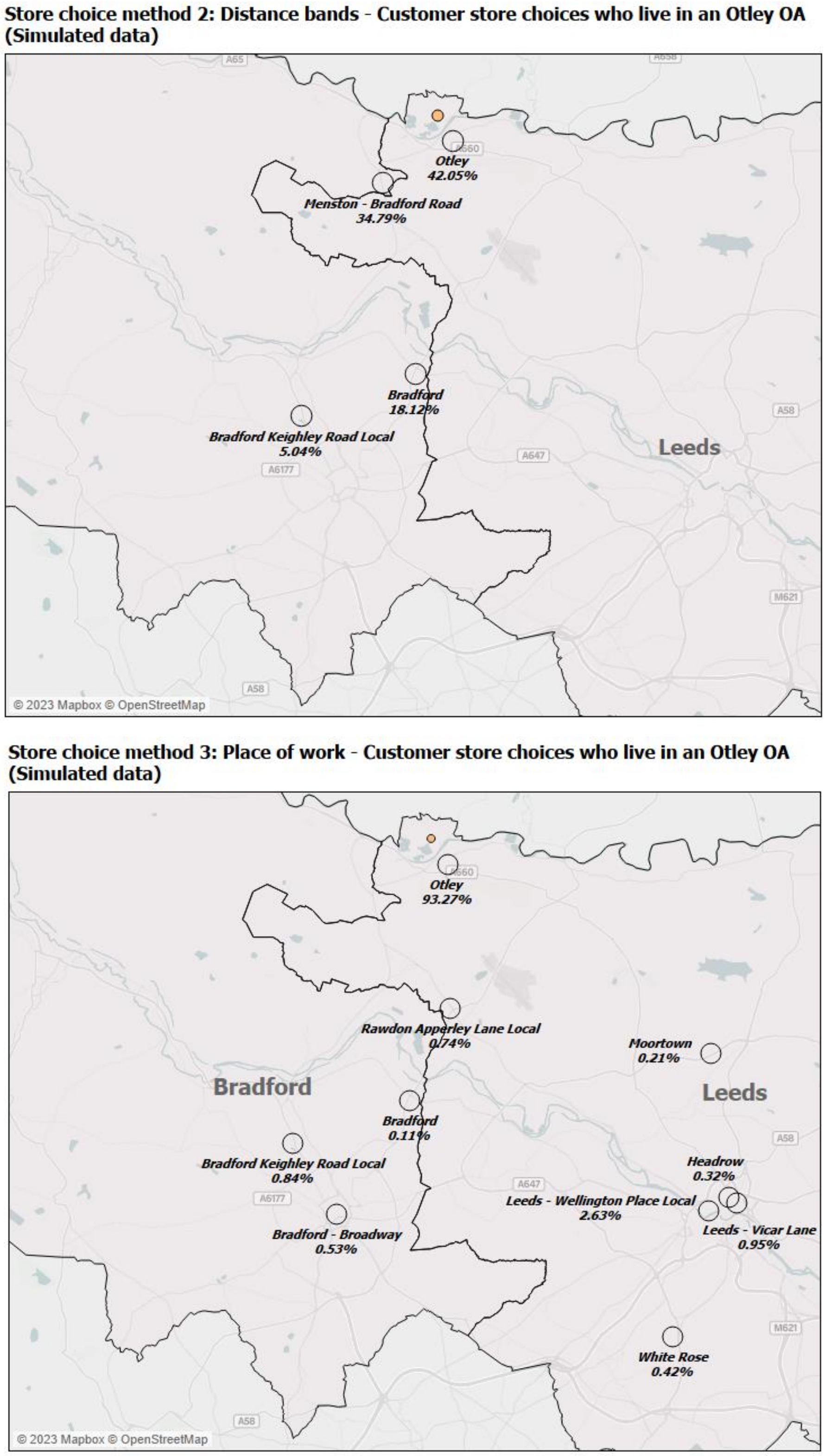



**Figure 6.23** Store choice method 3 comparison: Distance bands vs Place of work - stores visited by those in a single Otley OA. Percentages are the proportion of transactions that took place at each store in total. Black outlined dots represent Sainsbury's stores, yellow dots are customer home OAs.

### 6.3.6 Scenario test: Work-based store addition

This section of the chapter provides a brief example of how the model could be used for scenario testing in the future. Using the third store choice method that incorporates the workplace data from the 2011 Census, the example demonstrates the addition of a new convenience store to the Sainsbury's store network. The demonstration of the model in action takes place at the Dean Clough complex in Halifax, Calderdale. Dean Clough is a 22-acre mixed-use mill complex located on the edge of Halifax town centre and was once a group of large factory buildings (deanclough.com, 2023). Today, Dean Clough consists of 16 Grade II listed Victorian mills that were converted for various commercial and cultural uses; it is a leading example of successful urban regeneration (Stratton, 2000). Dean Clough offers a variety of things to do, boasting large workspaces, eateries, and boutique shops. However, the one thing it does not have is a small convenience store for those not-so-ostentatious quick eats.

In 2014, one of Calderdale's largest employers, Covéa Insurance, purchased two mills at Dean Clough, bringing over 1,000 employees into their offices (Covéa Insurance, 2014). These employees need their lunches, and not every employee will want to eat lunch at a quirky cafe every day. Therefore, this scenario tests what could happen if a new Sainsbury's Local opened at the Dean Clough complex. To model this scenario, a spatial interaction model could be used to estimate the area's demand for the store by disaggregating the number of customers in the area and estimating their expenditure. However, we know that some of those employees at Dean Clough are likely to already be Sainsbury's customers depending on the OA they reside in. Those customers will also belong to one of the seven consumer type groups identified in Chapter 5. Therefore, what potential does the IBM developed in this thesis have to estimate the behaviours expected at this new store? To model this, we need to know who is relocating to Dean Clough for work and where they reside.

However, there is a misalignment in the available statistics regarding worker movements, as each dataset is from a different time period. The 2011 Census captured worker flows before Covéa relocated to Dean Clough (in 2014); therefore, the census will not account for those employees in the area. The 2021 Census could be used; however, that census captured the population's information during the Covid-19 pandemic, and Covéa adopted a hybrid working environment. Additionally, the loyalty card dataset being used to link workers to consumer type groups were based on 2018 Sainsbury's customers. Therefore, there is quite a disparity between the available datasets.

Regardless, as this is an example demonstration, the thesis uses the 2011 Census to estimate how many Sainsbury's customers worked at the Dean Clough complex. The 2011 Census identified that 2,075 individuals worked at Dean Clough and resided within various OAs across Calderdale, Bradford, and Kirklees. Using the consumer type group proportions, the number of Sainsbury's customers in each OA was calculated, resulting in the assumption that ~100 of those individuals that worked in Dean Clough were already Sainsbury's customers with a loyalty card. The Sainsbury's customers in these OAs belonged to consumer type groups 3 (Convenience Suburbanites) and 7 (Hungry and Far From Home).

The model was initialised to run for 84-time steps, and customers were located at their home OAs using the census data. These customers were then assigned their place of work (Dean Clough) during the model's weekday mornings and afternoons. Therefore, if a customer were to transact (based on their consumer type group behavioural rules for temporality) during those times, they would transact as if they were at their place of work. The new store “Dean Clough” was added to the model using coordinates and was located on the site.

The model's results for this scenario test are presented in Figure 6.24, Figure 6.25 and Figure 6.26. Each figure presents a breakdown of the stores chosen by all customers for each day type, basket type, and time of day. As the model output cannot be validated against other data, general observations are made whether the relationships between consumer type groups and the model's results are logical, reflecting the expected behaviours. All 'for now' and 'top-up' transactions by these modelled customers occurred as expected at the newly placed 'Dean Clough' store on weekday mornings and afternoons (Figure 6.24 and Figure 6.25). All weekday morning and afternoon 'main' transactions occurred at the large Halifax Supermarket, aligning with the behavioural rule that only supermarkets provide enough product range for 'main' transactions (Figure 6.24,Figure 6.25 and Figure 6.26). Customers always choose one of the closest stores to their home OA for weekend transactions for all basket types, forming residential-based grocery demand. Customers are located in particular OAs based on their consumer type groups, with those living the furthest away from Dean Clough mainly belonging to consumer type group 7 (Hungry and Far From Home).

This demonstrative example of adding a new store into the model for simulating the potential behaviours of customers provides insight into the possible future scenarios that could be developed once accurate data has been obtained and implemented into the model. The model produces logical results with consumer type groups located in reasonable OAs, and their behaviours reflect the rules set out in the model. Although the model is in its early stages, it has demonstrated how it could be used for scenario testing, especially once consumer expenditure has been added along with other types of grocery demand. The model does generate sales values for each transaction using the loyalty card dataset's analysis and standard deviation, but this has yet to be included in the analysis due to the early stages of model development in other areas.

The model developed in this study provides retailers with insights into their current customers and the spatial and temporal behaviours in their grocery transactions. The model allows testing for 'what-if' scenarios at the individual level, capturing the nuanced behaviours of their customer base. This type of model is the first of its kind in the context of grocery location analytics and has generated interest from clients beyond Sainsbury's.

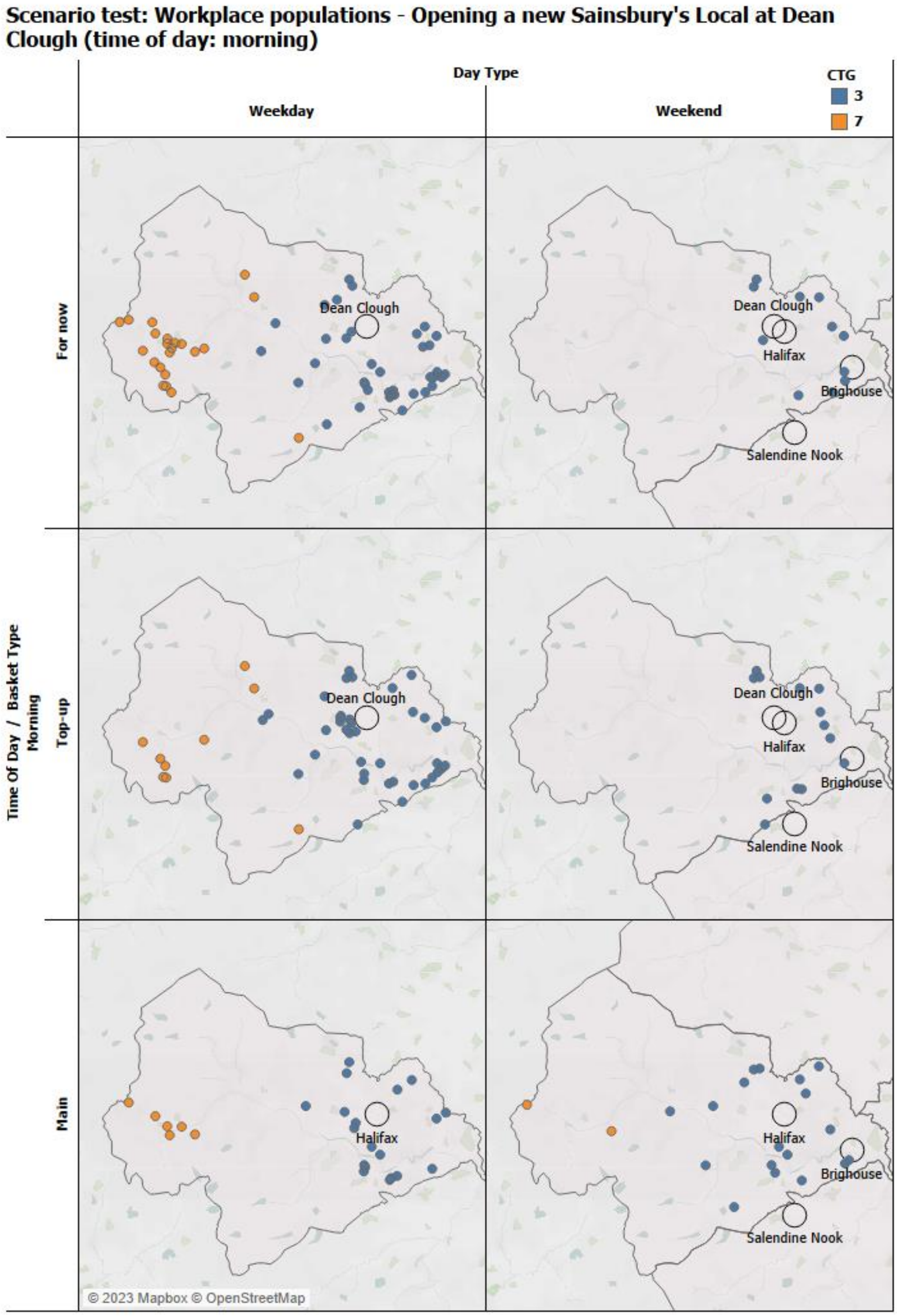



**Figure 6.24** Scenario test: Work-based store addition results (morning): black outlined dots are the stores visited, and the smaller dots are OAs where customers reside. The Dean Clough store is a new addition and is located on the Dean Clough complex.

