Harnessing social media data to explore urban tourist patterns and the implications for retail location modelling

Zi Ye

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

School of Geography

September 2021

The candidate confirms that the work submitted is his/her own, except where work which has formed part of jointly-authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

The work in Chapter 4 of the thesis has appeared in the publication as follows:

Ye, Z., Newing, A. and Clarke, G. 2020. Understanding Chinese tourist mobility and consumption-related behaviours in London using Sina Weibo check-ins. *Environment and Planning B: Urban Analytics and City Science*., p.239980832098074.

Zi Ye was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Graham Clarke and Andy Newing supervised this study, providing conceptual direction and advice on analyses and discussion. Andy Newing made edits and comments on the manuscript throughout. Graham Clarke provided advice on manuscript revision.

The work in Chapter 5 of the thesis has appeared in the publication as follows:

Ye, Z., Clarke, G. and Newing, A. 2021. Estimating small-area demand of urban tourist for groceries: The case of Greater London. *Journal of Retailing and Consumer Services*. 58, p.102263.

Zi Ye was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Graham Clarke provided advice on the scope of the paper. Graham Clarke and Andy Newing supervised this study, providing critical feedback and advice on analyses and discussion, made edits and comments on manuscripts.

The work in Chapter 6 of the thesis has appeared in the paper as follows:

Ye, Z., Clarke, G. and Newing, A. Adding urban tourism to retail location models. Submitted.

Zi Ye was responsible for the study design, data collection and analysis, preparation of figures, and writing the manuscript. Graham Clarke and Andy Newing supervised this study, providing critical feedback and advice on analyses and discussion, and made edits and comments on manuscripts. Andy Newing provided recommendation for model calibration.

Acknowledgements

This thesis has come to be with the kind support and assistance of many people. First, I would like to express my sincere gratitude to my supervisors Professor Graham Clarke and Dr. Andy Newing for their support, guidance and encouragement throughout the whole journey. They have led me to learn and to explore not only the route of this research project, but also the path to becoming an independent researcher. The thesis would not be in this shape without Graham's expertise and insights in this topic. It is also because of Andy's considerate support and spiritual encouragement that I could resume my confidence and enthusiasm at many points of the journey. I am also grateful for the valuable comments and constructive suggestions from my external examiner Paul Longley (University College London) and internal examiner Ed Manley on my thesis.

I would also like to thank my research support group panel members Professor Lex Comber and Dr Roger Beecham and transfer examiners Professor Alison Heppenstall for their constructive advice. The members of CSAP and other staff and postgraduates in the SoG have contributed a helpful and enjoyable study environment to provide me with help at various stages of my PhD. Particular thanks are expressed to Jacqui Manton, Alice Fletcher and Alison Suckall, who kindly offered help on many occasions during my PhD.

This work has been funded by the Chinese Scholarship Council and the University of Leeds (No. 201709370048). I am grateful for the fundings and the opportunity they offered for me to undertake this PhD.

I feel blessed to have all the amazing friends I met in the past four years. Thanks are due to Yi-Min Chang Chien, León Felipe Téllez Contreras, Claudio Bravo Lechuga, Yuanxuan Yang, Lili Xiang, Antonio Sliver, Nikee Groot, Marie Arnaud, Hoayong Lee, Sungje Jeon, Han Xu, Zhenzhen Wang, Ru Xu and Muteb Alotaibi. The joyful time we spent together inside and outside the office has become a great memory. Thanks in particular to Wanyun Ying, who shared every bit of the experience during my PhD journey, and it has been my fortune to have your companion and friendship. Thanks also go to Ryan Urquhart, Amanda Otley, Fran Pontin and Sedar Olmez in the LIDA; it has been a great experience to work with you.

Above all, I would like to express my deepest appreciation to my family, especially my husband Fei Qi and my son Yizhi Qi for their unconditional love and perpetual support, and the endless happiness they have brought to me. The thesis would not have been possible without their great support during the pandemic time. I am also indebted to my parents, their visits during the PhD, as well as those video chat, have incredibly brightened up our life. I would finally like to acknowledge my grandpa, who has always made me remind who I am and what I can achieve, even if he never talked about it on any phone time.

Abstract

The tourism landscape in urban destinations has been spatially expanded in recent years due to the increasing prevalence of sharing economy accommodation and other tourism trends. Tourists now mix with locals to form increasingly intricate population geographies within urban neighbourhoods, bringing new demand into areas which are beyond the conventional tourist locations. How these dispersed tourist demands impact local communities has become an emerging issue in both urban and tourism studies. However, progress has been hampered by the lack of fine granular travel data which can be used for understanding urban tourist patterns at the small-area level.

Paying special attention to tourist grocery demand in urban destinations, the thesis takes London as the example to present the various sources of LBSN datasets that can be used as valuable supplements to conventional surveys and statistics to produce novel tourist population estimates and new tourist grocery demand layers at the small area level. First, the work examines the potential of Weibo check-in data in London for offering greater insights into the spatial travel patterns of urban tourists from China. Then, AirDNA and Twitter datasets are used in conjunction with tourism surveys and statistics in London to model the small area tourist population maps of different tourist types and generate tourist demand estimates. Finally, Foursquare datasets are utilised to inform tourist grocery travel behaviour and help to calibrate the retail location model.

The tourist travel patterns extracted from various LBSN data, at both individual and collective levels, offer tremendous value to assist the construction and calibration of spatial modelling techniques. In this case, the emphasis is on improving retail location spatial Interaction Models (SIMs) within grocery retailing. These models have seen much recent work to add non-residential demand, but demand from urban tourism has yet to be included. The additional tourist demand layer generated in this thesis is incorporated into a new custom-built SIM to assess the impacts of urban tourism on the local grocery sector and support current store operations and trading potential evaluations of future investments.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API	Application Programming Interface
AOI	Areas of Interest
ATD	Average Trip Distance
CAZ	Centra Activities Zone
CRS	Comparative Relative Strength
DBSCAN	Density-Based Spatial Clustering of Applications with Noises
DCMS	Department for Digital, Culture, Media & Sport
DMO	Destination Management Organisation
EM	Expectation-Maximization
GBDVS	Great Britain Day Visitor Survey
GBTS	Great Britain Tourism Survey
GLA	Great London Authority
IPS	International Passenger Survey
IRTS	International Recommendations for Tourism Statistics
KDE	Kernel Density Estimation
LBSN	Location-based Social Networks
LCF	Living Costs and Food Survey
LDA	Latent Dirichlet Allocation
LISA	Local Indicator of Spatial Association
LSOA	Lower Layer Super Output Area
NLP	Natural Language Processing
NTS	National Travel Survey
OA	Output Area
OAC	Output Area Classification
ONS	Office for National Statistics
POI	Point of Interest
SDE	Standard Deviation Ellipses

SIM	Spatial Interaction Model
SOM	Self-Organising Map
TDI	Tourism Destination Images
TF-IDF	Term Frequency - Inverse Document Frequency
UGC	User-generated Content
UNWTO	United Nations World Tourism Organization
VFR	Visits to Friends and Relatives
WZS	Workplace Zone Statistics

Chapter 1
Introduction

1.1 Background

Over the last few decades, urban tourism has undergone huge growth and become an important contributor to the urban economy in many cities. According to the definition of International Recommendations for Tourism Statistics (IRTS) (United Nations World Tourism Organisation (UNWTO), 2008), a visitor refers to a person who visits a main destination outside his/her usual residence and spends less than a year for any main reasons including business, holidays or other personal purposes rather than being employed in a paid job (UNWTO, 2008, 2.9). A domestic, inbound or outbound visitor can either be an overnight tourist or a same-day visitor (day trip visitor). (UNWTO, 2008, 2.13). Considering the main purpose of a tourism trip, tourists can be typically classified as business and professional, holiday/leisure and recreation, visits to friends and relatives (VFR), education and training, health and medical care, religion/pilgrimages, shopping, transit and other (UNWTO, 2008, 3.14).

This thesis focuses on urban tourists who follow the definition of a visitor but only for tourism activity which takes place in an urban space. At these urban areas, tourists and day visitors mix with residents and commuters to form a complex fabric of small-area populations. But, how urban tourists use cities, in terms of travel routes and consumption patterns within the city, remain elusive questions in urban tourism research (Ashworth and Page, 2011). One of the main problems has traditionally been accessing data on tourist travel behaviour. Conventional statistical sources, from official and industry surveys, report tourist arrivals at the relatively coarse regional level and it is much more difficult to collect data of tourist space-time behaviour at finer spatial scales. However, advances in population tracking technologies have begun to generate new and novel location-based datasets which may capture some spatiotemporal trajectories of tourists at the individual level (Shoval and Ahas, 2016). For example, data is increasingly available from GPS devices and other tracking datasets generated from mobile roaming, Bluetooth and WiFi. However, these tracking datasets are either hard to access at the intra-urban level and it is usually difficult to distinguish tourists from local residents, which is a primary requirement in order to examine tourist travel patterns (J. Li et al., 2018). In contrast, continually produced and freely accessed location-based social networks (LBSN) data has been increasingly available to analyse human behaviour in space and time in an urban context (Martí, Serrano-Estrada, et al., 2019), including urban tourism research, with many examples of studies which have been able to explicitly identify tourist users. These LBSN data can indicate the presence and activities of tourists, enabling the space-time modelling of urban tourist distributions and dynamics and, in turn, offering the basis to quantify the local social and economic impacts of tourists at a range of geographical scales.

At the same time, within major urban tourist destinations, there is increasing evidence that tourists are expanding their activity patterns outside the usual core areas, dominated by hotels and principal tourist destinations (museums etc). (Maitland and Newman, 2014; Smith and Graham, 2019). For example, the proliferation of short-term accommodation rentals via platforms such as Airbnb, and on-demand tourist transport platforms such as bike-sharing service Mobike, allow tourists to explore and experience more peripheral urban areas which have traditionally served predominantly local residents and workers. Alongside other trends, including new event spaces being created throughout the city (attracting more day trippers) in urban destinations, tourists have become more intertwined with the daily lives of residents and the spatial extent of tourism has been expanded into suburban areas in many cities. These ongoing shifts and spatial expansions have been captured in various LBSN data sources by their users when they visit cities (in a continuous real-time way). Thus, these LBSN platforms offer enriched and refined datasets to help diagnose the changing spatiotemporal details of tourist travel and consumption behaviours.

As tourists penetrate into residential neighbourhoods, they become part of the ‘service population’, sharing local facilities and services with non-tourist populations. Among the array of facilities and services available, grocery and food stores are important to tourists, but tourist grocery shopping behaviours have rarely been examined as a formal topic in both retailing and urban tourism studies. This probably reflects the traditional tourist accommodation model – staying in central city hotels with no self-catering facilities. However, along with those tourists staying with friends and relatives, short-term accommodation rentals such as Airbnb (with often well-equipped self-catering facilities) have become a popular alternative to city centre hotels. These accommodations are more likely to encourage overnight tourists to purchase groceries and cook for themselves. Increasing volumes of urban day trippers have also heightened food shopping demand around key attractions that they visit. Against this background, a clear, quantified, and reliable understanding of the tourist population distribution and associated grocery demand could help service providers – such as retailers - to incorporate this demand into predictions of current and future store performance, which forms the basis of this thesis.