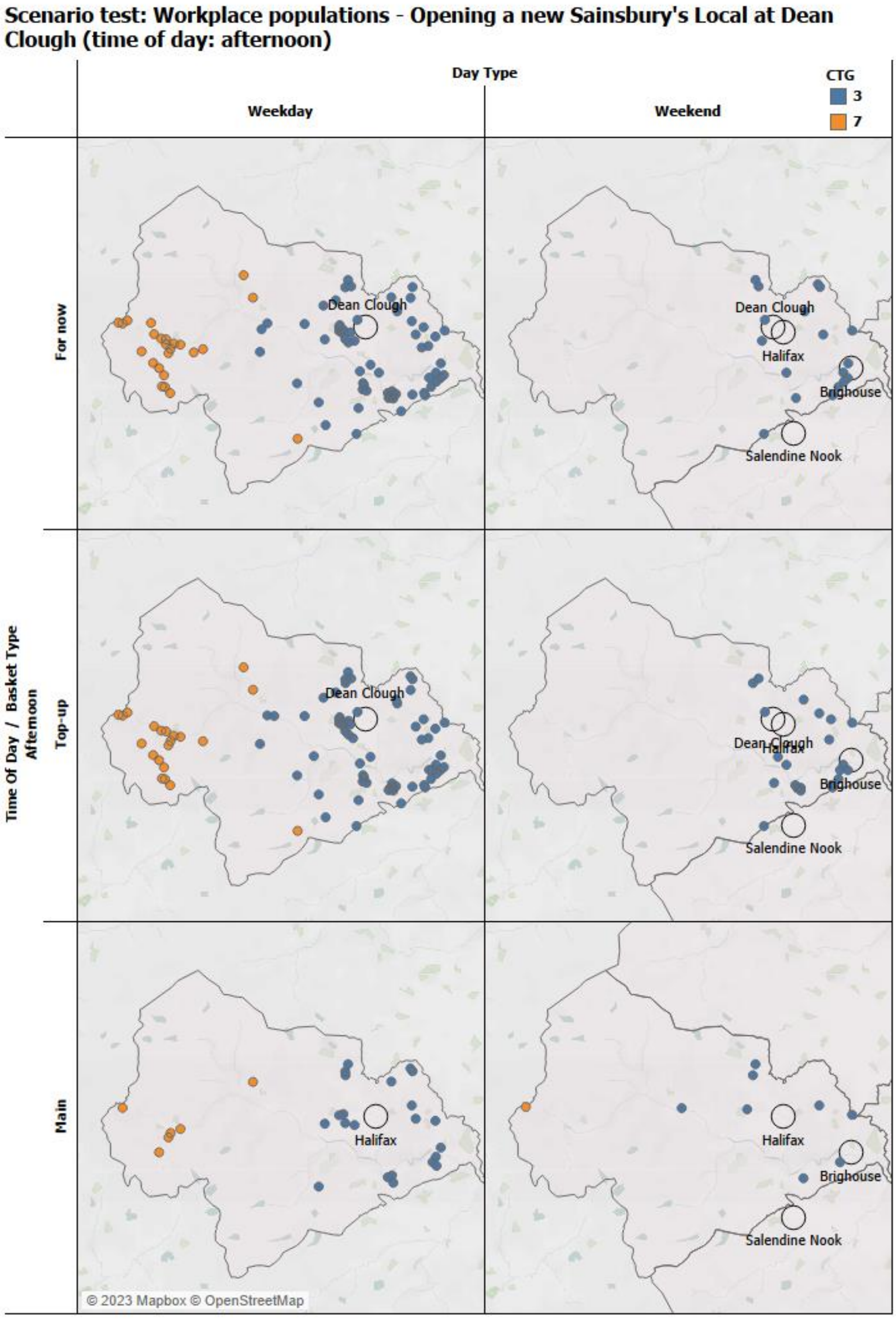



**Figure 6.25** Scenario test: Work-based store addition results (afternoon): black outlined dots are the stores visited, and the smaller dots are OAs where customers reside. The Dean Clough store is a new addition and is located on the Dean Clough complex.

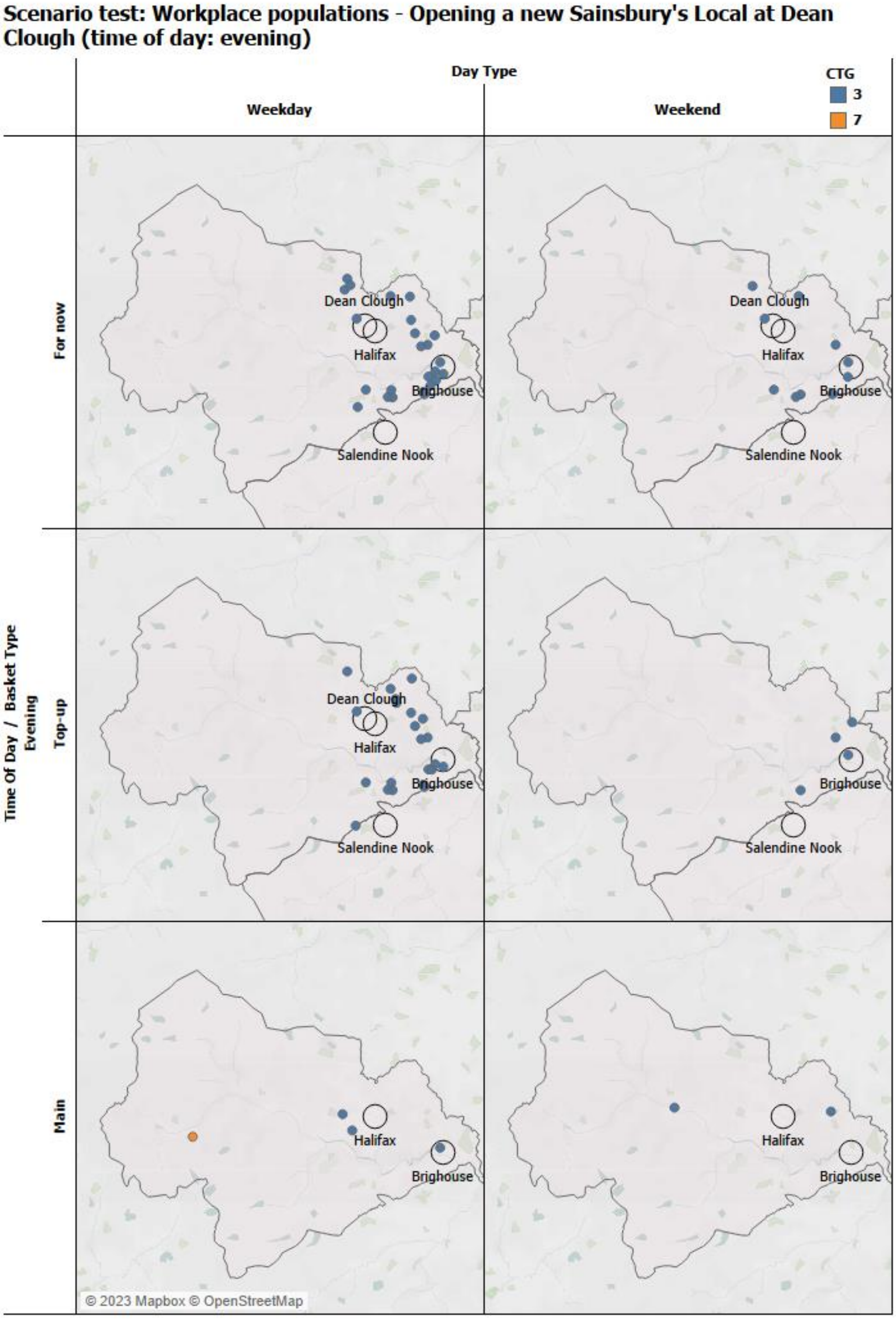



**Figure 6.26** Scenario test: Work-based store addition results (evening): black outlined dots are the stores visited, and the smaller dots are OAs where customers reside. The Dean Clough store is a new addition and is located on the Dean Clough complex.

## 6.4 Chapter 6 summary

The first part of this chapter focused on the temporal aspect of modelling consumer behaviours. A decision tree structure was developed to model the probabilities of different customers transacting based on the consumer type groups they belong to. Using the decision tree approach in an IBM, along with the roulette wheel selection function, the model could accurately simulate known behaviours of grocery customers (section 6.2.2). The developed model is a first attempt at refining spatial models used within the context of grocery retail, incorporating the individualised behaviours of known customers, which is missing in current spatial models, such as SIMs. As this thesis presents a novel approach, the KISS principle guided the model's creation, allowing for a logical development process that is entirely reproducible.

The second half of this chapter focused on modelling the spatial store choices of the simulated customers. Three store choice methods were presented, with each increasing in complexity, following the KISS principle. The first store choice method was inspired by Tobler's First Law, in which customers always transacted at their closest store. This naïve approach was used to test whether the model could allow customers to make store choices by following simple rules. As discussed in section 6.3.2, some behavioural rules were in place that linked customers' basket types to store types. This approach demonstrated that the model's spatiality function was working correctly as different stores were chosen based on consumers' OAs and basket types. As customers chose their closest store, this spatial approach captures residential-based demand well but overpredicts transactions at convenience stores due to the lack of stochastic behaviour in-store choice.

The second store choice model used distance bands directly derived from the consumer type group's store choices. This approach was used to see the extent to which the loyalty card data can be used to simulate customers' temporal and spatial behaviours. Whilst this approach captured variation in the customer's store choice behaviour, the results did not represent the observed customers. This modelling method may have potential with refinement and added complexity, such as incorporating an element of store attractiveness like that within SIMs. As the model presented in this thesis did not provide the customer with a direction to choose a store, some store choices were not representative of the observed data.

The third store choice method used the 2011 Census to relocate customers at different time periods to represent workplace populations. A small case study of an OA was presented in this thesis, but others had been tested too. This method was complex to model as assumptions had to be made. The loyalty card dataset does not provide information on *who* the customers are in the dataset; nothing is known about their demographics, employment status, or family size. The census does not provide information on who owns a Sainsbury's Nectar card; therefore, direct links cannot be made.

Nevertheless, loose assumptions based on the proportions of customers within an OA were used to estimate which customers will likely leave the OA for work. The output from this method captured simulated consumer store choices successfully, representing consumer spatiality and temporality compared to the observed dataset. Although the census was captured seven years before the loyalty card dataset, the model could still represent work-based grocery demand.

Finally, section 6.3.6 provided an example of how the model could be used for store location testing once a sound data source for estimating non-residential demand is available. The example provided context around the agent's behavioural rules if a new store was added in a new location. It is anticipated that future model developments beyond this thesis will be able to incorporate alternative data sources suitable for this type of IBM in grocery retail. Overall, this chapter has provided a fully reproducible model that is driven by a loyalty card dataset containing different consumer type groups. This model provides a foundational framework that can be expanded by applying new data sources, validated against other data, and appropriate for use within retail location planning for site location assessments.

# Chapter 7 Study findings, limitations, recommendations, and concluding comments

## 7.1 Introduction

The overarching research aim of this thesis is to develop a modelling framework that focuses on incorporating loyalty card-linked transaction data using IBM. The model aims to support grocery retailers' location planning and impact assessment process, providing a unique and enhanced methodology that considers their known customers' behaviours at the individual level. Current methods used in retail location analytics rely on top-down modelling methods, yet an abundance of individual data is being captured daily about customers. Therefore, this novel methodology sets the scene for developing and refining spatial models in grocery analytics.

The overarching research aim is split into three parts as outlined in section 1.2:

1. *To present a review of the historical and recent changes in consumer behaviour in the context of grocery retail planning and the tools used in location planning analytics.*
2. *To identify key segments of consumer type groups who exemplify particular consumer behaviours regarding when, where, how and what they purchase from grocery stores.*
3. *To design an individual-based model based on known customer behaviour observations, incorporating temporality and spatiality using a hybrid microsimulation and agent-based modelling methodology.*

These three aims are met with the findings presented in Chapter 2 to Chapter 6 via the completion of the objectives also outlined in section 1.2:

1. *To explore the literature around British grocery retail over the past decades, identifying key changes that impacted consumer behaviour and the key indicators of consumer behaviour.*
2. *To articulate the benefits and limitations of current grocery location planning analytical tools, including the potential data sources available.*
3. *To investigate the loyalty card linked transaction dataset provided by Sainsbury's, identifying general behaviours of their customers in relation to their store choice behaviours in West Yorkshire.*
4. *To segment the observed customers into consumer type groups based on their key behaviours regarding transaction frequency, store and channel choice, and purchase purpose, i.e., basket type.*
5. *To perform data mining on the customer segments, identifying the probability of making a transaction at any given time based on their linked loyalty card transactions.*
6. *To design and build a prototype individual-based model of consumer behaviours suitable for simulating their transactions by day type, time of day, channel, and basket type.*

This chapter achieves the final objective, bringing the study to a close:

7. *To discuss how the model's functionality and framework is beneficial in retail location analytics, and how it can be used and further developed by retailers and location professionals.*

Recapitulating, this thesis seeks to make three contributions: (1) A contribution to the refinement of spatial models in retail, (2) a contribution to the literature surrounding spatial modelling; and (3) a contribution to the field of location analytics academically and professionally.

This chapter discusses the research findings presented in this thesis, particularly in chapters 4, 5, and 6. Drawing on these discussions, section 7.4 concludes this study, highlighting the contributions of the thesis, the study's limitations, and recommendations for future work.

## 7.2 Inferring consumer typologies

This section focuses on the research produced in Chapter 4 and Chapter 5, in which preliminary analysis of the loyalty card-linked transaction dataset is presented, and consumer type groups were identified through the segmentation of individual customers within that dataset. The identification of such consumer type groups and separation based on specific patterns of behaviour satisfies aims 1 and 2 and are explored below.

### 7.2.1 Study findings

To begin, the first part of the study focused on the store network and loyalty card data provided by Sainsbury's (Chapter 4). Their extensive store network across West Yorkshire results from the supply-side developments within grocery retail, as discussed in section 2.1. Many of Sainsbury's largest supermarkets are located within a variety of areas that aligned with the Planning Policy Guidance changes (Guy, 1997; Wood et al., 2006), with stores located in a mix of 'out-of-town', 'within centre' and 'edge-of-centre' sites (Figure 4.2). Their convenience store network is also a result of policy updates and the complex range of changes in consumer demand (Wood and Browne, 2007), with 71% of their stores in West Yorkshire being associated with their "Local" fascia (Figure 4.4). Their larger stores that act as online delivery hubs tend to be located next to the most accessible road networks. They are mainly the superstores built before the PPG6 changes that limited store developments outside town centres (Wood et al., 2006).

By delving into the evolving patterns of grocery consumer behaviours (as reviewed in Chapter 2), from shopping at local greengrocers to shopping from the comfort of a sofa, it is clear that there is no *one size fits all* and that not all customers behave the same in their store and channel choices. It is integral for retailers and location planners to understand their customers' unique and distinctive behaviours, as these behaviours can profoundly impact store economic performances, especially in different contexts of space and time.

Current methodologies which utilise top-down approaches (Chapter 3) rely on disaggregating consumer demand and making inferences about consumer behaviours. The central assumption being that all consumers behave homogenously with no personal preferences. It is inferred that a store's attractiveness, surrounding population numbers and competitors account for its performance. This type of approach omits the main factor influencing store performance and revenue generation, and that is the individual *people.*

As described in Chapter 5, the loyalty card-linked transaction dataset was segmented in a two-step approach. Firstly, the consumers were separated based on their shopping channel choices (i.e., 'online', 'in-store', or 'multi-channel'). The 'online' and 'multi-channel' shoppers were assigned as unique clusters as they contained only 2% and 4% of the consumers in the dataset, respectively. The remaining consumers ('in-store' only channel) were segregated using the k-means clustering algorithm, and five distinctive clusters were classified. In total, seven consumer type groups were identified.

The low percentage of customers within the online and multi-channel consumer type groups demonstrated a curious finding that although retailers have progressed into the e-commerce era (Birkin et al., 2017; Beckers et al., 2022), most consumers continue to transact in-store. However, as the loyalty card-linked transaction dataset captured only transactions during 2018, consumer behaviour has shifted more towards online shopping since then. For example, the literature identifies a variety of driving forces for e-commerce growth, such as a shift towards convenience shopping (Sleeman, 2022) and the major impact of the Covid-19 pandemic on consumer channel choice (Meister et al., 2023). As the pandemic occurred after the dataset was recorded, the behaviours of the same customers may differ today. Additionally, the lower participation in e-commerce by these customers could be an artefact of these consumers residing within West Yorkshire. Research shows that customers within Yorkshire and the Humber transacted online the least nationally in 2018 (Kirby-Hawkins et al., 2018; Hood et al., 2020), which is therefore supported by the findings of this thesis.

The analysis of the in-store consumer type groups 3 to 7 presented diverse behaviours, with differing consumer demands and transaction habits. For example, distinctions in consumer type behaviour were observed across basket types (‘for now’, ‘top-up’, and ‘main’), time of day (‘morning’, ‘afternoon’, and ‘evening’), and day type (‘weekday’ and ‘weekend’) (Table 5.1). The different behaviours of these in-store shoppers could be due to several factors, such as brand loyalty (Wood and Browne, 2007), consumer lifestyles (De Kervenoael et al., 2006), and store accessibility (Vanhaverbeke and Macharis, 2011).

Consumer type group 3, “Convenience Suburbanites”, primarily transacted at Sainsbury’s stores for ‘for now’ and ‘top-up’ baskets on weekends and weekdays. The customers within this group rarely performed ‘main’ baskets and resided in the widest variety of geodemographic areas (Table 5.2). The lack of ‘main’ basket transactions performed by this consumer type group suggests that they may lack brand loyalty to Sainsbury’s and only transact at Sainsbury’s for smaller basket types due to store proximity to their homes. These customers may perform their ‘main’ baskets at other retailers due to budget (Retail Connections, 2023), culture (ABPL Group, 2016), or brand preference (Popkowski Leszczyc et al., 2004). Potentially for the same reasons (budget, culture and brand preference), those in consumer type groups 4, “Small Town Shoppers”, 5, “Supermarket Weekend Warriors”, and 6, “If I need it, Sainsbury’s has it!” mostly transact at Sainsbury’s for ‘main’ baskets. However, these three groups varied in their day type, time of day, and shopping frequencies; although there is some overlap between the basket types they performed, their temporalities differed. Consumer type group 7, “Hungry and far From Home”, stood out from the other in-store groups, as these customers travelled particularly far for their transactions, primarily being ‘weekday’ ‘morning’ and ‘afternoon’ for ‘for now’ and ‘top-up’ transactions. These customers are likely commuters purchasing groceries near their places of work, aligning with research around trip chaining (Primerano et al., 2008) and tour-based transactions (Daisy et al., 2018).

These distinctions demonstrate the versatility of consumer type behaviour at the micro-level, providing invaluable insight into the consumer “mind”. Although consumer behaviour (and human behaviour in general) is unpredictable at its core, the identified consumer type groups demonstrate that a level of aggregation can be applied to individual consumers to form a foundational understanding of the nuances in consumer type behaviour. In this way, the consumer type behaviour drives the modelling approach presented in Chapter 6, meaning that the temporal and spatial aspects of the model are instinctively based on actual consumer type behaviour, rather than inferred.

### 7.2.2 Limitations

It is acknowledged that whilst these findings are insightful in understanding and categorising consumer behaviour, there are various limitations to the data provided. Firstly, the time the loyalty card-linked transaction dataset was recorded was prior to significant world events such as the Covid-19 pandemic and only provides a snapshot of consumer behaviours during a 3-month period in 2018. The data does not account for seasonality which could impact the behaviours of those within the consumer type groups; for example, a new consumer type group could be identified who only transact at Sainsbury's during the Christmas holiday season (Newing et al., 2013).

Secondly, there are uncertainties associated with loyalty card transaction data, as discussed in section 3.3.1, in which customer addresses are not always updated, not all customers scan their loyalty card at check out, and customers may transact outside of the study area (Rains, 2019; Lloyd and Cheshire, 2019). Even where customers do scan their loyalty card, we cannot be certain that we are catching all Sainsbury's-based transactions for that household. Multiple members of the same household may be responsible for grocery shopping and use multiple loyalty cards.

Thirdly, the data used for this study only represents the behaviours of those with a Nectar card, omitting customers who may be frequent and loyal shoppers but do not partake in the loyalty card scheme. Non-loyalty cardholders may make up a significant proportion of all customers at Sainsbury's, and as they are not included in the consumer typologies, any model output will be missing the transactions by those individuals. Therefore, the model is limited to only simulating the behaviours of those with a loyalty card.

Finally, the clustering process did not incorporate demographic factors when identifying consumer type groups. Ideally, there would be a geodemographic link between the consumer type groups and the areas that customers reside in. Placing new customers within an IBM would be simple if each consumer type group had a particular association with a different demographic classification. For example, if customers in consumer type group 1 primarily lived in rural areas, new customers could be assigned to relevant OAs in the IBM. However, demographic elements were not incorporated into consumer type groups as they added little value to the clustering process. Most of the Sainsbury's customers in the dataset resided in OAs categorised within OAC groups 5 and 6 (Table 5.2). Therefore, the k-means clustering algorithm could not find associations between consumer type groups and OAC groups.

Despite these limitations, the loyalty card-linked transaction dataset provides invaluable insights into the behaviours of known customers, to the extent of identifying seven unique consumer type groups. These groups were suitable for developing an IBM that focuses on modelling spatiotemporal consumer behaviours, meeting the aims set out in this thesis. Therefore, the data limitations are not impactful for the purpose of developing a prototype IBM, and the dataset still provided a fascinating insight into consumer habits.

### 7.2.3 Future work and recommendations

To expand upon the work presented in Chapter 5, the most sensible and logical area to enhance the model would be to include extra data that complements the loyalty card-linked transaction data. Knowing the past actions of a set of customers provides a great wealth of information. However, there are significant attributes about these customers that are unknown and would be integral for future development work. To fully understand the customers used within the study, demographic information such as age, sex, household size, employment status and car ownership would provide a much insight into how lifestyle impacts transactional behaviour (Kirby-Hawkins, 2016; Hood et al., 2021). This would support the connection between population data and customer behaviour using this information. The simplest way to obtain this data would be to conduct customer surveys for Nectar cardholders willing to share their personal information and linked Nectar account ID, and perhaps their relationship with Sainsbury's, such as whether they consider Sainsbury's as their 'main' supermarket of choice (Benoit and Clarke, 1997). However, obtaining such data leads to ethical issues and data privacy limitations, which are always obstacles in these studies (Polding and Dieguez, 2021). Additionally, capturing a variety of Nectar card holders who transact at Sainsbury's is also a challenge, as only a particular group of customers may have the time, interest or willingness to complete such a survey.

Other public data sources, such as the 2021 Census or data held by the Consumer Data Research Centre, would provide these insights, and a more general assumption could be made regarding the customer's demographic. Studies that have previously utilised census and other data to create synthetic populations would greatly enhance populating a future agent-based model that incorporates behaviour beyond customer home location and store location journeys (Wickramasinghe et al., 2020; Wu et al., 2022). Linking the customers within this study to a synthetic version would optimise the model's accuracy and could then be validated against this study's output. If the Nectar card data were accompanied by customer surveys and census-based synthetic population data, it would be possible to identify relationships between customer shopping preferences, such as budget, distance between frequently visited locations and stores, online app usability and more.

Finally, a follow-up study using a different time period for the same subset of customers would be informative. If the same subset of customers were forced into the same consumer type groups presented in this thesis, how would each group's consumer behaviours differ between the 2018 dataset and more recent data? Additionally, if the clustering process was re-run from scratch using more recent data, how would the consumer type groups differ post Covid-19? Could a whole new consumer type group with a new set of behaviours have emerged since the pandemic?

These suggested developments are substantial, and obtaining such a variety of data is not always available due to costs, proprietary nature, time requirements, and ethical problems. However, any small developments would help enhance the framework provided in this thesis, which provides the groundwork to do so.

## 7.3 Building an Individual-based model of consumer store and channel choice behaviours

This section discusses the findings from the model development process presented in Chapter 6, focusing on incorporating temporal and spatial behaviours of the defined consumer type groups, which achieved aim 3.

### 7.3.1 Study findings

The first element of the model-building process was to incorporate temporality and refine this parameter to accurately reflect consumer type group transaction behaviours. As outlined in section 6.1.2.1, the model's foundation was built upon the development of a decision tree structure. The decision tree utilised the probabilities calculated in section 6.1.4 to drive the consumer type group transaction decisions once inserted into the model.

Previous attempts at modelling consumer behaviour, such as that of Sturley et al. (2018), fell short in successfully reproducing the temporal behaviours of known customers, as critically discussed in section 6.1.3 Their work produced a proof-of-concept model that did not correctly reflect the probabilities of certain transactions occurring and did not offer the agents an option to not transact. Despite these drawbacks, their work presented a logical process of building an IBM in the context of grocery retail but required refinement. The IBM created in this thesis was inspired by Sturley et al. (2018) by utilising probabilities for consumer transaction decisions and was expanded upon by creating a decision tree. Using the decision tree and weighted roulette wheel selection function, consumers were assigned behavioural choices based on their consumer type group and given a degree of flexibility in what transaction they carry out and when. The application of the roulette wheel selection provides the consumers with a choice of transaction behaviour at each decision node, thus incorporating the stochastic behaviour of consumers. The stochasticity element allows the model to capture dynamic transaction choices within each consumer type group. This is a novel application of a decision tree framework within the models used in the location analytics sector and provides a unique insight into modelling consumer temporality.

This research fills a gap in the academic literature and industrial practice. Tools and methods such as SIMs are robust and well-studied, allowing modellers to predict spatial flows and understand underpinning factors. However, as noted by Rowe et al. (2022), more progress has yet to be made in developing these models to incorporate an element of consumer heterogeneity. The individualised behaviours of consumers, notably in their temporal and spatial aspects, are incredibly valuable to understand for retailers, policymakers, and researchers. Incorporating the oddities and patterns of individual customers in SIMs is a challenging feat, even when additional methodologies are considered (Rowe et al., 2022). Therefore, the research presented here delivers a modelling framework that performs at the individual level, providing the ability to simulate individual customers based on observed, known behaviours. The model has been developed entirely on actual customer store and channel choice behaviours, notably focusing on the relationships between transaction temporality, channel, and basket type.

The second part of developing the IBM framework focused on the spatial aspect of consumer store choice behaviour. As highlighted in section 3.2.3, incorporating spatiality is a notorious challenge and requires vast data to simulate those behaviours accurately. The study has demonstrated how far a single loyalty card-linked transactional dataset can be used to develop an IBM and its limitations in locating customers temporally. The model fully simulates the temporal aspect of customer behaviour but requires further development regarding forming agent store choice rules. Multiple attempts at incorporating spatiality solely from the transactional dataset were included, finding sticking points when modelling non-residential grocery demand (as discussed in section 0).

Section 6.3.5 provided an example of utilising census data to assign customers to workplaces depending on the flows between OAs; this was achieved by using residential and workplace data from the 2011 Census at the OA level. However, the critical drawback of using census data is that it is seven years out of date compared to the transaction data, in which some customers will have likely moved homes or changed places of work over that period. The 2021 Census was considered as an updated data source; however, it created a new problem. The 2021 Census recorded population data after the Covid-19 pandemic, in which consumer working behaviours had changed since 2018 as remote working became incredibly popular for most office-based jobs (Adekoya et al., 2022). Therefore, no substantial data are available for the particular year of 2018 that can link customers to places of work, thus hindering the ability to simulate workplace-based grocery demand accurately.

Nevertheless, the third store choice method implemented in the model successfully captured an element of workplace-based grocery demand, and residential-based grocery demand. This was achieved by relocating the consumer agents in the model at different times of the day, replicating the working week. The output from this store choice method indicates the potential that census, or other types of workplace data can be used in conjunction with loyalty card data to simulate consumer store and channel choice behaviours. This was demonstrated via the example scenario test presented in section 6.3.6, in which a new store was placed in an established place of work and modelled the consumers interactions with this store. These consumers transacted at this store in the model at different times of the day, reflecting their consumer type group behaviours.

The research in Chapter 6 presents the successful development of an IBM that simulates the spatiotemporal transactional behaviours of different consumer type groups using loyalty card-linked transaction data. The research in this chapter provides a comprehensive process of creating such models, identifying the potential of using decision trees with loyalty card data, and the challenges of modelling store choice. The modelling techniques were chosen to fully incorporate the nuanced behaviours of grocery customers, providing location analysts with a tool to better understand their customer's behaviours. By modelling the customers at the individual-level, those in grocery location analytics can use these model outputs to support their location-based decisions via scenario testing, as changes can be made in the model on the supply and demand sides.

### 7.3.2 Limitations

Due to the complex nature of building and designing IBMs (as discussed in Chapter 3), various challenges were faced while creating the model presented in this thesis.