In terms of retail store performance, the spatial interaction model (SIM) has been accepted as a reliable and accurate spatial modelling technique earning a great reputation in grocery retail location analytics (Birkin et al., 2017). To adapt to increasingly complex demand side dynamics, SIMs have been extended over time from those capturing purely residence-based demand to those which now incorporate multiple types of non-residential demand present within store catchment areas. It has seen success in applications which capture the spatiotemporal fluctuation of demand driven by workplaces, schools and universities for example (Waddington et al., 2019). In terms of tourist demand, Newing (2013) and Newing et al. (2015) were the first to incorporate seasonal demand into SIMs focusing on spatiotemporal demand fluctuations in coastal holiday resorts. However, to date, SIMs have not been extended to include retail demand generated by tourists in urban areas. The reasons for this are likely to be multifaceted – as noted above, the lack of opportunities for tourists to cater for themselves in the past, the fact that grocery shopping activities have rarely been separately included in headline survey-based statistics (UNWTO, 2008) and the assumption that expenditures on groceries are usually relatively lower than other spending categories such as transport, accommodation, etc. However, in urban destinations which host a high volume of tourist visits, the spending of these urban tourists should no longer be neglected and excluded from the retail location modelling within the grocery sector. In that sense the work presented in this thesis aims to bridge the retail location modelling and urban tourism research literature for the first time.

Thus, overall, the aim of this thesis is to explore the potential of harnessing multisource LBSN data to more fully understand tourist patterns in urban destinations and thereby estimate tourist population at the small-area level. These insights, including the additional local demand ‘layer’ generated within this thesis can be added to demand variables in SIMs and help to improve decision-making in applied retail location analysis and planning.

1.2 Research question, aims and objectives

The main research question in this thesis is "***how can location-based social media data be used to explore the spatial behaviour and spending patterns of urban tourists and contribute to more accurate retail location modelling in the grocery sector of urban destinations?***" In light of this, the overarching research aims can be stated as follows:

1. To explore tourist activity and mobility patterns of a specific tourist source market from an under-exploited LBSN data with a particular focus on tourist shopping behaviour (Chapter 4 and Appendix A).
2. To model tourist population distributions by incorporating different tourist groups at the small-area level using collated data from multisource LBSN and conventional statistical sources, thereby developing a series of tourist demand layers for grocery shopping (Chapter 5 and Appendix B).
3. To develop a tourist SIM on the basis of the newly generated tourist grocery demand layer (and to calibrate the new model also with the help of novel LBSN data) to demonstrate the store revenue uplifts due to additional tourist demand in the catchment. (Chapter 6 and Appendix C).

To achieve these research aims, the research has the following objectives:

1. To review the literature on the use of LBSN data which offers insights into urban tourist patterns and spatial behaviour related to retailing in particular (Chapter 2).
2. To show the spatial expansion of urban tourism into residential areas, thus identifying the importance and challenges of linking urban tourism with retail location analysis (Chapter 3).
3. To design a 'LBSN data analytics' method to shed novel light on urban tourist travel behaviour in terms of activity preferences and mobility patterns (Chapter 4).
4. To explore tourist activity and mobility patterns of a specific tourist source market from an under-exploited LBSN data with a particular focus on tourist shopping behaviour (Chapter 4 and Appendix A).
5. To model tourist population distributions by incorporating different tourist groups at the small-area level using collated data from multisource LBSN and conventional statistical sources, thereby

developing a series of tourist demand layers of grocery shopping (Chapter 5 and Appendix B).

6. To develop a tourist SIM on the basis of the generated tourist grocery demand (and to calibrate the new model also with the help of novel LBSN data) to demonstrate the store revenue uplifts due to additional tourist demand in the catchment. (Chapter 6 and Appendix C).
7. To demonstrate that the SIM is capable of assessing tourist impacts on local grocery provision in urban destinations and assist store location planning via 'what-if' analysis (Chapter 6).
8. To, in summary, reflect on the findings and methodology of the work and discuss its practical implications for urban tourism and retail location planning and discuss the current limitations and potential improvement of the project with a future research agenda (Chapter 7).

1.3 Study area: defining and contextualising tourism in London

London is one of the most famous world cities. London is not only the home of 9 million residents and 800,000 commuters every day, it also attracts tourists and visitors with a diversity of travel purposes (Great London Authority (GLA), 2021). It receives 53% of inbound visits and 55% of inbound spending in the UK (VisitBritian, 2020). In the case of Chinese inbound visits, 94% of the direct flight seats between China and the UK are in London (VisitBritain, 2019a), which means London is a must-visit destination for most Chinese tourists to the UK. Tourism is of great importance to London's economy, making up 11.6% of GDP and accounting for one in seven jobs in the capital (London and Partners, 2017). In 2019, London hosted 21.7 million international tourists alongside 12.2 million domestic visitors, generating spending estimated at £15.73 billion and £3.03 billion respectively (International Passenger Survey (IPS), 2020; Great Britain Tourism Survey (GBTS), 2020). Meanwhile, each year, around 281 million day trippers also contribute approximately £14.46 billion to the economy of the City (Great Britain Day Visits Survey (GBDVS), 2020). The influx of visitors from outside Greater London interweaves with the city dwellers, boosting the daytime population to well over 10 million, among which tourists and visitors are estimated to constitute almost 11% (GLA, 2014). Hence, London provides a case study to develop the aims of this thesis. Although both the inbound and domestic visits in 2020/21 have considerably declined due to the recent Covid-19 pandemic, it has seen a resurgence in domestic tourism in 2021 and international tourism is expected to recover through the implementation of the ‘Tourism Recovery Plan’ (VisitBritain, 2021a; Department for Digital, Culture, Media & Sport (DCMS), 2021).

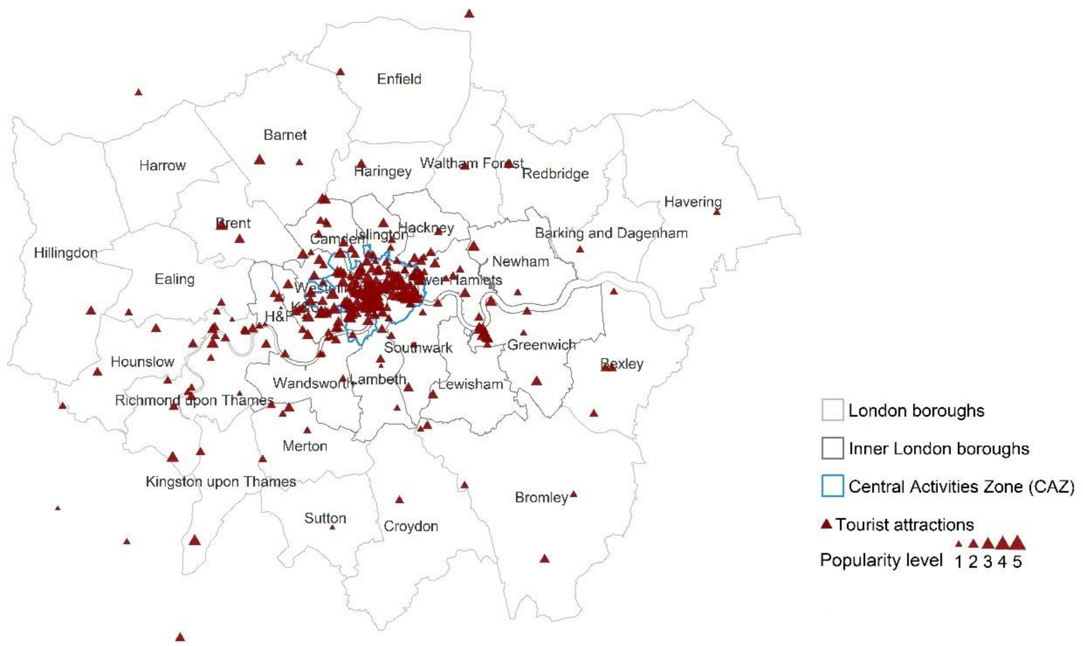


Figure 1.1 Overview of the study area – Greater London.

Tourism in London is no longer a separate activity but a pervasive, inextricably part of the everyday life of the city (Franklin and Crang, 2001; Maitland, 2013). Figure 1.1 provides an overview of the London boroughs and tourist attractions recommended by VisitLondon (popularity level was rescaled in height according to the number of TripAdvisor reviews as a proxy for visitor numbers). The unique collection of London’s iconic attractions is generally located in what is called the Central Activities Zone (CAZ). CAZ is the most vibrant district of London, offering an array of global significant places of culture, business, research and education, retailing, tourism and leisure, major places of worship and access to World Heritage Sites, the parks and the River Thames (GLA, 2021). However, tourism in London is no longer confined to the CAZ. The dense nature of CAZ and its concentrations of serviced accommodation has constrained the amount of space available for other strategic activities and land uses. Therefore, for many years the Greater London Authority (GLA) has sought to disperse tourism away from central London. The latest plan (GLA, 2021) still holds to the target to “promote tourism across the whole of the city”, and encourages new tourist accommodation in areas outside the CAZ, to help spread the benefits of tourism to the whole city (GLA, 2021).

> *"This Plan supports the enhancement and extension of London's attractions particularly to town centres and well-connected parts of outer London, complemented by supporting infrastructure including visitor accommodation, a high-quality public realm, public toilets and measures to promote access by walking, cycling and public transport."*
>
> *- London Plan 2021, GLA*

Coincident with the plan, new forms of short-term rental accommodations have begun to expand tourism into new spaces that were not previously regarded as tourism territories (GLA, 2021). According to Cromarty & Barton (2018), Airbnb, for example, now accounts for around a third of the accommodation sector in London. Airbnb reports that in London 72% of Airbnb properties are located outside the main hotel areas and 43% of visitor spending occurs in the neighbourhoods they stay in (Airbnb, 2018). GLA and Creative Tourist Consultants (2015) also reports a high proportion (58%) of repeat tourists who return to visit London regularly. These experienced travellers are familiar with the city and thus tend to explore areas beyond the conventional tourist precincts, which also makes London a special destination likely to disperse tourists outside the 'tourist bubble' (Judd, 2003; Maitland and Newman, 2009). Consequently, the mobilities as well as the experiences in London of tourists and locals are now blurring: tourists now appear in residential suburban neighbourhoods, parks and shopping malls in more peripheral suburban areas (Inkson, 2019). Whilst studies have been conducted to examine tourism in 'off the beaten track' areas in London, and investigate the associated social and economic impacts in qualitative ways (i.e. Smith and Graham, 2019), the analysis of the impacts of tourism spatial expansion is still under-researched.

As one of the most important global urban destinations, London is an ideal study area for urban tourism research. In fact, it was in London that for the first time the spatial expansion of tourism in urban destinations was elaborated, by Robert Maitland in his seminal studies regarding the changing faces of tourism and its implication for urban development (Maitland, 2007). A series of subsequent studies have untangled some of the trends that have reshaped the tourist landscape of London (see Section 3.3.2) and paved the way for this research to recognise and distinguish ongoing trends of tourism in London that are relevant to grocery retailing. As a quantitative study heavily relying on data, the author has obtained an abundance of administrative statistics and surveys regarding tourism in London, regarded as reliable traditional datasets. In addition, the author has sourced and accessed numerous big geospatial datasets generated across the city every day through various LBSN services. These LBSN data have already drawn considerable interest within urban studies and tourism research areas (see Chapter 2). It will be argued that they also offer great opportunities to examine issues related to urban tourism grocery shopping behaviours.

1.4 General methodology

1.4.1 Research design

The overall methodology of the research involves exploratory analysis of tourist travel behaviour from LBSN data, mapping tourist populations at small-area levels and modelling urban tourist demand in retail location analysis for the grocery sector. As a unique but under-researched LBSN data source in the region of London, Sina Weibo (Weibo) is explored first given its potential for providing detailed spatial travel behaviour of urban tourists from a specific source market: China.

Weibo is the most influential social media service in mainland China, launched in 2009, and has an average 230 million daily active users, 93% of which are mobile app users according to its latest report in 2021 (Weibo Corp., 2021). It allows users to generate microblogs (weibos) with check-ins at every venue location and allows users to add reviews of each venue that can then be publicly browsed. Thus, Weibo shares both the features of microblog services like Twitter and location-based networking services like Foursquare. Nearly 80% of Weibo users are under-30s, and females shows a higher propensity to use than males (Weibo Data Centre, 2021). But, according to the VisitBritain (2021b), Chinese inbound visitor to the UK shows a balanced gender ratio and only half are aged below 34. Therefore, Weibo users are over-represented in the females and younger generation visitor categories. Based on the location of user-generated weibos, 22% of the tourism check-ins concern outbound trips, among which the UK ranked the 6^{th} – higher than by volume of overnight stays, which suggests that Chinese inbound tourists are more likely to post on Weibo while visiting the UK (Weibo Data Centre, 2016; VisitBritain, 2021b). Meanwhile, according to Nielsen, (2017), 97% of Chinese tourists purchase a data package or make other such arrangements for using their smart device while overseas. The report also indicates that 91% of Chinese tourists use Wi-Fi hotspots, either those free to the public or by renting local Wi-Fi devices, while 69% of

Chinese tourists directly use mobile data packages. The high proportion of internet connections also enable Chinese visitors to share their experience on Weibo when they travel abroad.

As one of the largest international tourist source markets (and top global tourism expenditure group) Chinese inbound tourism has been one of the most important in London in the 2010s before the Covid-19 pandemic. They are especially notable for their expenditure on shopping as observed by (VisitBritain, 2021b). In fact, in 2019 Chinese inbound tourism represented the thirteenth largest group of inbound visitors in London but the second most important in terms of expenditure (VisitBritain, 2020). Whilst mainstream LBSN platforms such as Twitter build up valuable data relating to tourists from various countries, the Chinese tourists are more problematic in the sense that those services are neither commonly used by Chinese. Hence, although Weibo data sources cannot fully represent Chinese inbound tourist activity in London, it provides great opportunities to specifically explore Chinese tourist behaviour. Therefore, this research first exploits this LBSN dataset from Weibo to investigate the travel patterns of Chinese tourists in London with a specific interest on their shopping activities.

After the opportunities and limitations of one single LBSN data are discussed, the research then explores multisource LBSN data as indicators of tourist presences and links these data with statistical and survey data to map tourist population distributions at the small-area level. The tourist population modelling in this research is made up of both overnight tourists and day trip visitors, mapped at the Lower Layer Super Output Areas (LSOA) level across London. There are 4,836 LSOAs in London and the average population is 1,722 based on 2011 Census data. In line with the main choices of tourist accommodation according to the IPS and GBTS, overnight tourists mainly consist of three types: travellers choosing traditional serviced accommodation, free guests staying with friends and relatives or in their own second or holiday home, and tourists staying within short-term rental accommodation such as that provided via Airbnb. Along with day visitors informed by GBDVS, the four subgroups of tourists and visitors in London comprised the tourist population considered in this research.

The final fine-scale tourist population maps offer the possibility to assess tourism impacts at the local level in a quantitative way. By estimating the magnitude of tourism at the local level, and estimating those tourists' expenditures, this research produces a series of estimated tourist grocery demand layers across the LSOAs of London. These grocery demand layers can in turn be added into the retail location modelling process to develop a new SIM which can also be calibrated with the help of tourist dynamics extracted from LBSN data. Finally, the practical implications of the new tourist SIM in retail location decision-making can be explored.

1.4.2 Data collection

This work draws on a variety of data sources including administrative tourist and visitor statistics, census population and expenditure data, along with locally specific data from the open surveys, social media, and crowd-sourced datasets to support the spatial modelling of urban tourist patterns. For each of the four tourist types, the research first endeavours to find the latest official sources for the number of tourist visits and nights in London at the finest geographical scale. The standard tourist surveys and statistics are available only at the borough or even the city level and lack information at finer spatial scales. Therefore, tourism-related datasets from different types of LBSN services, as well as the point of interest (POI) at the tourism supply side, are considered as auxiliary to complement the survey data. These emerging datasets are offered as the travel behaviour of individual tourists or utilisation of each venue, which thereby can be aggregated into any possible geographical area level. All the datasets used in this thesis are listed in Table 1.1 with the description and source information. For each of these data sources, this research attempts to retrieve the latest version available.

Table 1.1 Datasets utilised in the thesis.

Range of datasets	**Data description**	**Data source and year**	**Used and detailed in the thesis**
Geography boundary and population data	London boundary (by LSOA)	London Datastore	Chapter 4, 5, 6, Appendix A, B and C
	Ethnic group of usual residents (by LSOA)	Census 2011	Chapter 5 and Appendix B
	Households (by LSOA)	Census 2011	Chapter 5, 6, Appendix B and C
	Living Costs and Food Survey (LCF)	ONS, 2017	Chapter 5, Appendix B and C

	Output Area Classification (OAC)	ONS, 2011	Appendix C
	Workplace population statistics	Census 2011	Appendix C
Tourism-related datasets	Inbound tourist nights, by accommodation type, by origin of country	IPS, 2018	Chapter 5 and Appendix B
	Domestic visitor nights, by accommodation type	GBTS, 2018	Chapter 5 and Appendix B
	Daytime population survey (by borough)	GLA, 2014	Chapter 5 and Appendix B
	Accommodation stock Audit (by borough)	VisitEngland, 2016	Chapter 5 and Appendix B
LBSN datasets	Weibo check-in data	Sina Weibo, Jan. 2016 – Aug. 2018	Chapter 4, Appendix A
	Foursquare venue	Foursquare Inc., 2019	Chapter 4, 6, Appendix A and C
	Foursquare movement	Foursquare Inc., 2019	Chapter 6 and Appendix C
	Geotweets	Twitter, Sept. 2018 – Oct. 2019	Chapter 5 and Appendix B
	Airbnb listing and reservation data	AirDNA, Jun. 2018	Chapter 5 and Appendix B
Point of Interests (POIs)	Service accommodation establishment	OpenStreetMap, Feb. 2020	Chapter 5 and Appendix B
		Ordnance Survey POI, Feb. 2020	Chapter 5 and Appendix B
	Grocery store	Geolytix Retail Points, Jan. 2021	Chapter 5, 6, Appendix B and C
		CACI grocery store points, 2014	Chapter 5, 6, Appendix B and C

1.4.3 Spatial modelling

Various spatial analysis and urban modelling techniques are utilised in this research to comprehensively examine the potential of LBSN data in analysing urban tourist patterns and to link urban tourist demand to retail location analytics in urban destinations. Three main spatial modelling methodologies are applied in the research corresponding with the three principal aims and the substantive chapters. The detailed application of each proposed methodology is thoroughly presented in Chapters 4, 5, and 6 respectively. Thus, this section only briefly introduces the adopted techniques.

1.4.3.1 Data preparation and exploratory analysis of LBSN data (Chapter 4)

First, a number of methods were needed to collect, clean, extract and analyse Chinese tourist spatial behaviour from Weibo check-in datasets. After distinguishing the tourist users from local and long-stay users in London, the spatial, temporal and contextual information associated with tourist whereabouts and movements is analysed by a series of spatiotemporal analysis techniques including geovisualisation, spatial clustering, point pattern analysis, network analysis and topic modelling. The 'LBSN data analytics' method shows the activity preferences and mobility patterns of Chinese tourists from their Weibo check-ins and informs a tourist segmentation to identify the divergent multipurpose travel behaviour of each cluster, with a special interest in their shopping activities.

1.4.3.2 Spatiotemporal population modelling and demand estimation (Chapter 5)

Spatial analysis techniques are used to compile finer scale tourist population spatial-temporal distributions for both the day and night time populations. The modelling process involves combining a wide range of disaggregated census data, survey data, together with LBSN data as covariate datasets to estimate tourist density for constructing tourist population maps across the LSOAs of London. Different grocery shopping expenditure rates are assumed for varied types of tourists to generate a set of small-area tourist grocery demand layers across London. The combined tourist grocery demand layer is then linked with the supply side of grocery stores by bivariate analysis to initially indicating the areas with unmatched tourist demand and grocery supply in London.

1.4.3.3 Spatial interaction modelling (Chapter 6)

A custom-built tourist SIM is developed and calibrated based on the tourist grocery demand layer in London produced in Chapter 5. The model calibration process is facilitated by utilising tourist shopping trips simulated from Foursquare movement datasets. The simulated tourist expenditure flows estimated by the new SIM are then used to inform a tourist grocery accessibility index for local grocery provision. 'What-if' analysis is applied in three realistic development plans to evaluate the practical implications of the new model.

1.5 Thesis structure

The rest of the thesis is organised with the following structure to meet the research aims and objectives as stated in Section 1.2.

Chapter 2 provides a comprehensive literature review around the potential of LBSN data for understanding tourist patterns in urban destinations (in line with objective 1), which mainly covers the application of different types of LBSN data sources in understanding the spatial patterns of tourist spatial behaviour in terms of their activity and mobility (particularly related to shopping). The advantages of using multisource LBSN data is highlighted and the limitation and concerns of LBSN data analytics in tourism research is also discussed.

Chapter 3 reviews the recent history of applied spatial interaction modelling and the attempts to disaggregate the demand side variable to include non-residential demand. It addresses objective 2 by highlighting the needs and challenges to incorporate tourist grocery demand in retail location modelling considering the new tourism trends in many urban destinations.

Chapter 4 addresses objectives 3 and 4 by using Weibo check-in datasets in London to explore the spatial behaviour of Chinese tourists during their stays in London. The analysis is applied both at the individual and collective level to fully exploit the potential of LBSN data in gaining knowledge of tourist travel behaviours regarding their activity preferences and mobility patterns. A tourist segmentation is created based on tourist multi-purpose travel behaviours. The shopping activity and associated location choice is a special focus in this section. This chapter was published as a peer-reviewed paper in EPB: Urban Analytics and City Science.

Chapter 5 addresses objective 5 by constructing tourist population layers in London using data collated from both conventional statistic sources and multiple novel LBSN data. The population density maps inform the tourist demand layer for grocery shopping to illustrate the spatial variations of demand uplifts and suggest areas with potential unfulfilled tourist needs. This chapter was published as a peer-reviewed paper in the Journal of Retailing and Consumer Services.

Chapter 6 addresses objectives 6 and 7 by developing and calibrating a tourist SIM based on the tourist demand layer generated in Chapter 5. The benefits of incorporating tourist demand in retail location modelling are presented and discussed by utilising the model outputs to estimate revenue uplifts, evaluate grocery provision changes and forecast the performance of new retail development plans, with the additional tourist demand included. Tourist grocery shopping trips extracted from Foursquare datasets are used to aid the model calibration. This chapter was reworked as a paper that is under editorial consideration.