Firstly, the data provided for this thesis only contains loyalty cardholders and their past transactions at Sainsbury's stores. Therefore, only the interactions between the customer and Sainsbury's stores are known, and their behaviours elsewhere cannot be derived. Modelling new customers is incredibly difficult, as they could belong to any consumer type group. Assumptions could be made by assuming those within the same OA would behave similarly, but there is no way to validate these postulations.

Secondly, no demographic data is available for the customers in the model. Linking these customers to other datasets, such as the census or transport data, is challenging but would support the ability to model other types of non-residential-based grocery demand. If we knew which customers had children in school, were retired, worked from home, or were university students, assumptions could be made to help identify other grocery demand types. If consumers could be linked to other data sources, agents in the model could be mobile, moving from point A to point B. Currently, the model cannot account for trip chaining transactions, like those at Leeds Train Station (section 4.2), which form a large part of grocery success in today's convenience-based society.

Thirdly, model validation is limited as data is only available to cross-validate the model's results within the loyalty card-linked transaction dataset. Sensitivity analysis and comparison to the observed data are sufficient, but ideally, the model could be compared to similar outputs that use a bottom-up approach. As this is the first attempt at such a model, it is challenging to find comparable outputs.

The fourth limitation is that the model only considers the closest store depending on the transaction scenario. This approach may be logical for some customers, but as the model does not incorporate competitor supermarkets, the logic requires more complexity. If the model were to incorporate other retailers, a more sophisticated store allocation methodology would be required, similar to that found in SIMs, where the customer makes a trade-off between store attractiveness and distance. In the model's current form, it is limited in its ability to consider competitors. It thus assumes that these agents are Sainsbury's customers – a logical assumption given that they have been derived from the Sainsbury's loyalty card dataset.

Finally, the most significant outstanding limitation of the model is its limited consideration for competition. This has been touched upon throughout the thesis and is an acknowledged limitation within the scope of this research. The UK grocery market offers consumers a variety of brand options (Clark et al., 2022). As a result of convenience culture, increasing living costs, and expanded choice, Barclay's Bank (2024) found that two in five customers are shopping at multiple supermarkets. Customers no longer transact solely with one grocery retailer; obtaining data to capture the variety of retailers they interact with is incredibly difficult at the individual level. Other retail studies, such as Heppenstall et al. (2006), built a hybrid agent-based model of the petrol market in West Yorkshire. Their work uses an agent-based methodology to simulate the openings and closures of petrol forecourts based on simulated customer expenditure. Customer behaviour is incorporated using a spatial-interaction model to determine the geographical source of demand. Like the research presented here, consumer mobility is based on home-work networks as it is the most readily available data with an origin-destination structure. Heppenstall et al. (2013) utilised a dataset containing daily petrol prices across various forecourt retailers over a 3-month period. The dataset was used to control the price parameters of the model and allowed for the incorporation of competitor retailers. If such a dataset existed for the grocery market, the model presented in this study could begin to consider consumers' behaviours outside of Sainsbury's. Competitor data would allow for a complete comprehension of the extent to which consumers behave differently when purchasing groceries from each other.

These model limitations indicate the core challenges that bottom-up models face, especially when working with a limited behavioural dataset. As consumer behaviours are personal, it is challenging to quantify behaviours representative of reality outside of what is already known. Despite these limitations, the following section provides suggestions and directions for future work in this study.

### 7.3.3 Future work and recommendations

To further expand the work presented in Chapter 5, key areas of improvement should be considered. The first improvement includes the incorporation of spatial data. The most readily available dataset that could be implemented in this model is the 2021 Census. Future work could take a larger sample of past consumer transactions for the same 2021 time-period to create a model of consumer store and channel choice behaviours. This would provide a more accurate representation of consumer movements than the previous census. Additional data from the census should also be considered, such as those mentioned in Waddington et al. (2019) that account for other types of grocery demand, for instance, school-based demand (representing parents being near schools during dropping off and picking up children), university demand, and leisure demand.

The model could also be enhanced by utilising mobility data to help locate optimal store locations, which would inherently alter customer behaviour, as demonstrated in the small case study presented in section 6.3.6. Mobility data provided by companies such as Huq can have a two-fold application in this model. Mobility data records how consumers move and interact with the places around them. It is ideal for observing customer movements and provides a better understanding of consumer trends. The first data enhancement for this model would be to link the Nectar cardholders to the mobility data to gauge a loose insight into the types of movements the customers could be making (Nijenhuis et al., 2022; Kaziyeva et al., 2023). However, linking individual customers to the mobility data would be laborious. The second usage would be to enhance the store choice methodology further. The current model uses travel-to-work census data to place customers at different locations at different times. If a causally informed synthetic population was applied to the model, mobility data could be used to have these agents travelling for various reasons other than work. Additionally, this data helps identify customer hotspots that help place stores in the model for scenario testing. Obtaining this data for a large study area can be expensive, though, unless it is provided in collaboration with academic institutions (such as the Urban Big Data Centre or CDRC).

The most aspirational recommendation would be calculating the customer's probability of transacting elsewhere at non-Sainsbury's supermarkets. Currently, the model only accounts for transactions at Sainsbury's, which limits the ability to consider new customer behaviours. For example, a customer may transact at Sainsbury's three times a week for 'for now' and 'top-up' transactions and spend around £20 a week. This customer will likely transact at another retailer for the remainder of their weekly grocery transactions. Using the LCFS, a researcher could calculate the likely amount of money the customer would spend on groceries and link that value to a basket type. In the example, the customer may be expected to have £50 a week left over that is being spent at other supermarkets that are not Sainsbury's. That £50 could be spent on a ‘main’ basket type at the customer’s local supermarket or online with a competitor. What would happen if Sainsbury’s opened a new supermarket near that customer’s home or made their e-commerce offer more attractive to that consumer? Would that customer begin transacting for their ‘main’ basket-type shop at Sainsbury’s? This information is unknown and is incredibly difficult to calculate accurately. If these data were available, the model could simulate various consumer type groups and their interactions at non-Sainsbury’s retailers.

The next recommendation would be to attempt to incorporate and apply a SIM-like model for agent store location choices. As demonstrated in this thesis, the KISS principle can only be used to an extent in recreating consumers’ complex and sophisticated behaviours. If the modelled customers had further preferences beyond distance in their rules, they would have the capacity to make further choices, such as choosing a store that offers a broader range of services over the closest store.

In conjunction with customer, population, and mobility data, an integral dataset would be the Retail Place Boundaries dataset created by Geolytix (geolytix.com, 2020). These data provide the exact boundaries of retail places across the UK, covering train stations, retail parks, shopping centres, urban centres, cities, towns and much more. As part of their GeoData packs, they also provide retail points of interest, including various retail brands and their exact coordinates. Using a combination of these data, a store retail score can be calculated and used to determine a store's attractiveness depending on what is located nearby; if a supermarket is located near a train station or football stadium, would this store attract more multi-purpose transactions? Or would this help us understand the times or days of the week when demand will likely be present near that store? This would allow for further model expansion and support the incorporation of competitor stores, recreating the benefits of SIMs whilst maintaining the temporal benefits of IBM.

These datasets and logic enhancements would create a more rounded model for understanding consumer behaviour and allow for a complete agent-based model approach. Using a synthetic population, as suggested in section 7.2.3, the need for distinct individuals based on empirical data is less critical if we can link these populations to customer-type groups. Then, a more complex method of assigning store choice rules could be implemented, including a trade-off between store proximity and attractiveness within SIMs. An agent-based model can allow agent interactions that adapt as their environment changes by modelling agents' temporal and spatial behaviours using the synthetic population and Nectar card data.

A more developed modelling framework would allow those undertaking a Retail Impact Assessment to fully account for individuality between consumers and their interactions with new and existing retail stores. With future model developments incorporating retail place locations and their attractiveness and competitors, the method presented could provide property developers with a more accurate tool for measuring retail impact. Whilst current SIM approaches in RIAs adequately allocate expected demand, they cannot account for temporal variations in demand based on consumer home locations.

Whilst these sections identified limitations and areas for further refinement, this thesis has met its stated aims. It adds considerable value to the academic theory and practical model building within this context, as summarised in the next concluding section.

## 7.4 Concluding comments

Retail location analytics has been a long-studied research topic in the field of geography. Past studies have heavily focused on the spatial aspect of consumer behaviour, drawing inferences from various demand layers and linking these to store locations. These studies have primarily been used to estimate revenue for grocery retailers at the store level based on the population catchments. However, as discussed throughout this thesis, consumer grocery behaviours are individualised and heterogeneous, which is not captured by current spatial models. Therefore, this thesis set out to design an IBM framework that incorporates the spatial and temporal behaviours of grocery customers.

Through the analysis of rarely accessed loyalty card-linked transaction data, seven unique consumer typologies were identified based on customer purchasing patterns within the Sainsbury's store network, specifically relating to transaction temporality, channel, mission, and store choice. The analysis of these customers alone provides an integral insight into the ways in which consumer behaviours differ, notably regarding transaction frequency, willingness to travel, the types of baskets purchased, and the interrelations between these factors.

The study then utilised these observed consumer type groups to build an IBM, using methodological approaches found in decision trees, MSMs, and ABMs. The model built provides an early-stage framework of how an entirely data-driven IBM could be developed that is representative of actual consumer transaction behaviours in grocery retail. The model presented in this thesis provides a step-by-step process that is fully reproducible and can be further developed and expanded on in future research, particularly pertaining to adding more complex spatial consumer behaviours. The reproducible workflow of this research is presented in Figure 7.1.

Future work based on the research presented in this thesis is primarily concerned with the addition of other data sources. The model presented heavily used loyalty card-linked transaction data, providing integral insight into Sainsbury's consumer behaviours. However, further information, such as demographics or consumer movement data, would enhance the model presented to better capture all types of grocery demand. Additionally, competitor data would benefit the model development to simulate consumer store choices beyond Sainsbury's. Finally, once a more sophisticated store choice method has been identified, future work should focus on thoroughly validating the model via rigorous scenario testing, especially as consumer behaviours continue to evolve with the proliferation of e-commerce. Future scenario tests have been recommended throughout this chapter.

As alluded to throughout the thesis, multiple challenges are faced when creating IBMs of individual consumer behaviour. These challenges are primarily related to the inaccessibility of alternative data sources that would support the loyalty card dataset in modelling spatial consumer store choices. Obtaining personal and demographic data about the customers is notoriously difficult due to GDPR guidelines. Linking customers to other data sources, such as mobility data, is also a unique challenge. For the IBM in a grocery context to fully simulate consumer behaviours, other research fields are required, such as transport. Many transport studies focus on individuals' mobility rather than individual grocery customers' mobility. A collaborative research study with transport experts could support the enhancement of the IBM of grocery consumers.

The challenges and limitations discussed in this chapter do not suggest that IBMs are inadequate for modelling consumer behaviours. In fact, with access to the appropriate data, expertise from other disciplines, and data to validate against, IBMs can further enhance and refine the spatial models used in consumer spatial analytics. Developing such models is not an easy win for retailers but would provide them with a bespoke modelling tool that can finally help them better understand and predict the individualised behaviours of grocery consumers in an ever-evolving convenience-based climate.

The research presented in this thesis contributes to three key areas. Firstly, it provides an IBM framework for future studies to expand upon the spatial models in retail. Secondly, the exploration of the customer data and implementation into an IBM contributes to the literature within spatial modelling and grocery spatial analytics. Thirdly, it contributes to the wider field of location analytics academically and professionally. The research outputs in chapters 4, 5, and 6 are suitable for journal publication, as discussed in section 1.4, with the first paper regarding consumer typologies being ready for submission.

Ultimately, this thesis has achieved the core aims set out in section 1.2 by attaining all seven objectives. The research presented in this thesis has successfully designed and built a working IBM that simulates individual customers based on their consumer type group, capturing their spatial and temporal transactional behaviours driven directly by loyalty card data. There is still considerable work to be done to enhance these models for them to be ‘retailer ready’. Nevertheless, this thesis presents the foundations of a modelling framework that successfully simulates spatiotemporal consumer behaviours, with *individuals* at its core.

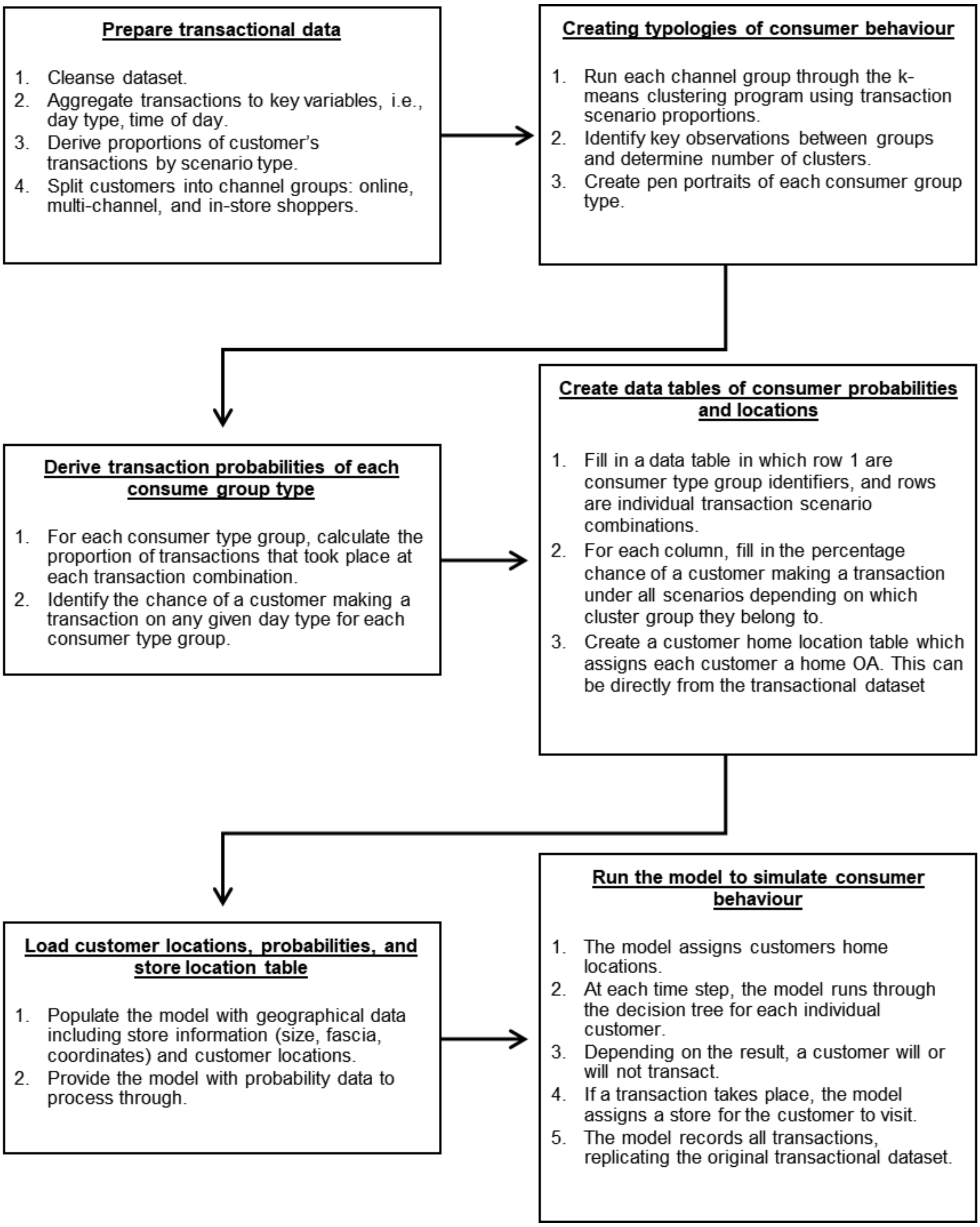



**Figure 7.1** Reproducible workflow of the thesis from segmenting customers, mining behavioural rules, and inserting into the IBM.

# List of References


ABPL Group 2016. Retailing to the British Asian. *Medium*. [Online]. Available from: https://medium.com/@abplgroup/retailing-to-the-british-asian-supermarkets-or-hyper-local-stores-4483c8070edb.

Adekoya, O.D., Adisa, T.A. and Aiyenitaju, O. 2022. Going forward: remote working in the post-COVID-19 era. *Employee Relations: The International Journal*. **44**(6), pp.1410–1427.

Akehurst, G. 1984. 'Checkout': The Analysis of Oligopolistic Behaviour in the UK Grocery Retail Market. *The Service Industries Journal*. **4**(2), pp.189–242.

Anderson, W.P., Chatterjee, L. and Lakshmanan, T. r. 2003. E-commerce, Transportation, and Economic Geography. *Growth and Change*. **34**(4), pp.415–432.

Axtell, R.L. and Farmer, J.D. 2022. Agent-Based Modeling in Economics and Finance: Past, Present, and Future. *INET Oxford Working Paper*. (No. 2022-10).

Bae, J.W., Paik, E., Kim, K., Singh, K. and Sajjad, M. 2016. Combining Microsimulation and Agent-based Model for Micro-level Population Dynamics. *Procedia Computer Science*. **80**, pp.507–517.

Ballas, D., Broomhead, T. and Jones, P.M. 2018. Spatial Microsimulation and Agent-Based Modelling *In*: H. Briassoulis, D. Kavroudakis and N. Soulakellis, eds. *The Practice of Spatial Analysis* [Online]. Cham: Springer International Publishing, pp.69–84. [Accessed 24 March 2024]. Available from: http://link.springer.com/10.1007/978-3-319-89806-3_4.

Barclays 2024. Consumer card spending grew just 3.1 per cent in January as Brits opted for nights in and home-cooked meals, yet confidence in personal finances reached two-year high | Barclays. [Accessed 27 March 2024]. Available from: https://home.barclays/news/press-releases/2024/01/consumer-card-spending-grew-just-3-1-per-cent-in-january-as-brit/.

Beaumont, J. 1991. GIS and market analysis *In*: *Geographical information systems: principles and applications*. London: Longman, pp.139–151.

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A. and Urquhart, R. 2022. Incorporating E-commerce into Retail Location Models. *Geographical Analysis*. **54**(2), pp.274–293.

Bell, D. and Mgbemena, C. 2018. Data-driven agent-based exploration of customer behavior. *SIMULATION*. **94**(3), pp.195–212.

Bell, M. 2015. Demography, time and space. *Journal of Population Research*. **32**(3/4), pp.173–186.

Benoit, D. and Clarke, G.P. 1997. Assessing GIS for retail location planning. *Journal of Retailing and Consumer Services*. **4**(4), pp.239–258.

Berry, T., Newing, A., Davies, D. and Branch, K. 2016. Using workplace population statistics to understand retail store performance. *The International Review of Retail, Distribution and Consumer Research*. **26**(4), pp.375–395.

Bevan, J. 2006. *Trolley wars: the battle of the supermarkets*. London: Profile Books.

Birkin, M. 2018. Big Data for Social Science Research: Big Data (Ubiquity symposium). *Ubiquity Symposium*. **2018**, pp.1–7.

Birkin, M. 2021. Microsimulation *In*: W. Shi, M. F. Goodchild, M. Batty, M.-P. Kwan and A. Zhang, eds. *Urban Informatics* [Online]. The Urban Book Series. Singapore: Springer, pp.845–864. [Accessed 18 December 2023]. Available from: https://doi.org/10.1007/978-981-15-8983-6_44.

Birkin, M. and Clarke, G. 2012. The enhancement of spatial microsimulation models using geodemographics. *The Annals of Regional Science*. **49**(2), pp.515–532.

Birkin, M., Clarke, G. and Clarke, M. 2010. Refining and operationalising entropy maximizing models for business applications. *Geographical Analysis*. **42**, pp.422–445.

Birkin, M., Clarke, G. and Clarke, M. 2017. *Retail Location Planning in an Era of Multi-Channel Growth*. Routledge.

Birkin, M., Clarke, G.P. and Clarke, M. 1999. GIS for business and service planning. *Geographical Information Systems*. (2), pp.709–722.

Birkin, M. and Clarke, M. 2019. Applied spatial modelling in the twenty-first century: the Wilson legacy. Looking back and looking forward. *Interdisciplinary Science Reviews*. **44**(3–4), pp.286–300.

Birkin, M. and Heppenstall, A. 2011. Extending Spatial Interaction Models with Agents for Understanding Relationships in a Dynamic Retail Market. *Urban Studies Research*. **2011**, pp.1–12.

Bland, J.M. and Altman, D.G. 1999. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*. **8**(2), pp.135–160.

Blythman, J. 2004. *Shopped: the shocking power of British supermarkets*. London: Harper Perennial.

Boero, R., Castellani, M. and Squazzoni, F. 2008. Individual behavior and macro social properties. An agent-based model. *Computational & Mathematical Organization Theory*. **14**, pp.156–174.

Bombaij, N.J.F. and Dekimpe, M.G. 2020. When do loyalty programs work? The moderating role of design, retailer-strategy, and country characteristics. *International Journal of Research in Marketing*. **37**(1), pp.175–195.

Bonabeau, E. 2002. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*. **99**(Supplement 3), pp.7280–7287.

Borgers, A. and Timmermans, H. 1986. A Model of Pedestrian Route Choice and Demand for Retail Facilities within Inner-City Shopping Areas. *Geographical Analysis*. **18**(2), pp.115–128.

Boyle, P., Bond, R., Martinez Carracedo, J., Simmons, G., Mulvenna, M. and Hollywood, L. 2022. The impact of the COVID-19 pandemic on grocery shopper behaviour: Analysis of shopper behaviour change using store transaction data. *Journal of Consumer Behaviour*. **21**(2), pp.259–271.

Brown, S. 1993. Retail location theory: evolution and evaluation. *International Review of Retail, Distribution and Consumer Research*. **3**(2), pp.185–229.

Bruch, E. and Atwell, J. 2015. AGENT-BASED MODELS IN EMPIRICAL SOCIAL RESEARCH. *Sociological methods & research*. **44**(2), pp.186–221.

Bulman, J., Davies, R. and Carrel, O. 2017. Living costs and food survey: Technical report for the surbey year 2015/16. [Accessed 13 September 2020]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/methodologies/livingcostsandfoodsurvey.

Burns, L., See, L., Heppenstall, A. and Birkin, M. 2018. Developing an Individual-level Geodemographic Classification. *Applied Spatial Analysis and Policy*. **11**(3), pp.417–437.

Burt, S. 2000. The strategic role of retail brands in British grocery retailing. *European Journal of Marketing*. **34**(8), pp.875–890.

Burt, S. and Sparks, L. 1994. Structural change in grocery retailing in Great Britain: a discount reorientation? *The International Review of Retail, Distribution and Consumer Research*. **4**(2), pp.195–217.

Burt, S. and Sparks, L. 2001. The Implications of Wal-Mart's Takeover of ASDA. *Environment and Planning A*. **33**, pp.1463–1487.

Burt, S. and Sparks, L. 1995. Understanding the arrival of limited line discount stores in Britain. *European Management Journal*. **13**(1), pp.110–119.

Burt, S., Sparks, L. and Teller, C. 2010. Retailing in the United Kingdom - a Synopsis. *European Retail Research*. **24**(1), pp.173–194.

Burt, S.L. and Sparks, L. 2003. Power and Competition in the UK Retail Grocery Market. *British Journal of Management*. **14**(3), pp.237–254.

Byrom, J., Hernández, T., Bennison, D. and Hooper, P. 2001. Exploring the geographical dimension in loyalty card data. *Marketing Intelligence & Planning*. **19**(3), pp.162–170.

Campaign Magazine 1999. AGENDA: Sainsbury's Central stores promises a radical format for the busy city-centre and lunchtime shopper. [Accessed 27 July 2023]. Available from: https://www.campaignlive.co.uk/article/agenda-sainsburys-central-push-sainsburys-new-central-stores-promises-radical-format-busy-city-centre-lunchtime-shopper-simply-trailing-behind-its/59268?utm_source=website&utm_medium=social.

Castle, C.J.E. and Crooks, A.T. 2006. Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations. *UCL Working Papers Series*. (110).

Chen, L. 2012. Agent-based modeling in urban and architectural research: A brief literature review. *Frontiers of Architectural Research*. **1**(2), pp.166–177.

Chen, Q. and Crooks, A.T. 2021. Delineating a '15-minute city': an agent-based modeling approach to estimate the size of local communities *In*: *Proceedings of the 4th ACM SIGSPATIAL International Workshop on GeoSpatial Simulation* [Online]. Beijing China: ACM, pp.29–37. [Accessed 3 February 2023]. Available from: https://dl.acm.org/doi/10.1145/3486184.3491080.

Cheshire, P.C., Hilber, C.A.L., Montebruno, P. and Sanchis-Guarner, R. 2022. (In)convenient stores? What do policies pushing stores to town centres actually do?

Ciari, F., Löchl, M. and Axhausen, K.W. 2008. Location decisions of retailers: an agent-based ap- proach. , p.34.

Clark, S., Hood, N. and Birkin, M. 2021. A hedonic model of the association between grocery brand provision and residential rental prices in England. *International Journal of Housing Markets and Analysis*. **14**(4), pp.680–700.

Clark, S., Hood, N. and Birkin, M. 2022. Identifying the effect of retail brands on private residential rental prices in Great Britain. *Journal of Housing and the Built Environment*. **37**(3), pp.1489–1509.

Clarke, G. 1998. Changing methods of location planning for retail companies. *GeoJournal*. **45**, pp.289–298.

Clarke, G., Thompson, C. and Birkin, M. 2015. The emerging geography of e-commerce in British retailing. *Regional Studies, Regional Science*. **2**(1), pp.371–391.

Clarke, G.P. and Hayes, S. 2013. GIS and Retail Location Models *In*: *Geomarketing* [Online]. John Wiley & Sons, Ltd, pp.165–186. [Accessed 29 July 2023]. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118614020.ch7.

Clarke, I. 2000. Retail power, competition and local consumer choice in the UK grocery sector. *European Journal of Marketing*. **34**(8), pp.975–1002.

Clarke, M. and Birkin, M. 2018. Spatial Interaction Models: from Numerical Experiments to Commercial Applications. *Applied Spatial Analysis and Policy*. **11**(4), pp.713–729.

Clarkson, R.M., Clarke-Hill, C.M. and Robinson, T. 1996. UK supermarket location assessment. *International Journal of Retail & Distribution Management*. **24**(6), p.22.

Consumer Intelligence 2022. How is the cost-of-living crisis impacting shopping and switching? [Accessed 27 July 2023]. Available from: https://www.consumerintelligence.com/articles/how-is-the-cost-of-living-crisis-impacting-shopping-and-switching.

conveniencestore.co.uk 2007. Sainsbury's gives c-stores a rethink. *Convenience Store*. [Online]. [Accessed 27 July 2023]. Available from: https://www.conveniencestore.co.uk/news/sainsburys-gives-c-stores-a-rethink/191094.article.

Cooley, P. and Solano, E. 2023. Agent-Based Model (ABM) Validation Considerations.

Covéa Insurance 2014. Covéa Insurance expands Halifax offices to accommodate surge in recruitment | Covéa Insurance. [Accessed 25 July 2023]. Available from: https://www.coveainsurance.co.uk/news-hub/press-releases/covea-insurance-expands-halifax-offices-to-accommodate-surge-in-recruitment/.

Crols, T. and Malleson, N. 2019. Quantifying the ambient population using hourly population footfall data and an agent-based model of daily mobility. *GeoInformatica*. **23**(2), pp.201–220.

Crooks, A.T. and Heppenstall, A.J. 2012. Introduction to agent-based modelling *In*: *Agent-based models of geographical systems*. Springer, pp.85–105.

Cuthbertson, R., Reynolds, J., Kar, M., Blake, N. and Hastings, T. 2012. *UK Retail: Leading Globally, Serving Locally* [Online]. Lonson: British Retail Consortium. [Accessed 14 April 2023]. Available from: http://www.brc.org.uk/downloads/UK_Retail_Leading_Globally_Serving_Locally.pdf.