Chapter 7 provides an extended discussion around the findings and a final conclusion of this work (meeting objective 8). The methodological opportunities and challenges, practical implications, limitations and future research agenda of the study are discussed in this chapter.

Chapter 2 Harnessing LBSN data in tourist travel behaviour research

2.1 Introduction

There is increasing interest in the opportunities that social media data present for researchers in many areas of the social sciences, including tourism. Given the amount of social media data becoming available, it is not surprising that it is one of the most prominent forms of volunteered Big Data (Kitchin, 2013). Nearly one third of the world's population are currently social media users and with the development of the mobile internet, 42% access social media through mobile platforms (Statista, 2018). Nearly all these social media services have been developed with location-based features and the built-in GPS equipped smartphones augment user-generated content (UGC) with a spatial dimension. These services allow users to opt in to attach the geolocation to their UGC, but It is worth noting that not everyone agrees to reveal their geolocation. These UGC with geolocation information are referred to as LBSN data (Roick & Heuser, 2013; Steiger, Albuquerque, & Zipf, 2015).

Most LBSN data are generated in one of three different ways. The first is known as geotagging, which is the annotation of location information from microblog posts (Twitter), photographs (Flickr and Instagram) and videos (Youtube). In the context of tourism, users can share their experiences by texts, photos or video clips with the geolocation on social media platforms. The second is geosocial networking, which aims to share activities along with current whereabouts, such as via Foursquare, Weibo and Facebook Places (Roick and Heuser, 2013). The services of this type encourage users to check-in at venues when they physically close to these places and allow tourists to record these visits at destinations. At the same time, traditional online review sites such as TripAdvisor, Booking.com and Yelp, or online peer-to-peer marketplaces such as Airbnb, contain a large amount of location-based UGC and have also been developed with LBSN features. Tourists use these websites to exchange travel experiences by reading and generating reviews and book services.

The tourist in this thesis refers to the overnight tourist or day trip visitor. Based on the tourism statistics of IPS (2019) and GBTS (2019), the overnight tourists in London, can be either overseas inbound tourists or domestic visitors staying overnight in London and are further classified by their temporary accommodation types, including short-term rentals, hotels and other serviced accommodation and staying with relatives and friends / own home. On the other hand, the day trip visitors are mainly domestic tourists or Londoners who have a longer than 3 hours same-day visit in London, according to the definition by GBDVS (2020). Tourists may use any of the aforementioned three LBSN services to record and share their travel experiences, but the overnight tourists are more likely to post reviews and rate the services on tourist websites with day trip visitors more likely to only use geotagged social media and geosocial networking services.

The value of LBSN data in addressing different tourism issues has been partly reviewed by a number of excellent, complimentary papers. Shoval and Ahas (2016) offered a useful historical account of the various data sets used for tracking tourists to date, beginning with GPS and Bluetooth and then considering smartphones and finally LBSN data. In a comprehensive literature review of big data in tourism research by Li et al. (2018), the focus was largely on the characteristics, analytics techniques and research focuses of different big data sets including LBSN data. In addition, the analysis framework of some potential applications of LBSN data relating to more specific tourism research questions have been reported: for example, tour itinerary recommendations (Lim et al., 2018), travel demand modelling (Abbasi et al., 2015) and tourist behaviour analytics (Miah et al., 2017). However, to date, there is limited review about how different types of LBSN data could contribute insights into the tourist spatial patterns in urban destinations, although in the urban studies domain Martí, Serrano-Estrada, et al. (2019) contributed a review of the opportunities and challenges of different social media datasets in the context of urban phenomena research. Therefore, this chapter attempts to fill this gap in the literature to present a comprehensive review on using different kinds of principal LBSN data in tourism research, with a focus on how these data can help to understand tourist behaviour in greater detail. The review focuses on the principal LBSN data sources, geotagging social media data, location-based check-in data and tourism service website data, followed by articles which have used combinations of these in tourism research. Following the review some key issues around the application of LBSN data for tourist research (i.e. some pros and cons) are discussed.

2.2 Understanding tourist spatial behaviour using LBSN data

The short length of stay, limited knowledge and preconceived expectation of tourists result in the severely restricted space-time budgets of tourists compared to local users (Ashworth and Page, 2011). Prior work has identified that tourists use the host city in a geographically selective and temporally changeable way (Shoval et al., 2011; Grinberger et al., 2014). Thus, tourist travel behaviour in urban destinations is usually hard to capture or predict (Lew and McKercher, 2006). Traditionally, tourist travel data are collected by tourism authorities through surveys, which are limited in both scope and nature. There is a deficiency of official tourism statistics at a fine spatial and temporal resolution capturing tourist travel behaviour. By contrast, LBSN data supplies a vast amount of digital tracking data of tourist spatial behaviour at the individual level, spatially and temporally disaggregated, enabling a sizeable tourism literature to distil knowledge for destination management organisations (DMOs) (Li et al., 2018).

2.2.1 Geotagged social media data: incidence and dynamics

Geotagged social media data is a major data source for the analysis of tourist travel behaviour. The geotags are the exact locations of individuals, thus often being taken as tourist footprints, representing the places visited at the time of the posts being shared. Therefore, a fundamental contribution of this type of LBSN data in tourist spatial behaviour studies is to highlight the most visited locations and understand tourist behaviours and movements in more detail.

Although different social media platforms generate distinctive LBSN datasets there are generally three common stages undertaken in relation to their use: (1) tourist identification: data cleaning and pre-processing to identify who are likely to be tourist users; (2) geovisualisation and hotspot analysis: tools to detect spatially-dependent patterns and conduct exploratory analysis; 3) trajectory analysis: construction of tourist itineraries and movements over time and space. Each of these tasks will be elaborated on further in the following sections, in which we consider the key LBSN platforms in tourism research, Twitter and Flickr.

Twitter is one of the most widely used social media platforms globally. The high granularity of spatiotemporal information harvested from the geotags enables researchers to investigate tourist behaviour and locations at both the individual level and for any aggregations required. Twitter allows users to generate messages of up to 280 characters and provides geographic coordinate tags from the GPS sensor of their mobile devices or associated venues sourced from Foursquare (see below). For analysts, Twitter provides a stable and flexible APIs to filter real-time and historical geo-tagged Tweets samples (these geo-tags can relate to countries, towns, places within cities or even actual cartesian coordinates). Using the API, it is possible to download a collection of the most recent 3,200 tweets of each user or a sample of Tweets within a given locality.

On the other hand, Flickr is the most extensively used photo-sharing service in academic research due to the open accessibility of the data stream. Spyrou and Mylonas (2016a,b) suggest that among all the social science applications of Flickr, travel and tourism are the most discussed topics and that Flickr has gained more extensive usage than other LBSN sources within the tourism academic community. Flickr generates substantially less data than Twitter in a given time span, but the photo density is considerably higher at tourist attractions than Twitter, especially at recreational and protected areas (L. Li et al., 2013; Tenkanen et al., 2017; Seresinhe et al., 2018). Twitter only offers sample data and only 1-2% are geotagged, but Flickr offers flexible and advanced APIs allowing users to download nearly all the public content for free. A proportion of these photos are geotagged (varying from 5% to 15% by locations (Hauff, 2013)) with precise coordinates and semantic textual tags, along with other photo metadata, enabling additional analysis for locations (Bahrehdar et al., 2020).

There are a number of application types within tourism research which have used geotagged social media data. First, it is possible to use geotagged social media data to develop metrics to estimate tourist visits around attractions or during major events, potentially offering reliable estimates of tourist numbers for specific localities or events (Wood et al., 2013; Steiger, Westerholt, et al., 2015; Kim et al., 2019). This is useful, as often there is limited published data on tourist movements and total visitor numbers (outside the major attractions) (Kádár, 2014; Kim et al., 2019). For instance, the geotagged Flickr photos have been used to quantify tourist activities around coral reef tourism at a global scale (Spalding et al., 2017). Sessions et al. (2016) used Flickr to predict monthly visitor numbers to national parks. The continuous tracking of geotagged social media have also been addressed to monitor how tourist visitor numbers change over time (Hu et al., 2015; Barchiesi et al., 2015; Tenkanen et al., 2017; Payntar et al., 2021). Importantly, Hamstead et al. (2018) showed that daily park visits estimated from geolocated Flickr and Twitter data were highly correlated with the empirical observed visitors.

The variations in the geographical distribution of geotagged social media data can also show the density surface of tourist distributions within a locality. This can be used to pinpoint the areas of intensive visits in the destination cities. Geovisualisation tools such as grid density or Kernel Density Estimation (KDE) can help map those variations. To avoid the data skewness caused by spambots and extremely active users, usually, unique tourist users rather than individual geotagged posts are used to create the tourist density surfaces. In a study of Cilento (an important tourist area in southern Italy), Chua et al. (2016) utilised geotagged Twitter data to procure an overview of tourist density distribution based on the total number of unique visitors in a fixed raster grid (Figure 2.1). Moreover, different types of tourist activity can produce very different spatial patterns or concentrations. The temporal profile of the data serves to present the changes of tourist distribution during the different time of a day (Vu et al., 2015; Zhou and Zhang, 2016). Figure 2.2 shows various KDE surfaces for four different leisure activities according to the topics within geotweets in London identified by Lansley and Longley (2016b). The darker tone in each activity type indicates a higher density of geotweets belonging to the corresponding topic.

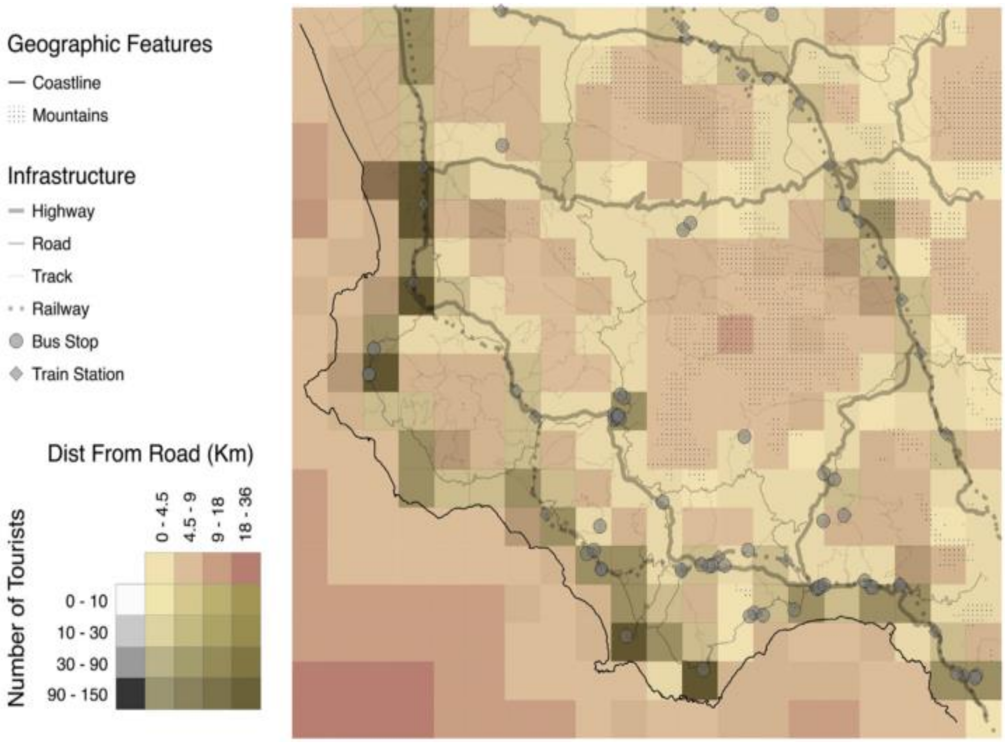


Figure 2.1 Location popularity and accessibility to main transport infrastructure based on the total number of unique visitors in each cell of the grid. Source: Chua et al. (2016).