D. Polding, R. and Eizaguirre Dieguez, M. 2021. An Investigation into the Effectiveness of Big Data in Organizations, the Use of Customer Data, and the Ethical Implications of the Data Economy *In*: *2021 International Symposium on Electrical, Electronics and Information Engineering* [Online]. ISEEIE 2021. New York, NY, USA: Association for Computing Machinery, pp.599–607. [Accessed 25 March 2024]. Available from: https://doi.org/10.1145/3459104.3459201.

Daisy, N.S., Millward, H. and Liu, L. 2018. Trip chaining and tour mode choice of non-workers grouped by daily activity patterns. *Journal of Transport Geography*. **69**, pp.150–162.

Davies and Rogers 1984. *Store Location and Assessment Research* [Online] 1st ed. Chichester: Wiley. [Accessed 29 July 2023]. Available from: https://www.wiley.com/en-us/Store+Location+and+Assessment+Research-p-9780471903819.

De Kervenoael, R., Hallsworth, A. and Clarke, I. 2006. Macro-level change and micro level effects: A twenty-year perspective on changing grocery shopping behaviour in Britain. *Journal of Retailing and Consumer Services*. **13**(6), pp.381–392.

deanclough.com 2023. Dean Clough, Halifax: home to the arts, leisure & business. *Dean Clough*. [Online]. [Accessed 25 July 2023]. Available from: https://www.deanclough.com/about-us/.

DeAngelis, D.L. and Diaz, S.G. 2019. Decision-Making in Agent-Based Modeling: A Current Review and Future Prospectus. *Frontiers in Ecology and Evolution*. **6**.

DeAngelis, D.L. and Grimm, V. 2014. Individual-based models in ecology after four decades. *F1000Prime Reports*. **6**, p.39.

DoE 1996. Town Centres and Retail Development. Revised Planning Policy Guidance Note 6.

Dolega, L. and Lord, A. 2020. Exploring the geography of retail success and decline: A case study of the Liverpool City Region. *Cities*. **96**, p.102456.

East, R. 2022. Online Grocery Sales after the Pandemic. *International Journal of Market Research*. **64**(1), pp.13–18.

Edmonds, B., Le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root, H. and Squazzoni, F. 2019. Different Modelling Purposes. *Journal of Artificial Societies and Social Simulation*. **22**(3), p.6.

Edmonds, B. and Moss, S. 2005. From KISS to KIDS – An ‘Anti-simplistic’ Modelling Approach *In*: P. Davidsson, B. Logan and K. Takadama, eds. *Multi-Agent and Multi-Agent-Based Simulation*. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, pp.130–144.

Elms, J., Canning, C., de Kervenoael, R., Whysall, P. and Hallsworth, A. 2010. 30 years of retail change: where (and how) do you shop? J. Fernie, ed. *International Journal of Retail & Distribution Management*. **38**(11/12), pp.817–827.

England, J. 2000. *Retail Impact Assessment: A Guide to Best Practice*. London: Routledge.

Foot, D. 1981. *Operational Urban Models: An Introduction*. London: Routledge.

Free, C. 2008. Walking the talk? Supply chain accounting and trust among UK supermarkets and suppliers. *Accounting, Organizations and Society*. **33**, pp.629–662.

Gale, C.G., Singleton, A.D., Bates, A.G. and Longley, P.A. 2016. Creating the 2011 area classification for output areas. *Journal of Spatial Information Science*. (12), pp.1–27.

Garcia-Magarino, I., Gray, G., Lacuesta, R. and Lloret, J. 2018. Survivability Strategies for Emerging Wireless Networks With Data Mining Techniques: a Case Study With NetLogo and RapidMiner. *IEEE Access*. **6**, pp.27958–27970.

Gauri, D.K., Sudhir, K. and Talukdar, D. 2008. The Temporal and Spatial Dimensions of Price Search: Insights from Matching Household Survey and Purchase Data. *Journal of Marketing Research*. **45**(2), pp.226–240.

Gautheir, T.D. 2001. Detecting Trends Using Spearman's Rank Correlation Coefficient. *Environmental Forensics*. **2**(4), pp.359–362.

geolytix.com 2020. Geolytix UK : Retail Place Boundaries. *Geolytix*. [Online]. [Accessed 25 March 2024]. Available from: https://geolytix.com/blog/uk-retail-places/.

Getis, A. and Getis, J. 1966. Christaller's Central Place Theory. *Journal of Geography*. **65**(5), pp.220–226.

Gomez, J. 2019. Consumer Behavior Using an Agent Based Model: Analyzing 4 Theories.

Gonçalves, M. and Nott, G. 2022. How supermarket loyalty schemes are aiming to trump inflation. *The Grocer*. [Online]. [Accessed 27 July 2023]. Available from: https://www.thegrocer.co.uk/marketing/how-supermarket-loyalty-schemes-are-aiming-to-trump-inflation/664162.article.

gov.uk 2022. Trading hours for retailers: the law. *GOV.UK*. [Online]. [Accessed 24 October 2022]. Available from: https://www.gov.uk/trading-hours-for-retailers-the-law.

Grimm, V., Berger, U., DeAngelis, D.L., Polhill, J.G., Giske, J. and Railsback, S.F. 2010. The ODD protocol: A review and first update. *Ecological Modelling*. **221**(23), pp.2760–2768.

Grimm, V. and Railsback, S.F. 2012. Designing, Formulating, and Communicating Agent-Based Models *In*: A. J. Heppenstall, A. T. Crooks, L. M. See and M. Batty, eds. *Agent-Based Models of Geographical Systems* [Online]. Dordrecht: Springer Netherlands, pp.361–377. [Accessed 16 May 2023]. Available from: https://link.springer.com/10.1007/978-90-481-8927-4_17.

Guo, L. and Wang, Z. 2019. Ratio Analysis of J Sainsbury plc Financial Performance between 2015 and 2018 in Comparison with Tesco and Morrisons. *American Journal of Industrial and Business Management*. **9**(2), pp.325–341.

Guy, C. 2007. Planning for Retail Development. *Journal of Retail & Leisure Property*. **6**(3), pp.269–270.

Guy, C.M. 1997. Fixed Assets or Sunk Costs? An Examination of Retailers' Land and Property Investment in the United Kingdom. *Environment and Planning A: Economy and Space*. **29**(8), pp.1449–1464.

Guy, C.M. 1980. *Retail location and retail planning in Britain*. Farnborough: Gower.

Hägerstrand, T. 1970. What about people in regional science? *Papers in Regional Science*. **24**(1), pp.7–24.

Hamad, H. and Schmitz, M. 2019. Journal of Marketing and Consumer Research www.iiste.org ISSN.

Hammond, R.A. 2015. *Considerations and Best Practices in Agent-Based Modeling to Inform Policy* [Online]. National Academies Press (US). [Accessed 14 April 2020]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK305917/.

Harland, K. and Birkin, M. 2015. Microsimulation Modelling for Social Scientists *In*:, pp.78–96.

Harris, L.C. and Ogbonna, E. 2001. Competitive advantage in the UK food retailing sector: past, present and future. *Journal of Retailing and Consumer Services*. **8**(3), pp.157–173.

Harris, P., Robinson, H., Riley, F.D. and Hand, C. 2017. Consumers' Multi-Channel Shopping Experiences in the UK Grocery Sector: Purchase Behaviour, Motivations and Perceptions: An Extended Abstract *In*: P. Rossi, ed. *Marketing at the Confluence between Entertainment and Analytics*. Developments in Marketing Science: Proceedings of the Academy of Marketing Science. Cham: Springer International Publishing, pp.103–107.

Hasan, S., Schneider, C.M., Ukkusuri, S.V. and González, M.C. 2013. Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics*. **151**(1–2), pp.304–318.

Hassan, A. and Parvez, M. 2013. A Comparative Case Study Investigating the Adoption of Customer Relationship Management (CRM) The Case of Tesco and Sainsbury's. *International Journal of Managing Value and Supply Chains*. **4**(1), pp.1–10.

Heppenstall, A., Crooks, A., Malleson, N., Manley, E., Ge, J. and Batty, M. 2021. Future Developments in Geographical Agent-Based Models: Challenges and Opportunities. *Geographical Analysis*. **53**(1), pp.76–91.

Heppenstall, A., Evans, A. and Birkin, M. 2006. Using hybrid agent-based systems to model spatially-influenced retail markets. *Journal of Artificial Societies and Social Simulation*. **9**(3).

Heppenstall, A., Malleson, N. and Crooks, A. 2016. "Space, the Final Frontier": How Good are Agent-Based Models at Simulating Individuals and Space in Cities? *Systems*. **4**(1), p.9.

Heppenstall, A.J., Harland, K., Ross, A.N. and Olner, D. 2013. Simulating Spatial Dynamics and Processes in a Retail Gasoline Market: An Agent-Based Modeling Approach. *Transactions in GIS*. **17**(5), pp.661–682.

Hernández, T. and Bennison, D. 2000. The art and science of retail location decisions. *International Journal of Retail & Distribution Management*. **28**(8), pp.357–367.

Hernandez, T., Bennison, D. and Cornelius, S. 1998. The organisational context of retail locational planning. , p.10.

Hildebrand, A. 2012. Analysis: Retail CCTV - from surveillance to shopper analysis. *Retail Week*. [Online]. [Accessed 30 July 2023]. Available from: https://www.retail-week.com/tech/analysis-retail-cctv-from-surveillance-to-shopper-analysis/5044197.article.

Hincks, S., Kingston, R., Webb, B. and Wong, C. 2018. A new geodemographic classification of commuting flows for England and Wales. *International Journal of Geographical Information Science*. **32**(4), pp.663–684.

Hood, N., Clarke, G. and Clarke, M. 2016. Segmenting the growing UK convenience store market for retail location planning. *The International Review of Retail, Distribution and Consumer Research*. **26**(2), pp.113–136.

Hood, N., Clarke, G., Newing, A. and Rains, T. 2021. Applications of store loyalty card big data in the location planning process. *Big Data Applications in Geography and Planning*., pp.107–120.

Hood, N., Urquhart, R., Newing, A. and Heppenstall, A. 2020. Sociodemographic and spatial disaggregation of e-commerce channel use in the grocery market in Great Britain. *Journal of Retailing and Consumer Services*. **55**, p.102076.

Huff, D.L. 1963. A Probabilistic Analysis of Shopping Center Trade Areas. *Land Economics*. **39**(1), pp.81–90.

Hughes, C. 2019. *Sainsbury's Future of Food Report*. London.

Humby, C., Hunt, T. and Phillips, T. 2004. *Scoring Points: How Tesco is Winning Customer Loyalty*. Kogan Page Publishers.

Hummel, P., Braun, M. and Dabrock, P. 2021. Own Data? Ethical Reflections on Data Ownership. *Philosophy & Technology*. **34**(3), pp.545–572.

Humphery, K. 1998. *Shelf Life: Supermarkets and the Changing Cultures of Consumption*. Cambridge University Press.

IBISWorld 2023. Onlinge Grocery Retailers in the UK - Market Size 2011-2029. [Accessed 22 July 2023]. Available from: https://www.ibisworld.com/default.aspx.

IGD 2022. Growth forecast across all retail channels. [Accessed 30 March 2023]. Available from: https://www.igd.com/articles/article-viewer/t/growth-forecast-across-all-retail-channels/i/29820.

Ing, D. and Mitchell, A. 1994. Point of Sale Data in Consumer Goods Marketing: Transforming the Art of Marketing into the Science of Marketing *In*: *The Marketing Information Revolution*. Boston: Harvard Business School Press, pp.30–57.

Jackson, P., del Aguila, R.P., Clarke, I., Hallsworth, A., de Kervenoael, R. and Kirkup, M. 2006. Retail Restructuring and Consumer Choice 2. Understanding Consumer Choice at the Household Level. *Environment and Planning A: Economy and Space*. **38**(1), pp.47–67.

Jones, P. 1981. Retail Innovation and Diffusion: The Spread of Asda Stores. *Area*. **13**(3), pp.197–201.

Jordan, R., Birkin, M. and Evans, A. 2014. An agent-based model of residential mobility. *Computers, Environment and Urban Systems*. **48**, pp.49–63.

Jurgens, D., Finethy, T., McCorriston, J., Xu, Y. and Ruths, D. 2015. Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice. *Proceedings of the International AAAI Conference on Web and Social Media*. **9**(1), pp.188–197.

Kahn, B.E. and Schmittlein, D.C. 1989. Shopping trip behavior: An empirical investigation. *Marketing Letters*. **1**(1), pp.55–69.

Kansal, T., Bahuguna, S., Singh, V. and Choudhury, T. 2018. Customer Segmentation using K-means Clustering *In*: *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*., pp.135–139.

Kantar, W. 2022. Grocery Market Share - Kantar. [Accessed 30 October 2022]. Available from: https://www.kantarworldpanel.com/en/grocery-market-share/great-britain/snapshot/23.01.22/.

Kanungo, T., MounV, D.M., Netanyahu, N.S., Piatko, C., Silverman, R. and Wu, A.Y. 2000. The Analysis of a Simple k-Means Clustering Algorithm. , p.10.

Karamshuk, D., Noulas, A., Scellato, S., Nicosia, V. and Mascolo, C. 2013. Geo-Spotting: Mining Online Location-based Services for Optimal Retail Store Placement *In*: *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* [Online]., pp.793–801. [Accessed 8 July 2023]. Available from: http://arxiv.org/abs/1306.1704.

Kaziyeva, D., Stutz, P., Wallentin, G. and Loidl, M. 2023. Large-scale agent-based simulation model of pedestrian traffic flows. *Computers, Environment and Urban Systems*. **105**, p.102021.

Kerr, C.C., Stuart, R.M., Mistry, D., Abeysuriya, R.G., Rosenfeld, K., Hart, G.R., Núñez, R.C., Cohen, J.A., Selvaraj, P., Hagedorn, B., George, L., Jastrzębski, M., Izzo, A.S., Fowler, G., Palmer, A., Delport, D., Scott, N., Kelly, S.L., Bennette, C.S., Wagner, B.G., Chang, S.T., Oron, A.P., Wenger, E.A., Panovska-Griffiths, J., Famulare, M. and Klein, D.J. 2021. Covasim: An agent-based model of COVID-19 dynamics and interventions. *PLOS Computational Biology*. **17**(7), p.e1009149.