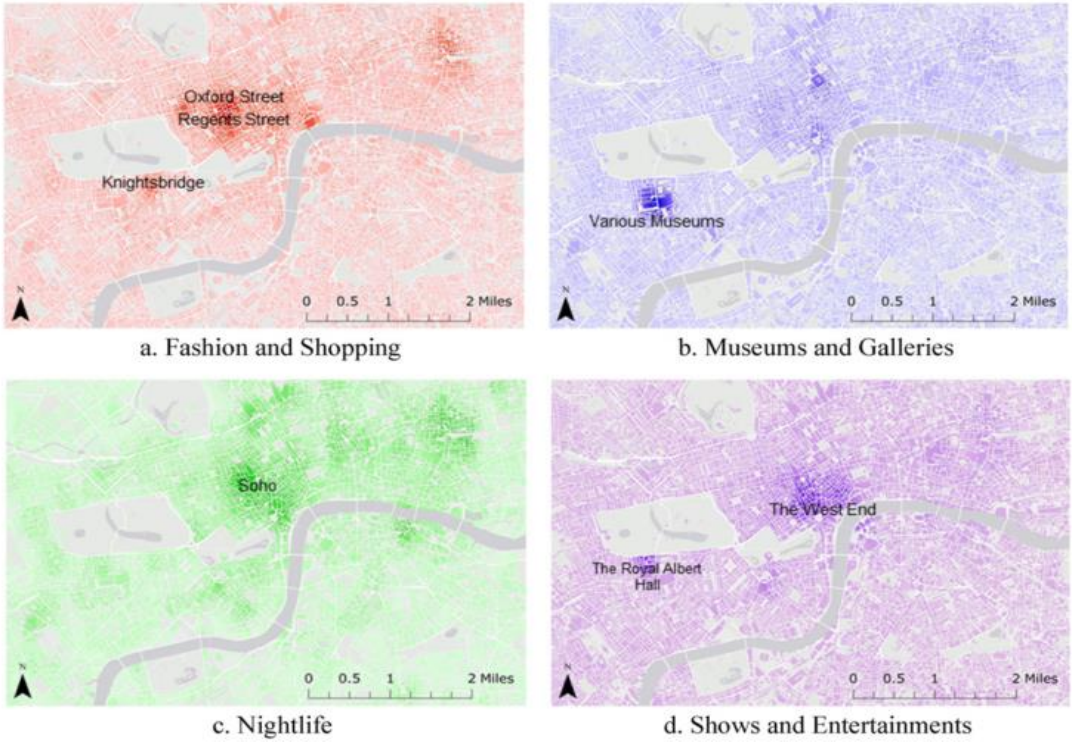


Figure 2.2 A kernel density smoothing of georeferenced Tweets for 4 subgroups of ‘Leisure and Attractions’. Source: Lansley and Longley (2016b).

The geovisualisation of urban tourist density reflects the selective usage of tourists in the host city. However, the density surfaces of tourists fail to offer deeper insight into tourist activity patterns and are ineffective for outlining movements at the locations. Spatial clustering algorithms have been applied in a finer way to recognise tourist concentration areas in destinations, which are usually named tourism hotspots or Areas of Interest (AOI). The spatial clustering of Flickr geo-photos has been especially effective in identifying sightseeing spots in many urban destinations, whereas geotweet clusters indicate a variety of tourist activity locations (Spyrou and Mylonas, 2016a).

Spatial clustering algorithms have been widely applied in this respect. The density-based clustering algorithm DBSCAN (Density-Based Spatial Clustering of Applications with Noises) and its derivative versions, along with other model-based clustering algorithms such as EM (Expectation-Maximization) and SOM (Self-Organising Map) and spatial autocorrelation tools like LISA (Local Indicator of Spatial Association) and Getis Ord Gi*, have been adopted to identify AOIs (Majid et al., 2015; Steiger et al., 2016; Comito et al., 2016; Miah et al., 2017; Salas-Olmedo et al., 2018; Chen et al., 2019; Giglio et al., 2019). Figure 2.3 shows who used the most adopted three clustering algorithms to detect the most visited AOIs by tourists and residents according to their geotweets collected in Florida (Hasnat and Hasan, 2018).

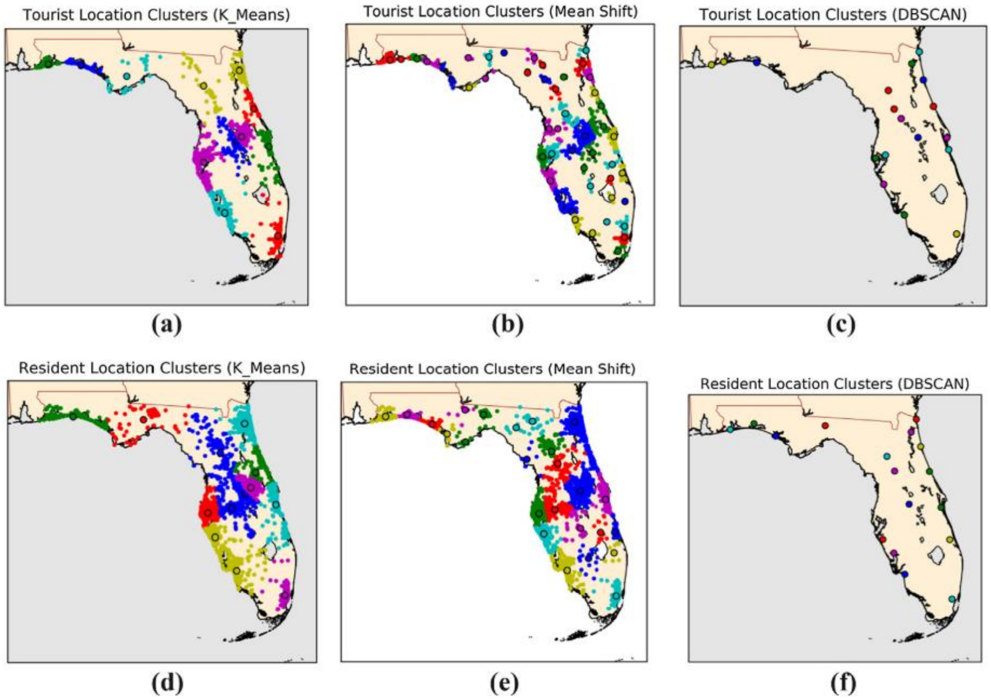


Figure 2.3 Clustering of tourist and resident geotweets in Florida to detect AOIs by three spatial clustering algorithms: K-means, mean-shift and DBSCAN. Source: Hasnat and Hasan (2018).

After identifying the AOIs at the urban destination, AOI popularity can be measured by the geotagged social media data density (Kádár ,2014; Vu et al., 2015) or the comparative relative density of one AOI with respect to the overall activity of the city, referred to as Comparative Relative Strength (CRS) (Girardin et al., 2009). For example, Figure 2.4 shows an example of how the popularity of venues can be calculated based on the geo-photo users, showing the presence of tourists (yellow polygons) and areas of intense activity (the red heat map). The height of the polygon shows the number of individual tourists present, by which the popularity of a POI can be measured. If these tourism hotspots are associated with the textual details and representative photos, it is possible to offer further insights into the sentiment perspective and tourism destination images (TDI) in regards of tourist experience and perception (Memon et al., 2015; D. Li et al., 2018; Liu et al., 2019; Ferreira et al., 2020) For example, Girardin, Calabrese, et al. (2008) revealed the most ancient parts of the city by mapping the distribution of the tag “ruins” in Rome. Thus, many studies have employed Natural Language Processing (NLP) techniques such as TF-IDF (Term Frequency - Inverse Document Frequency) to analyse the representative tags of AOIs and generate annotations for tourist attractions (Hollenstein and Purves, 2010; Hu et al., 2015; Spyrou and Mylonas, 2016b).

Figure 2.4 Presence of tourists in the main AOIs in downtown Florence. Source: Girardin, Fiore, et al. (2008).

The chronological sequences of individual user's geotagged social media data, although not able to reflect every detail of tourist movements, can be used at a collective level to capture the spatiotemporal features of tourist flows (Girardin, Calabrese, et al., 2008; Rashidi et al., 2017; Payntar et al., 2021). Tourist trajectories from geotagged social media data have been used to inform how different tourists connect an assortment of key attractions in their travel routes. Comito et al. (2015) detected 20 popular tourist attractions in London and mapped the most frequent travel routes based on a sequential pattern mining algorithm, by which they constructed a 'reachability graph' of these major attractions, shown in Figure 2.5. Along with the temporal information, these empirical travel routes have been suggested to benefit tourist traffic management during rush hour (Vu et al., 2015; Shi et al., 2017). Taking the detected AOIs as nodes and individual tourist movements between AOI pairs as edges, spatial network analysis (with graph theory) is an effective approach to construct tourism networks around locations, which helps to characterise tourist mobility patterns and location connectivity in cities (Shao et al., 2017; Hu et al., 2018). This graph-based approach helps to study tourist movement patterns by identifying nodal attractions, significant transition hubs, core-periphery districts and itineraries associated clusters from such networks. These can provide insights into the destination management and operations of tourism stakeholders (Liu et al., 2017; Kang et al., 2018; Agryzcov et al., 2019).

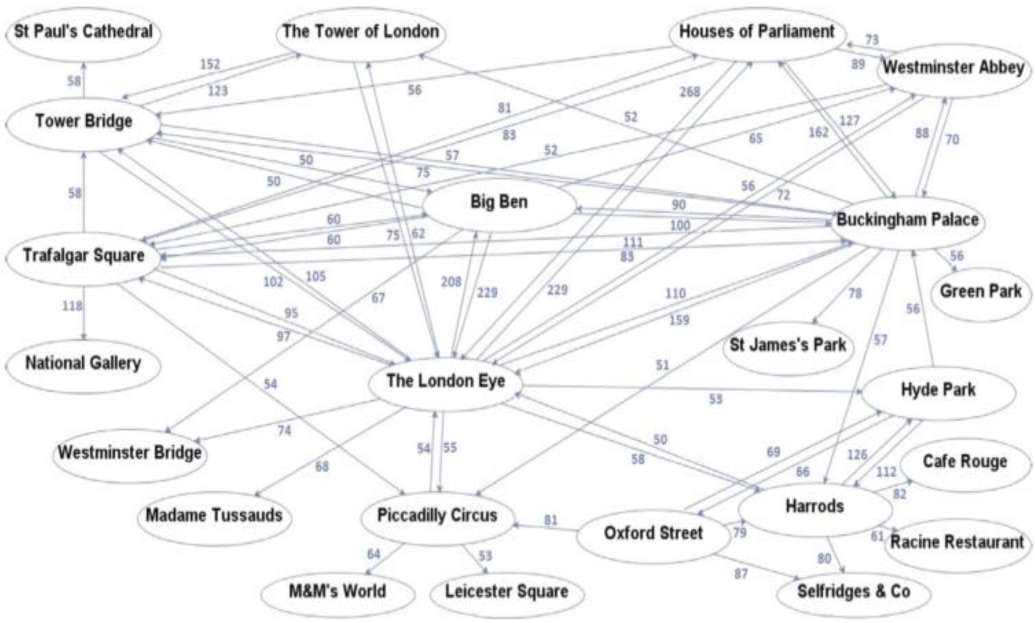


Figure 2.5 Reachability of major attractions in London. Source: Comito et al. (2015).

When the origin of tourists has been gleaned in the tourist identification phase, tourist flows can be disaggregated by demographic features to reveal different travel behaviour among tourists from different countries or in comparison with the locals (Girardin, Fiore, et al., 2008; Gunter and Önder, 2021; Chua et al., 2016). Su et al. (2016) used the intensity of Flickr data to reveal that tourists from different origin countries have spatially and temporally varying travel preferences, and are attracted by different local influences at the destination city (which includes the cultural and natural attractions, congestion and accessibility, infrastructure level, economy and safety factors). The research of Chua et al., (2016) also highlighted how Twitter data can be used to disaggregate the origin of tourists by country, revealing that Greek and Dutch visitors accounted for more than half of the primary inbound tourists in Cilento, and that they had higher mobility and travelled longer distances than tourists from other countries. Girardin, Calabrese, et al. (2008) used the chronologically organised geotagged Flickr data to uncover each user’s travel path through Rome and thereby indicated dominant tourist trajectories. Figure 2.6 shows the paths taken by Flickr users in Rome, (a) for Italian tourists and (b) for American visitors, which revealed that the American tourists tended to have a narrower travel path with less AOIs visited.

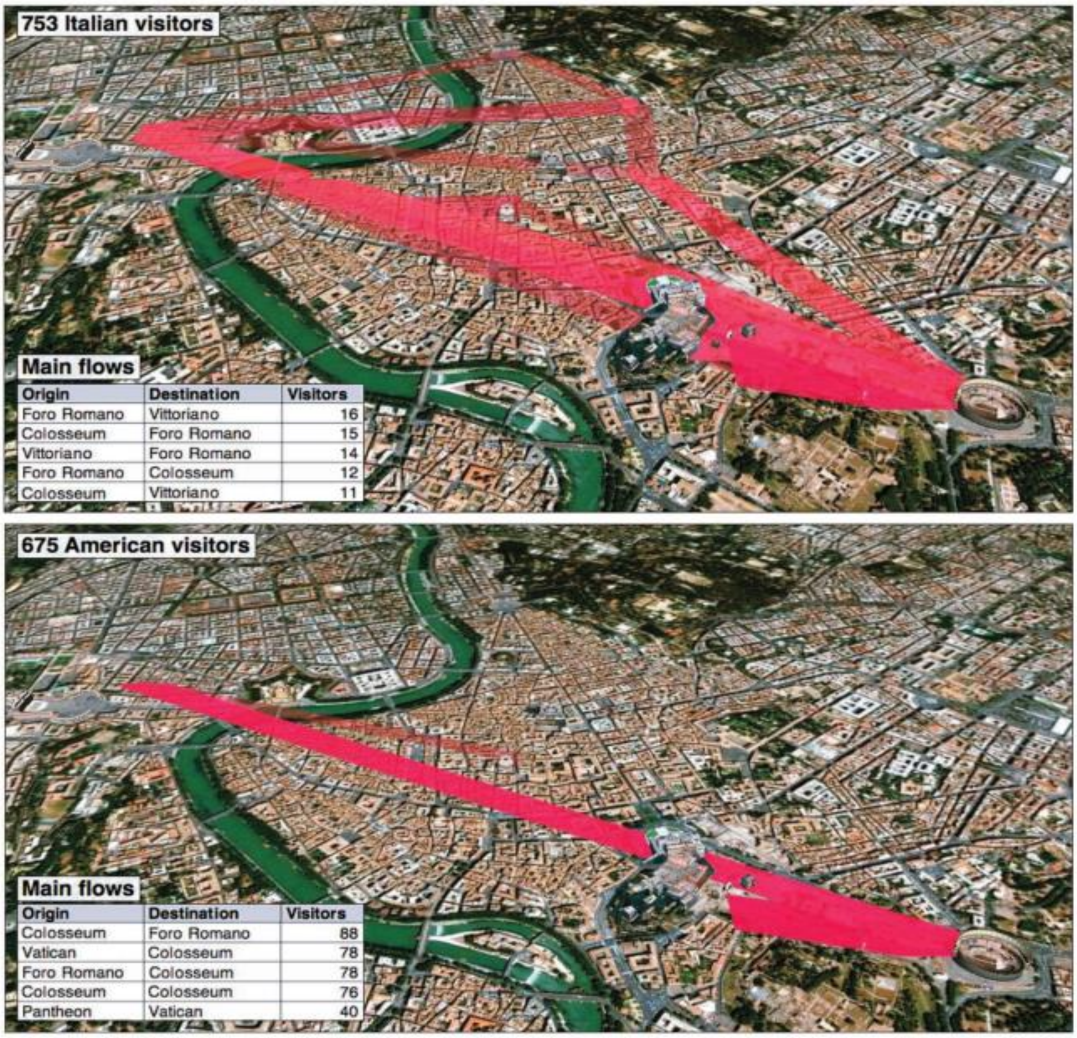


Figure 2.6 Geovisualiation of the main paths taken by tourist and local photographers between AOIs in Rome. Source: (Girardin, Calabrese, et al., 2008).

However, although sufficient observations of spatiotemporal movements from tourists at the individual level have been offered by Twitter, the contextual information of these geolocations are not necessarily contained in these geotweets (unlike the check-in at POIs or hashtag of geo-photos). Therefore, apart from linking geotweets with Foursquare venues (see below), studies have assigned a place category to geotweets by additionally associating the locations with OpenStreetMap (OSM) category or land use type (García-Palomares et al., 2018; Bustamante et al., 2019).

Another approach to infer tourist activity from Twitter, is to explore the nature and contents of the tweets themselves. Whilst the examples presented below do not focus directly on tourism activities, they highlight the nuanced activity-based patterns and insights that can be extracted from these data and which can be applied in tourism studies. Lansley and Longley (2016b) offered a good instance of the use of text analysis in relation to many activities, some more obviously associated with tourism (i.e. photography and insights, leisure and attractions, place and check-ins). Lloyd and Cheshire (2017) extracted retail-related geotweets which they then located within retail centre boundaries in a UK example. The density and spatial clustering of these geotweets revealed the areas with elevated shopping activity, while the mobilities of the Twitter users indicated the accessibility and catchment areas of retail centres. Their work suggests that the shopping activities extracted from geotweets are a robust indicator of retail clusters in urban areas and the retail flows reflect consumer movements in the real world. Also, Lovelace et al. (2016) extracted the geotagged Twitter trips from home locations to retail centres to construct a shopping flow matrix. This retail flow matrix was used to inform the classic gravity model and radiation model and was compared against the spatial interaction flows generated from the theoretical models.

To conclude, it has been shown that geotagged social media data can provide rich and detailed insights on tourist behaviour at a range of spatial scales. However, one of the challenges of academic research is to link those geotagged posts or images to known attractions and other places visited. Foursquare has the potential to add some of that context and its contribution to date is discussed in the next section.

2.2.2 Location-based check-in data: activity and preference

Location-based check-in data is useful as it is the only type of LBSN data that innately relates to human activities which usually occur at popular venues in the city, suggesting various human activities taking place there (Noulas, 2013; Hecht and Stephens, 2014; Fekete, 2015). Foursquare generates the most extensively used location-based check-in datasets in academic research, accessible through collaboration schemes or streaming via the Twitter portal (Foursquare, 2019). Other similar geolocation social networking services to Foursquare also offer location-based check-in datasets, such as Gowalla, Brightkite and Facebook Places (Cheng et al., 2011; Ma et al., 2017; Chen et al., 2018). There are also some specialist LBSN data sources that offer important insights into people from specific locations, such as Sina Weibo (Weibo) for the study of the Chinese and Mastodon for Indians.

1. Foursquare

An advantage of leveraging Foursquare check-in data is that it provides contextual information about POIs in cities (predominantly venues such as restaurants and leisure facilities) and the check-ins can only be generated when the user is physically located close to the venues (Twitter and Flickr do not have such a rigid requirement). Hence, the study of Foursquare check-ins can offer useful insights into tourist activity locations (Ferreira et al., 2016; Maeda et al., 2018; Vu et al., 2019) and the popularity of places with a tendency to focus on consumption related locations (Karamshuk et al., 2013; Yu et al., 2013).

Foursquare venues are hierarchically classified into ten main categories using a broad spectrum of sub-categories (Foursquare, 2019). The ten main categories are: Arts & Entertainment; College & University; Event; Food; Nightlife Spot; Professional & Other Places; Travel & Transport; Outdoors & Recreation; Shop & Service; and, Residence. In the context of urban analytics, it is common that the ten main categories are interpreted as ten human activities as shown in Table 2.1. When tourist users are distinguished from others, they also represent typical tourist activities.

Table 2.1 Ten Foursquare main categories and corresponding activities.

Foursquare main category	**Selected typical sub-categories**	**Activity type**
Arts & Entertainment	Museum, Historic site, Concert hall, Art gallery, Theatre	Visiting arts and entertainment venues
College & University	College academic building, College library, College Residence Hall	University-related activity
Event	Christmas Market, Conference, Music Festival, Parade	Participating events
Food	Café, Tea room, Steak house, Chinese restaurant, Burger Joint	Dining out
Nightlife Spot	Bar, Brewery, Night Market	Nightlife
Professional & Other Places	Monument/landmark, Church, Library, Observatory	Visiting landmarks/buildings

Outdoors & Recreation	Beach, Park, Palace, Scenic lookout, Bridge, Castle	Outdoor sightseeing
Shop & Service	Department store, Supermarket, Souvenir shop, shopping mall, Grocery Store	Shopping
Travel & Transport	Train station, Bus station, Airport, Pier, Taxi Stand	Travelling
	Hotel	Accommodation
Residence	Home, Residential Building	Staying at home

A number of studies have taken advantage of such interpretation to depict the spatiotemporal distribution of various human activities. For instance, Xie et al. (2018) and Vu, Li, et al. (2019) identified the discrepancy between tourist preferences in different destinations of the same tourism source market, according to the categories of their Foursquare check-ins. Ferreira et al. (2020) identified a clear Electronics Enthusiastic tourist group in Tokyo from the collected Foursquare check-ins, which has a strong presence of visiting electronics stores. Martí et al. (2020) diagnosed multi-activity clusters across the city with the aids of Foursquare POIs as the touristic offer and check-ins as the utilisation of tourism services. Yu et al. (2013) used Foursquare check-in data in Paris to estimate and rank the business area popularity of certain categories (food, entertainment, shop, nightlife and cafés) in different market areas. Vu et al. (2018) explored the Foursquare check-ins to find that tourists in Hong Kong are more likely to go shopping on the arrival and departure days and they also identified the most popular shopping malls visited by tourists.