Khawaldah, H., Birkin, M. and Clarke, G. 2012. A Review of Two Alternative Retail Impact Assessment Techniques: The Case of Silverburn in Scotland. *The Town Planning Review*. **83**, pp.233–260.

Kirby-Hawkins, E. 2016. DESIGNING A LOCATION MODEL FOR FACE TO FACE AND ON-LINE RETAILING FOR THE UK GROCERY MARKET. , p.246.

Kirby-Hawkins, E., Birkin, M. and Clarke, G. 2018. An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Environment and Planning B: Urban Analytics and City Science*. **46**(6), pp.1148–1164.

Klein, D., Filipiak, K., Calinski, T. and Międzynarodowe Centrum Matematyczne im. Stefana Banacha (eds.). 2018. *International Conference on Trends and Perspectives in Linear Statistical Inference with celebration of Tadeusz Caliński's 90th birthday: LINSTAT 2018, Będlewo, Poland: book of abtracts, August 20-24, 2018, Będlewo, Poland*. Poznań: Bogucki Wydawnictwo Naukowe.

Knox, S.D. and Denison, T.J. 2000. Store loyalty: its impact on retail revenue. An empirical study of purchasing behaviour in the UK. *Journal of Retailing and Consumer Services*. **7**(1), pp.33–45.

Kotsiantis, S.B. 2013. Decision trees: a recent overview. *Artificial Intelligence Review*. **39**(4), pp.261–283.

Krumme, C., Llorente, A., Cebrian, M., Pentland, A. ('Sandy') and Moro, E. 2013. The predictability of consumer visitation patterns. *Scientific Reports*. **3**(1), p.1645.

Langston, P., Clarke, G.P. and Clarke, D.B. 1997. Retail saturation, retail location, and retail competition: an analysis of British grocery retailing. *Environment and Planning A*. **29**, pp.77–104.

Langston, P., Clarke, G.P. and Clarke, D.B. 1998. Retail Saturation: The Debate in the Mid-1990s. *Environment and Planning A: Economy and Space*. **30**(1), pp.49–66.

Lansley, G. and Cheshire, J. 2018. Challenges to representing the population from new forms of consumer data. *Geography Compass*. **12**(7), p.e12374.

Lansley, G. and Longley, P. 2017. Opprtunities for retail data and their geographic integration in social science *In*: *Big Data for Regional Science*. Routledge.

Lauer, J. 2020. Plastic surveillance: Payment cards and the history of transactional data, 1888 to present. *Big Data & Society*. **7**(1), p.2053951720907632.

Lee, J.-G. and Kang, M. 2015. Geospatial Big Data: Challenges and Opportunities. *Big Data Research*. **2**(2), pp.74–81.

Lee, J.-S., Filatova, T., Ligmann-Zielinska, A., Hassani-Mahmooei, B., Stonedahl, F., Lorscheid, I., Voinov, A., Polhill, J.G., Sun, Z. and Parker, D.C. 2015. The Complexities of Agent-Based Modeling Output Analysis. *Journal of Artificial Societies and Social Simulation*. **18**(4), p.4.

Leyland, A. 2022. Why is it grocers 'match' their prices to Aldi as opposed to Lidl? *The Grocer*. [Online]. [Accessed 27 July 2023]. Available from: https://www.thegrocer.co.uk/leader/why-is-it-grocers-match-their-prices-to-aldi-as-opposed-to-lidl/668597.article.

LIDA 2023. Sainsbury's: Property Analytics. *Leeds Institute for Data Analytics*. [Online]. [Accessed 31 January 2023]. Available from: https://lida.leeds.ac.uk/partnerships/sainsburys-property-analytics/.

Lipowski, A. and Lipowska, D. 2012. Roulette-wheel selection via stochastic acceptance. *Physica A: Statistical Mechanics and its Applications*. **391**(6), pp.2193–2196.

Lloyd, A. and Cheshire, J. 2019. Detecting Address Uncertainty in Loyalty Card Data. *Applied Spatial Analysis and Policy*. **12**(2), pp.445–465.

Lloyd, S. 1982. Least squares quantization in PCM. *IEEE Transactions on Information Theory*. **28**(2), pp.129–137.

Lorscheid, I., Berger, U., Grimm, V. and Meyer, M. 2019. From cases to general principles: A call for theory development through agent-based modeling. *Ecological Modelling*. **393**, pp.153–156.

Lovelace, R., Birkin, M., Cross, P. and Clarke, M. 2016. From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows. *Geographical Analysis*. **48**(1), pp.59–81.

Lovelace, R. and Dumont, M. 2016. *Spatial Microsimulation with R*. New York: Chapman and Hall/CRC.

Ma, X., Liu, C., Wen, H., Wang, Y. and Wu, Y.-J. 2017. Understanding commuting patterns using transit smart card data. *Journal of Transport Geography*. **58**, pp.135–145.

Macqueen, J. 1967. Classification and analysis of multivariate observations. *5th Berkeley Symp. Math. Statist. Probability*., pp.287–297.

Malleson, N., Birkin, M., Birks, D., Ge, J., Heppenstall, A., Manley, E., McCulloch, J. and Ternes, P. 2022. Agent-Based Modelling for Urban Analytics: State of the Art and Challenges. *AI Communications*. **35**(4), pp.393–406.

Martin, D., Cockings, S. and Leung, S. 2015. Developing a Flexible Framework for Spatiotemporal Population Modeling. *Annals of the Association of American Geographers*. **105**(4), pp.754–772.

Meister, A., Winkler, C., Schmid, B. and Axhausen, K. 2023. In-store or online grocery shopping before and during the COVID-19 pandemic. *Travel Behaviour and Society*. **30**, pp.291–301.

Miller, H.J. 2008. Time Geography *In*: S. Shekhar and H. Xiong, eds. *Encyclopedia of GIS* [Online]. Boston, MA: Springer US, pp.1151–1156. [Accessed 25 July 2023]. Available from: https://doi.org/10.1007/978-0-387-35973-1_1383.

Mintel 2019. Brits spent £12.3 billion on online groceries in 2018. *Mintel*. [Online]. [Accessed 5 April 2022]. Available from: https://www.mintel.com/press-centre/retail-press-centre/brits-spent-12-3-billion-on-online-groceries-in-2018.

Mitchell, M. 1998. Common Selection Methods *In*: *An introduction to genetic algorithms* [Online]. MIT Press. [Accessed 27 July 2023]. Available from: https://dl.acm.org/doi/10.1145/2907674.2907678.

Morphet, C.S. 1991. Applying multiple regression analysis to the forecasting of grocery store sales: an application and critical appraisal. *The International Review of Retail, Distribution and Consumer Research*. **1**(3), pp.329–351.

Nagpal, A. and Gabrani, G. 2019. Python for Data Analytics, Scientific and Technical Applications *In*: *2019 Amity International Conference on Artificial Intelligence (AICAI)*., pp.140–145.

Nakaya, T., Fotheringham, A.S., Hanaoka, K., Clarke, G., Ballas, D. and Yano, K. 2007. Combining microsimulation and spatial interaction models for retail location analysis. *Journal of Geographical Systems*. **9**(4), pp.345–369.

nectar360.co.uk 2023. Nectar Loyalty Programme. *Nectar360*. [Online]. [Accessed 29 June 2023]. Available from: https://www.nectar360.co.uk/what-we-do/nectar-loyalty-programme/.

Newing, A., Clarke, G. and Clarke, M. 2018. Applied spatial modelling for retail planning in tourist resorts. *International Journal of Retail & Distribution Management*. **46**(11/12), pp.1117–1132.

Newing, A., Clarke, G. and Clarke, M. 2014. Exploring small area demand for grocery retailers in tourist areas. *Tourism Economics*. **20**(2), pp.407–427.

Newing, A., Clarke, G. and Clarke, M. 2013. Identifying seasonal variations in store-level visitor grocery demand. *International Journal of Retail & Distribution Management*. **41**(6), pp.477–492.

Newing, A., Clarke, G.P. and Clarke, M. 2015. Developing and Applying a Disaggregated Retail Location Model with Extended Retail Demand Estimations. *Geographical Analysis*. **47**(3), pp.219–239.

Newing, A., Hood, N. and Sterland, I. 2020. Planning Support Science (PSS) for Retail Location Planning *In*: *Handbook on Planning Support Science*. In Handbook of Planning Support Science: Edward Elgar Publishing.

Newing, A., Hood, N., Videira, F. and Lewis, J. 2022. 'Sorry we do not deliver to your area': geographical inequalities in online groceries provision. *The International Review of Retail, Distribution and Consumer Research*. **32**(1), pp.80–99.

Newing, A.D. 2013. Incorporating seasonal visitor demand in retail location modelling. , p.316.

Nijenhuis, B., Doumen, S.C., Hönen, J. and Hoogsteen, G. 2022. Using mobility data and agent-based models to generate future E-mobility charging demand patterns *In*: *CIRED Porto Workshop 2022: E-mobility and power distribution systems* [Online]. Hybrid Conference, Porto, Portugal: Institution of Engineering and Technology, pp.214–218. [Accessed 25 March 2024]. Available from: https://digital-library.theiet.org/content/conferences/10.1049/icp.2022.0697.

North, M.J., Macal, C.M., Aubin, J.S., Thimmapuram, P., Bragen, M., Hahn, J., Karr, J., Brigham, N., Lacy, M.E. and Hampton, D. 2010. Multiscale agent-based consumer market modeling. *Complexity*. **15**(5), pp.37–47.

ODPM 2005. Planning Policy Statement 6: Planning for Town Centres. Available from: http://www.communities.gov.uk/documents/planningandbuilding/.

Office for National Statistics 2012. 2011 Census - Office for National Statistics. [Accessed 11 September 2020]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforsmallareasinenglandandwales/2012-11-23.

Office for National Statistics 2014. 2011 Census: Workplace Population. [Accessed 28 June 2023]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/articles/workplacepopulationanalysis/2014-05-23#data.

Office for National Statistics 2018. 2011 Residential-based area classifications, Pen portraits and radial plots. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications/penportraitsandradialplots.

Office for National Statistics 2020a. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland - Office for National Statistics. [Accessed 11 September 2020]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.

Office for National Statistics 2023. How life has changed in Bradford: Census 2021. [Accessed 22 July 2023]. Available from: https://www.ons.gov.uk/visualisations/censusareachanges/E08000032/.

Office for National Statistics 2020b. Living Costs and Food Survey, 2017-2018. [Accessed 8 November 2022]. Available from: https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=8459#!#3.

Office for National Statistics 2011. Location of usual residence and place of work (WF01BEW). [Accessed 10 July 2023]. Available from: https://www.nomisweb.co.uk/census/2011/wf01bew.

Office for National Statistics 2017. Output Area Classification, 2011. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclas sifications/datasets.

Office for National Statistics 2015. Pen Portraits for the 2011 Area Classification for Output Areas.

Office for National Statistics 2016. Rural/urban classification 2011. [Accessed 25 June 2023]. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification.

Office for National Statistics 2021. West Yorkshire facts and figures - ONS. Available from: https://www.ons.gov.uk/visualisations/areas/E11000006/.

Office of Rail and Road 2018. Estimates of Station Usage 2018-19.

Owen, G. 2003. Corporate strategy in UK food retailing 1980–2002 *In*: London: London School of Economics. [Accessed 27 July 2023]. Available from: https://web.archive.org/web/20040721174741/http://cep.lse.ac.uk/seminarpapers/24-05-04%20-%20Background%20paper%20by%20Geoffrey%20Owen.pdf.

Pacione, M. 1979. The in-town hypermarket: An innovation in the geography of retailing. *Regional Studies*. **13**(1), pp.15–24.

Pantano, E., Pizzi, G., Scarpi, D. and Dennis, C. 2020. Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research*. **116**, pp.209–213.

Parrott, L. 2008. Adaptive Agents *In*: S. E. Jørgensen and B. D. Fath, eds. *Encyclopedia of Ecology* [Online]. Oxford: Academic Press, pp.47–51. [Accessed 24 March 2024]. Available from: https://www.sciencedirect.com/science/article/pii/B9780080454054001415.

van der Ploeg, T., Austin, P.C. and Steyerberg, E.W. 2014. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology*. **14**(1), p.137.

Poole, R., Clarke, G.P. and Clarke, D.B. 2002a. Grocery Retailers and Regional Monopolies. *Regional Studies*. **36**(6), pp.643–659.

Poole, R., Clarke, G.P. and Clarke, D.B. 2002b. Growth, Concentration and Regulation in European Food Retailing. *European Urban and Regional Studies*. **9**(2), pp.167–186.

Popkowski Leszczyc, P.T.L., Sinha, A. and Sahgal, A. 2004. The effect of multi-purpose shopping on pricing and location strategy for grocery stores. *Journal of Retailing*. **80**(2), pp.85–99.

Primerano, F., Taylor, M.A.P., Pitaksringkarn, L. and Tisato, P. 2008. Defining and understanding trip chaining behaviour. *Transportation*. **35**(1), pp.55–72.

Puga-Gonzalez, I., Bacon, R., Voas, D., LeRon Shults, F., Hodulik, G. and Wildman, W. 2022. Adapting Cohort-Component Methods to a Microsimulation: A case study. *Social Science Computer Review*. **Review 40**(4), pp.1054–1068.

Pullum, L. and Cui, X. 2012. Techniques and Issues in Agent-Based Modeling Validation.

Putler, D.S. and Krider, R.E. 2012. *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R* [Online] 0 ed. Chapman and Hall/CRC. [Accessed 12 May 2022]. Available from: https://www.taylorfrancis.com/books/9781466503984.

Qi, J., Yu, Y., Wang, L. and Liu, J. 2016. K-Means: An Effective and Efficient K-Means Clustering Algorithm *In*: *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)* [Online]. Atlanta, GA, USA: IEEE, pp.242–249. [Accessed 24 October 2022]. Available from: https://ieeexplore.ieee.org/document/7723700/.

Quach, S., Thaichon, P., Martin, K.D., Weaven, S. and Palmatier, R.W. 2022. Digital technologies: tensions in privacy and data. *Journal of the Academy of Marketing Science*. **50**(6), pp.1299–1323.