Foursquare has been considered heavily related to the places of consumption such as restaurants, retailers and other leisure facilities which relate to users' consumption activity (Noulas et al., 2013; Fekete, 2015; Agryzcov et al., 2015; Vu et al., 2018a). In retail analysis, the total sum of check-ins at a venue acts as a proxy of the attractiveness of the venue and the retail flows identified from the check-in data provide a real-world dataset regarding consumer movements. Zhou and Zhang (2016) used the Foursquare check-ins to capture the spatiotemporal changes of shopping activity and examined how it interacts with other main human activity. By taking the most visited areas as the activity centre of a user, Qu and Zhang (2013) depicted the catchment area of the retail outlets based on the activity areas of the Foursquare users who have checked-in at those stores. The applications of Qu and Zhang (2013) and Daggitt et al. (2016) made use of the increase of the check-in counts at the venue to signify the potential uplift of business performance. They took user check-ins at different venues belonging to the same category as a means to understand venue competitiveness. Similarly, Sklar et al. (2017) regarded the rise of the check-ins in a commercial area as the growth of urban activity of that district. Given this thesis' objectives in relation to tourist shopping activities, the focus on specific commercial activity types afforded by these data is important for this study.

An added advantage in such applications is to use the retail check-ins and their flows to inform inputs of machine learning models and thereby to predict store performance change and assist site selection in retail planning. The check-ins and flows are valuable empirical data that could quantify the geolocation, visitation and mobility features of a site and its neighbourhood area. These features have been taken as variables to formulate prediction models to forecast the popularity change of different locations by the number of check-ins or visitors (Noulas et al., 2010; Long et al., 2013; Y. Li et al., 2013; Yu et al., 2016; Al Sonosy et al., 2018) or by considering socio-spatial interactions in the region (Scellato et al., 2011; Noulas et al., 2012a; Yu et al., 2013; Doan and Lim, 2019). For example, Karamshuk et al. (2013) used Foursquare check-ins to derive geographic and mobility features of the retail locations in New York. The mobility features included location density, centrality, area competitiveness, and the spatial interactions with other types of venue in the same area; and the mobility features were area popularity, change of transition density in the area, area incoming flow, and the area transition quality accumulated by the probability of transitions between all the type of venues. These features combined to represent the attractiveness, competitiveness and interactions of the sites for location selection. Other research has also generated temporal profiles of venues to add a temporal layer regarding place attractiveness ((Sklar et al., 2017; D'Silva et al., 2018; Hsieh et al., 2019), or have considered the proximity of venues to main attractions or food providers (Georgiev et al., 2014). When using these variables to formulate prediction models, the trained models can forecast the future check-in trends of the candidate locations (increasing check-ins indicating potential prosperous store performance) and also identify the influential factors of the check-in growth (D'Silva et al., 2018; Hsieh et al., 2019).

Similar to traditional geodemographics built on the category of residency, the Foursquare check-in records facilitate a taxonomy categorisation at the level of the venue (or POI in Weibo and place in Facebook) to generate customer profiles or abstract their lifestyle patterns (Noulas et al., 2011; McKenzie et al., 2015; Ferreira et al., 2020). Joseph et al. (2012) grouped users into clusters according to the places they checked in on Foursquare via topic modelling. Hasan and Ukkusuri (2015) used the contextual information of users' check-ins to infer their personal interests and cluster users by the visitation patterns and neighbourhood choices. In research on tourist travel behaviour, individual tourist's Foursquare check-in records (placed in a sequenced time series) can be treated as travel itineraries or diaries made up by the venue-category suggested activity and temporal details. For example, Vu et al. (2018a)'s analysis on Foursquare check-in data in Hong Kong confirmed that Asian tourists are more likely to travel by public transportation for shopping and also to visit theme parks whereas Western tourists have much higher preferences for nightlife spots. Hu and Jin (2017) used the Foursquare check-ins in Macau to identify the gambling tourist and examined how their travel behaviour was different from other sightseeing tourists. Meanwhile, tourist activity profiles and itineraries generated by their origin of sources have been used as the important inputs for destinations recommendation prediction (Kesorn et al., 2017; Kotiloglu et al., 2017; Xie et al., 2018). These studies suggest that Foursquare check-in data has the potential to reflect empirical tourist movements between various activities, with an exclusive advantage associated with tourist consumption activities, such as shopping and dining out (Salas-Olmedo et al., 2018; Vu et al., 2018). In this regard, this thesis also makes use of the Foursquare movement datasets, aggregated from users' chronological check-in records in London, to simulate the possible grocery shopping flows of tourists (see Chapter 6).

2. Weibo

Sina Weibo, equivalent to Twitter, is the main LBSN platform in mainland China. It allows users to generate microblogs with check-in at locations. Weibo offers a great opportunity to study Chinese activity and mobility in a similar way to Foursquare. Weibo check-ins have drawn great attention from researchers in China especially for the study of migration and culture ties (Wu et al., 2016), urban agglomeration (Zhen et al., 2017), urban function and structure (Wang, Wang, et al., 2016; Shen and Karimi, 2016) and human activity (Yuan and Wang, 2018). In common with the Foursquare examples presented above, Weibo check-ins have also been utilised in retail analysis in China. For example, Wang et al. (2016) adopted Weibo check-ins to estimate user activity centres and calculate the probability of these users visiting retail agglomerations in Beijing. The results were used to further calibrate a spatial interaction model to delineate the trade areas of these retail agglomerations.

The applications of Weibo check-ins in domestic tourism destination research in China has also taken advantage of Weibo check-ins as the indicator of tourist activity and mobility. Shao et al. (2017) delimited the underlying tourism communities as tourism spatial structure by extracting the main tourism areas from geotagged Weibo data in Huangshan City. Meanwhile, the timestamp and location categorisation of Weibo check-ins have also been used to analyse and compare the activity patterns of foreign and domestic visitors (Maeda et al., 2018), tourists and residents (Shen et al., 2019; Khan et al., 2020) or observed gender differences in check-in behaviour (Rizwan et al., 2018). For example, Xue and Zhang (2020) incorporated the expenditure levels at checked-in restaurants and hotels to show the diverse spatial behavioural and consumption patterns of the short-haul tourists, long-haul tourists and local residents in Suzhou. Zhang et al. (2020) compared the Weibo check-in records from visitors and locals as indicators to uncover the (lesser known) restaurants only popular among locals in Beijing.

Weibo check-in data located out of mainland China forms a specific big spatial dataset which can be used to identify overseas Chinese populations (Liu and Wang, 2015). After distinguishing tourist users from migrants and long-term students, this dataset has valuable potential in understanding overseas Chinese tourist activity and mobility patterns, as highlighted in the analysis presented in Chapter 4. Recent research by Chen et al. (2020) also used Weibo data to track the global mobility of the Chinese visitors who had visited Sydney.

To conclude, location-based check-in data offers a vast number of tourists' check-ins at tourist sites together with the contextual information concerning these locations, which allow researchers to rank venues by popularity, understand place attractiveness by their geographic and mobility features, and characterise destinations or AOIs and profile tourists visiting there. These advantages can support research related to destination marketing and management in many aspects: can create tourist travel dairies, uncover tourist preferences for activities and venues, identify popular or busy travel itineraries, and understand the spatial interaction between urban places from a tourism lens.

2.2.3 Tourism service website data: offer and utilisation

Tourism service websites provide the most popular UGC used in tourism studies. These review data are originating from a series of social media reviews sites like TripAdvisor and Yelp and third-party booking platforms such as Booking.com, Expedia, Airbnb and Google Places, etc., covering nearly all the main tourist activities of travelling, eating and lodging (Schuckert et al., 2015; Filieri, 2015; Jia, 2020). The text reviews from diverse service platforms offer valuable and authentic data sources from the perspective of tourists (Xiang et al., 2017). Hence, research has been conducted to uncover knowledge about tourist travelling experiences and sentiments via a range of textual analysis approaches, which are popular for evaluating sentiments and improving service management to attract more tourists (J. Li et al., 2018).

From a geographical perspective, these major worldwide online tourism services, even though their completeness cannot be fully warranted, have been regarded as valuable geospatial data sources of the most precise and up-to-date overviews of the touristic offers in destinations so far (Batista e Silva et al., 2018; Martí, Serrano-Estrada, et al., 2019). Many of them provide the metadata of accommodation establishments including the location, capacity, and proxies for utilisation. These datasets could offer indicators of tourist night-time population distributions and have been used as an important complement to create temporal dasymetric maps (Batista e Silva et al., 2020). For example, the location and capacity of accommodation establishments from Booking.com and TripAdvisor, in conjunction with tourist statistics from Eurostat, produce a set of high-resolution overnight tourist density grids over the EU on a monthly basis. This novel dataset of tourist population has been used to assess the spatiotemporal pattern of tourism in the EU on both local and regional scales (Batista e Silva et al., 2018).

Meanwhile, while the location detailed information from these tourism service data illustrates the supply side of tourism sectors (including hotels, short-term rentals, restaurants and attractions), the collection of reviews and ratings reflect the utilisation of these facilities and services suggesting the popularity from the demand side. van der Zee et al. (2020) compared the spatial patterns of tourist dining out choices and the restaurant distribution in a series of urban heritage destinations in Belgium by using the restaurant reviews on TripAdvisor. The outcomes have been fed back to the DMO directors to inspect the effectiveness of local tourism policies which aim at spreading tourism from core central areas and thus for planning secondary tourism product offerings. Furthermore, these tourism service supply and usage data provide a chance to examine the relationships between services and the city. Gutiérrez et al. (2017) measured the spatial association between the location of Airbnb, hotels and tourist attractions. Their analysis indicated that Airbnb accommodation intensified the existing pressures on housing (and overcrowding) in the residential areas of Barcelona. The information on Google Places listing has also been used to demonstrate the economic activities on offer in urban destinations, and to help characterise the urban tourism function areas (Adelfio et al., 2020; Martí et al., 2020). The evolution of these venue datasets also offers opportunities to examine the spatial extension of tourism occurring in many urban destinations, which is detailed more in Section 3.4.2.

To sum up, while geotagged social media and location-based check-in data shed light on tourist activities and mobilities at the individual level, a set of diverse tourism service websites offer valuable geolocated datasets of the touristic offer and utilisation at the destinations. These data sources help to depict the landscape of the tourism supply side, as well as the popularity of tourist choices on the demand side. It is clear now that the three main types of LBSN data all have their unique potential to contribute to tourism research from different aspects. Therefore, it is useful to employ multiple LBSN datasets at the same destination. The major advantages of doing this are discussed in the following section.

2.3 Multisource LBSN data incorporation

Given the diverse purposes of the LBSN platforms and the varied characteristics of the datasets they offer (Table 2.2), there is a potential for interlinking different types of LBSN data in tourism research (Stock, 2018).