Rains, T. 2019. *Big Consumer Data: Understanding shopping patterns with loyalty card data*. UCL.

Rains, T. and Longley, P. 2021. The provenance of loyalty card data for urban and retail analytics. *Journal of Retailing and Consumer Services*. **63**, p.102650.

Reinsel, D., Gantz, J. and Rydning, J. 2018. The Digitization of the World from Edge to Core.

Retail Connections 2023. Cost-of-living prompts discounter switching: 56% of UK shoppers have switched entire weekly shop to Aldi or Lidl. *Retail Connections*. [Online]. [Accessed 27 July 2023]. Available from: https://www.retailconnections.co.uk/articles/cost-of-living-prompts-discounter-switching-56-of-uk-shoppers-have-switched-entire-weekly-shop-to-aldi-or-lidl/.

RetailWeek 2005. Sainsbury's Legacy supermarket build on c-store arm. *Retail Week*. [Online]. [Accessed 27 July 2023]. Available from: https://www.retail-week.com/sainsburys-buys-shaw/41748.article.

Reynolds, J. and Wood, S. 2010. Location decision making in retail firms: evolution and challenge. *International Journal of Retail & Distribution Management*. **38**(11/12), pp.828–845.

Richiardi, M. 2014. The Missing Link: AB Models and Dynamic Microsimulation *In*: S. Leitner and F. Wall, eds. *Artificial Economics and Self Organization* [Online]. Lecture Notes in Economics and Mathematical Systems. Cham: Springer International Publishing, pp.3–15. [Accessed 26 January 2023]. Available from: http://link.springer.com/10.1007/978-3-319-00912-4_1.

Rigby, C. 2022. Are shoppers moving away from quick commerce apps as they put cost before convenience? RESEARCH. *Internet Retailing*. [Online]. [Accessed 22 July 2023]. Available from: https://internetretailing.net/are-shoppers-moving-away-from-quick-commerce-apps-as-they-put-cost-before-convenience-research/.

Rogers, D.S. 2016. Location planning research in a new location landscape.

Rosés, R., Kadar, C. and Malleson, N. 2021. A data-driven agent-based simulation to predict crime patterns in an urban environment. *Computers, Environment and Urban Systems*. **89**, p.101660.

Rowe, F., Lovelace, R. and Dennett, A. 2022. *Spatial Interaction Modelling: A Manifesto* [Online]. Open Science Framework. [Accessed 7 June 2023]. Available from: https://osf.io/xcdms.

Roy, J.R. and Thill, J.-C. 2004. Spatial interaction modelling. *Papers in Regional Science*. **83**(1), pp.339–361.

Sainsbury's Archive 2019. Sainsbury's Archive Timeline. [Accessed 27 July 2023]. Available from: https://sainsburyarchive.org.uk/timeline/over-150.

Sainsbury's.co.uk 2023a. Grocery delivery times & charges. Available from: https://help.sainsburys.co.uk/help/delivery-collection/grocery-delivery-times.

Sainsbury's.co.uk 2023b. Preliminary Results for the 52 weeks ended 4 March 2023. [Accessed 30 July 2023]. Available from: https://www.about.sainsburys.co.uk/news/latest-news/2023/27-04-23-preliminary-results.

Sainsbury's.co.uk 2022. Ranges | Sainsbury's. [Accessed 2 November 2022]. Available from: https://www.sainsburys.co.uk/shop/gb/groceries/get-ideas/our-ranges/our-ranges.

Sainsbury's.co.uk 2023c. Sainsbury's arrives in Henfield with launch of new Neighbourhood Hub store. [Accessed 22 July 2023]. Available from: https://www.about.sainsburys.co.uk/news/latest-news/2023/31-05-2023-sainsburys-arrives-in-henfield-with-launch-of-neighbourhood-hub-store.

Sainsbury's.co.uk 2020. Sainsbury's opens the doors to new format Neighbourhood Hub stores. [Accessed 27 July 2023]. Available from: https://www.about.sainsburys.co.uk/news/latest-news/2020/12-11-20-neighbourhood-hubs.

Sajjad, M., Singh, K., Paik, E. and Ahn, C.-W. 2016. A Data-Driven Approach for Agent-Based Modeling: Simulating the Dynamics of Family Formation. *Journal of Artificial Societies and Social Simulation*. **19**.

Santos, J.M. and Embrechts, M. 2009. On the Use of the Adjusted Rand Index as a Metric for Evaluating Supervised Classification *In*: C. Alippi, M. Polycarpou, C. Panayiotou and G. Ellinas, eds. *Artificial Neural Networks – ICANN 2009* [Online]. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.175–184. [Accessed 14 December 2021]. Available from: http://link.springer.com/10.1007/978-3-642-04277-5_18.

Schenk, T.A., Löffler, G. and Rauh, J. 2007. Agent-based simulation of consumer behavior in grocery shopping on a regional level. *Journal of Business Research*. **60**(8), pp.894–903.

Scott, S. and Koehler, M. 2011. *A Field Guide to NetLogo* Version 1. George Mason University, Department of Computational Social Science.

Seaborn, C., Attanucci, J. and Wilson, N.H.M. 2009. Analyzing Multimodal Public Transport Journeys in London with Smart Card Fare Payment Data. *Transportation Research Record*. **2121**(1), pp.55–62.

Seth, A., Pringle-Pattison, A.S. and Randall, G. 1999. *The Grocers: The Rise and Rise of the Supermarket Chains*. Kogan Page Publishers.

Shukla, P., Banerjee, M. and Adidam, P.T. 2013. The moderating influence of socio-demographic factors on the relationship between consumer psychographics and the attitude towards private label brands. *Journal of Consumer Behaviour*. **12**(6), pp.423–435.

Simkin, L.P. 1990. EVALUATING A STORE LOCATION. *International Journal of Retail & Distribution Management*. **18**(4).

Sleeman, J. 2022. Impact of the pandemic on the demand for logistics and distribution capacity. [Accessed 17 April 2023]. Available from: https://thesla.org/2022/01/logistics-and-the-future-of-retail-real-estate-10th-feb-4pm/.

Smith, E.R. and Conrey, F.R. 2007. Agent-Based Modeling: A New Approach for Theory Building in Social Psychology. *Personality and Social Psychology Review*. **11**(1), pp.87–104.

Sparks, L. 1990. Spatial-Structural Relationships in Retail Corporate Growth: A Case-Study of Kwik Save Group P.L.C. *The Service Industries Journal*. **10**(1), pp.25–84.

spring-board.info 2021. About | Springboard Research & Retail Intelligence. Available from: https://www.spring-board.info/aboutuk/.

Statistica.com 2023. Tesco Clubcard members 2023. *Statista*. [Online]. [Accessed 30 July 2023]. Available from: https://www.statista.com/statistics/1400244/tesco-clubcard-data/.

Stratton, M. (ed.). 2000. *Industrial Buildings: Conservation and Regeneration*. London: Taylor & Francis.

Sturley, C., Newing, A. and Heppenstall, A. 2018. Evaluating the potential of agent-based modelling to capture consumer grocery retail store choice behaviours. *The International Review of Retail, Distribution and Consumer Research*. **28**(1), pp.27–46.

Syakur, M.A., Khotimah, B.K., Rochman, E.M.S. and Satoto, B.D. 2018. Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*. **336**, p.012017.

Thill, J.-C. and Thomas, I. 1987. Toward Conceptualizing Trip-Chaining Behavior: A Review. *Geographical Analysis*. **19**(1), pp.1–17.

Thompson, C., Clarke, G., Clarke, M. and Stillwell, J. 2012. Modelling the future opportunities for deep discount food retailing in the UK. *The International Review of Retail, Distribution and Consumer Research*. **22**(2), pp.143–170.

Timmermans, H.J.P. 2001. Spatial Choice Models *In*: N. J. Smelser, J. Wright and P. B. Baltes, eds. *International Encyclopedia of the Social & Behavioral Sciences*. Oxford: Elsevier, pp.14768–14771.

Timmins, K.A. and Edwards, K.L. 2016. Validation of spatial microsimulation models: a proposal to adopt the Bland-Altman method. *International Journal of Microsimulation*. **9**(2), pp.106–122.

Tisue, S. and Wilensky, U. 2004. NetLogo: A Simple Environment for Modeling Complexity.

Tobler, W.R. 1987. Experiments In Migration Mapping By Computer. *The American Cartographer*. **14**(2), pp.155–163.

Torrens, P.M. 2023. Agent models of customer journeys on retail high streets. *Journal of Economic Interaction and Coordination*. **18**(1), pp.87–128.

Torrens, P.M. 2012. Moving Agent Pedestrians Through Space and Time. *Annals of the Association of American Geographers*. **102**(1), pp.35–66.

Trewern, J., Chenoweth, J., Christie, I., Keller, E. and Halevy, S. 2021. Are UK retailers well placed to deliver 'less and better' meat and dairy to consumers? *Sustainable Production and Consumption*. **28**, pp.154–163.

Tsou, M.-H., Zhang, H., Nara, A. and Han, S.Y. 2018. Estimating hourly population distribution change at high spatiotemporal resolution in urban areas using geo-tagged tweets, land use data, and dasymetric maps.

Urquhart, R., Newing, A., Hood, N. and Heppenstall, A. 2022. Last-Mile Capacity Constraints in Online Grocery Fulfilment in Great Britain. *Journal of Theoretical and Applied Electronic Commerce Research*. **17**(2), pp.636–651.

Vanhaverbeke, L. and Macharis, C. 2011. An agent-based model of consumer mobility in a retail environment. *Procedia - Social and Behavioral Sciences*. **20**, pp.186–196.

Waddington, T., Clarke, G., Clarke, M.C., Hood, N. and Newing, A. 2019. Accounting for Temporal Demand Variations in Retail Location Models. *Geographical Analysis*. **51**(4), pp.426–447.

Waddington, T.B.P., Clarke, G.P., Clarke, M. and Newing, A. 2018. Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand. *The International Review of Retail, Distribution and Consumer Research*. **28**(1), pp.1–26.

Walker, B. and Johnson, T. 2019. NetLogo and GIS: A Powerful Combination *In*:, pp.257–248. [Accessed 21 December 2022]. Available from: https://easychair.org/publications/paper/HTNF.

Which.co.uk 2023. Supermarket price comparison. Available from: https://www.which.co.uk/reviews/supermarkets/article/supermarket-price-comparison-aPpYp9j1MFin.

Wickramasinghe, B.N., Singh, D. and Padgham, L. 2020. Building a large synthetic population from Australian census data.

Wilkinson, P.D.J. 2023. Spatial Interaction Models in a Big Data Grocery Retailing Environment.

Willem, L., Verelst, F., Bilcke, J., Hens, N. and Beutels, P. 2017. Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015). *BMC Infectious Diseases*. **17**(1), p.612.

Wilson, A. 1970. *Entropy in Urban and Regional Modelling* [Online]. London: Pion. [Accessed 4 April 2023]. Available from: https://www.routledge.com/Entropy-in-Urban-and-Regional-Modelling-Routledge-Revivals/Wilson/p/book/9780415696319.

Wilson, A.G. 1974. *Urban and regional models in geography and planning* [Online]. London: Wiley. [Accessed 4 April 2023]. Available from: https://trid.trb.org/view/1172294.

Wood, S. and Browne, S. 2007. Convenience store location planning and forecasting – a practical research agenda. *International Journal of Retail & Distribution Management*. **35**(4), pp.233–255.

Wood, S., Lowe, M. and Wrigley, N. 2006. Life after PPG6—recent UK food retailer responses to planning regulation tightening. *The International Review of Retail, Distribution and Consumer Research*. **16**(1), pp.23–41.

Wood, S. and McCarthy, D. 2014. The UK food retail ‘race for space’ and market saturation: a contemporary review. *The International Review of Retail, Distribution and Consumer Research*. **24**(2), pp.121–144.

Wood, S. and Tasker, A. 2008. The importance of context in store forecasting: The site visit in retail location decision-making. *Journal of Targeting, Measurement and Analysis for Marketing*. **16**(2), pp.139–155.

Working, H. and Hotelling, H. 1929. Applications of the Theory of Error to the Interpretation of Trends. *Journal of the American Statistical Association*. **24**(165A), pp.73–85.

Wright, T. 2008. Geography at Sainsbury’s. *University of Leeds Alumni Newsletter*.

Wrigley 1987. The Concentration of Capital in UK Grocery Retailing. *Environment and Planning A*. **19**, pp.1283–1288.

Wrigley, N. 1994. After the store wars:: Towards a new area of competition in UK food retailing? *Journal of Retailing and Consumer Services*. **1**, pp.5–20.

Wrigley, N. 1992. Antitrust Regulation and the Restructuring of Grocery Retailing in Britain and the USA. *Environment and Planning A*. **24**, pp.727–749.

Wrigley, N. 1991. Is the ‘Golden Age’ of British Grocery Retailing at a Watershed? *Environment and Planning A: Economy and Space*. **23**(11), pp.1537–1544.

Wrigley, N. and Lowe, M. 1996. *Retailing, consumption and capital: towards the new retail geography* [Online] (N. Wrigley & M. Lowe, eds.). Longman. [Accessed 19 April 2023]. Available from: https://eprints.soton.ac.uk/184035/.

Wu, G., Heppenstall, A., Meier, P., Purshouse, R. and Lomax, N. 2022. A synthetic population dataset for estimating small area health and socio-economic outcomes in Great Britain. *Scientific Data*. **9**(1), p.19.

Yang, L. and Gilbert, N. 2008. Getting away from numbers: using qualitative observation for agent-based modeling. *Advances in Complex Systems*. **11**(02), pp.175–185.

Yang, Y., Cao, J., Liu, Xuefeng and Liu, Xiulong 2018. Wi-Count: Passing People Counting with COTS WiFi Devices *In*: *2018 27th International Conference on Computer Communication and Networks (ICCCN)*., pp.1–9.

Yeung, K.Y. and Ruzzo, W. 2001. Details of the Adjusted Rand index and Clustering algorithms Supplement to the paper ‘An empirical study on Principal Component Analysis for clustering gene expression data’ (to appear in Bioinformatics). *Science*. **17**.