Firstly, different LBSN data can be fused to increase the breadth of the datasets in the same study area. This could also to some extent alleviate the data biases and representativeness issue when using a single source (see Section 2.4) (Li et al., 2013; Malleson and Birkin, 2014). Location-based check-ins and tourism service website data are clearly associated with key venues but may ignore some less attractive destinations, while geotagged social media data have the advantage to unveil new venues and activity spaces since users are able to post from any place (Zhang et al., 2020). Also, while geotagged social media data and location-based check-ins tend to be more related to tourist activity in the daytime (Longley and Adnan, 2016), accommodation service website data supplies an up-to-date overview of tourist overnight accommodations when official or industry data is inaccessible (Batista e Silva et al., 2020). Population biases are universal issues in social media datasets (Olteanu et al., 2019). First of all, social media services usually have a bias towards urban populations (Hecht and Stephens, 2014). In the US, Twitter users in urban areas are 5.3 times more than rural users, and this figure increases to 24.4 times in the case of Foursquare (Hecht and Stephens, 2014). Second, Twitter and Foursquare show a higher proportion of male users than female (Baeza-Yates, 2020), whereas the reports of Facebook and Weibo indicate they have more female users than male (Baeza-Yates, 2020; Weibo Data Centre, 2021). When it comes to the age of the user, Twitter is believed to over-represent younger adults (Wojcik and Hughes, 2019); 18-29 year olds are the most active users of the geotagged functions in Twitter. For example, the mean age of Twitter users in the UK is 34 (Mellon and Prosser, 2017). But Twitter still has an older user age group than Instagram, which has 65% of the users in 18-34 years old category (Baeza-Yates, 2020). Airbnb millennials make up about 60% of all guests (iPropertymanagement, 2020) (Deane, 2021). In regards to ethnicity, research has found that there are no racial or ethnic differences for Twitter users (Hargittai, 2020), when limited to the geotagged Twitter users, there is however still a population bias towards White, Black, Asian, and Hispanic/Latino groups (Jiang et al., 2019) (Malik et al., 2015). Twitter also has been recognised as biased towards higher income groups and those with higher education levels (Mellon and Prosser, 2017). More details of the population biases of LBSN datasets are reported in Table 2.2.

Furthermore, people use different social media platforms in different ways (Oh and Syn, 2015). For example, Manikonda et al. (2016) found that Twitter users express more negative emotions and are more likely to post around work, news and opinions, whereas more personal daily activities and social pastimes are shared on Instagram. Hausmann et al. (2018) argued that Flickr was more popular among experienced nature enthusiastic tourists who shared high-quality professional photographs while Instagram users appeared to be less experienced and younger tourists typically sharing more personal activity. Hence, the combination of multiple sources of LBSN data brings more perspectives to investigate human activities. In the research of

Tenkanen et al. (2017), the authors analysed social media data collected from Instagram, Twitter and Flickr to assess how they can be used to estimate visitation patterns in national parks. They testified that Instagram outperforms Twitter and Flickr in monitoring the number of visitors in natural areas, but the combination of the three platforms yields the most robust matches with the official visitor statistics. Therefore, the authors advocated an ‘all data revolution’ to consider all different data sources available for different areas of the world to inform global conservation research.

Secondly, the cross-referencing of different LBSN data enable the delineation of human activities in the ways that a single data source cannot capture, and therefore inform a more robust interpretation of the research topic (Hamstead et al., 2018; Martí et al., 2020). For example, the comparison of Airbnb listing with TripAdvisor hotel offers in Martí, García-Mayor, et al. (2019) clearly illustrates the decentralisation of tourism to non-touristic zones in Alicante. When overlapping Airbnb data with Foursquare, it also demonstrates that some of the Airbnb lodging clustering areas lack other human activity opportunities, suggesting future potentials to attract extra economic activity. Salas-Olmedo et al. (2018) employed three LBSN data sources to identify spatial patterns of urban tourist activities in Madrid, in which the data from a photo sharing service (Panoramio in this instance) helped to understand sightseeing activities, Foursquare illustrated consumption-related activities and supplementary Twitter data helped to identify the location of tourist’s accommodation. All three sources of LBSN data provided useful new information on the spatiotemporal distribution of tourist activities in one destination, but differences still showed. As shown in Figure 2.7, Twitter showed the most dispersed tourist activity, highest in the historic centre and accommodation areas. In contrast, the geo-photos occurred more in the historic centre and main sightseeing spots than simply the city centre. Foursquare showed more tourists at the luxury shopping area near the historic centre and other shopping centres or individual stores. The authors also revealed the similarity between each pair of LBSN data. They found that Foursquare and Twitter had a certain similarity between the spatial patterns of tourists, whereas geo-photos had the lowest similarity with Foursquare check-ins and a slightly higher similarity with geotweets. Their research demonstrates that the similarities and differences among the three kinds of data sources can be viewed as complementarity when examining the multifunctional tourist areas and specialised areas in cities. Similarly, Martí et al. (2020) provided an example of harnessing the unique offer from different LBSN data to identify the tourist activity zones of specialised functions on the basis of users' sharing contents.

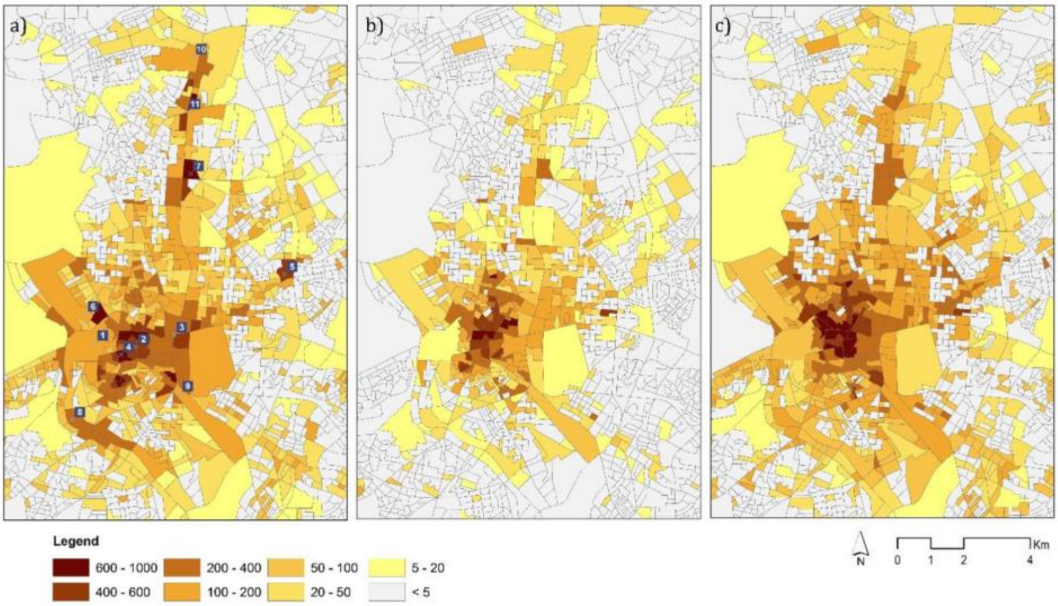


Figure 2.7 Tourist density according to: a) Panoramio, b) Foursquare, c) Twitter. Source: Salas-Olmedo et al. (2018).

In Section 2.2, the thesis reviewed the applications of each of the three types of LBSN datasets in tourist travel behaviour research. This section discusses the necessity and the potential to combine various LBSN datasets in the same study area. Different social media platforms have unique strengths in terms of the data they offer. Nonetheless, the methods and techniques developed for LBSN data analysis are usually transferable. The extracted geospatial big datasets are mostly structured to include geolocation, timestamps, textual contents and other measurable metadata. In light of these, Martí, García-Mayor, et al. (2019) suggest that the interlinking of different LBSN data has facilitated "a transversal view of urban dynamics" to uncover unexpected nuances. Table 2.2 presents a critical summary of the merits, implications and pitfalls of the three different types of LBSN datasets in tourism research, particularly the topics about tourist travel behaviour. In conclusion, the combination of data from different LBSN sources not only increases the breadth of the data, but also enriches the support of analysis and allows a better interpretation of tourism phenomena. This thesis also benefits from the application of multiple sources of LBSN data, as elaborated in Chapter 5.

Table 2.2 Summary of the three types of LBSN datasets in tourism studies.

Dataset	**Geotagged social media data**	**Location-based check-ins**	**Tourism service websites**
Typical services **(<u>used in the thesis</u>)**	• <u>Twitter</u> • Flickr • Instagram	• <u>Foursquare</u> • Facebook Places • <u>Weibo</u>	• <u>TripAdvisor</u> • <u>Airbnb</u> • Booking • Expedia
Benefit	Tourist spatiotemporal densities and dynamics	Tourist activities at venues	Destination tourism offer and utilisation
Implications for tourist travel behaviour	• Correlation with observed visitor numbers • Proxies of tourist distribution density • Identify tourism hotspots • Measure AOI popularity and insight of tourist experience and perception • Individual tourist trajectories • Aggregated tourist flows, movement patterns • Comparison between different origin of countries or with locals	• Tourist activities, especially consumption-related • Travel itineraries and tourist diaries • Characterise destinations by tourist choices • multipurpose trips at destination • Tourist preferences by source market • Tourism spatial structure • POI popularity and the change over time • Indicators of place attractiveness and retail flows • Weibo: global mobility of Chinese visitor	• Tourist experience and sentiment • Most complete, precise and up-to-date overview of tourism supply side in destinations • Utilisation, popularity and experience from the perspective of the demand side • Delineate tourism function areas in cities • Examine the relationship between tourism services and the city, including tourism expansion • Accommodation datasets: tourist night-time population distribution
shortcomings	• Lack of contextual information to infer tourist activity	• Only at popular venues	• Venue level rather than tourist level

Population bias	• Over-represent male, urban population, with higher income and education level • Geotagged Twitter users: 18-29 year olds are the most active users; bias at White, Black, Asian, and Hispanic/Latino groups • Instagram: young adults (18-35)	• Foursquare: Highly over-represent urban population, more male than female, 47% of the user are age 25-34, 29% are 35-44 years old • Weibo: 54.6% female, 80% under-30s	• Airbnb: 54% female; 36% of the users are 25 to 34 years old; 23% are 35 to 44 years old •
Skewness	Skewed by prolific user, at urban area	Skewed at urban area and popular venue; Skewed by prolific user	Skewed at popular venue
Accuracy	Low	High	High

2.4 Discussion

The examples given above show what has been possible so far in terms of applications of LBSN data in tourism research. This section discusses some of the pros and cons of using LBSN data. A major advantage is the fact that LBSN data provides a huge volume and velocity of data. In contrast, most available surveys are limited in their scope and nature. Data on tourist arrivals and demographic information from official government statistics are also limited because they usually offer few opportunities to examine tourist behaviour at an individual level. Individual data enables analysis which is far more spatially and temporally disaggregate than official survey-derived sources (with LBSN data uniquely time and geo-location ‘stamped’) enabling the construction of detailed tourist trajectories at different times of the day and between different groups of tourists. LBSN data has thus been used as a cost-effective but reliable alternative to surveys to support tourist spatial behaviours at a range of geographical scales and to generate knowledge for the destination tourism stakeholders.

LBSN data is not also purely based on a limited number of cases, unlike many surveys, thus avoiding some of the issues around researcher selection bias. Also, LBSN users create the contents and locations in a voluntary way.

This is an advantage to reduce the possibility of behaviours altering when the survey subjects realise that they are under observation (McCarney et al., 2007). However, the bias nature of social media usage still brings a major issue regarding the application of LSBN data – how well does the data represent the entire population? Although there is a high volume of data, the proportion of active users is usually a small percentage compared to the engaged population under research (Steiger, Westerholt, et al., 2015). Many studies address the representation issue that social media platforms present according to people's interests and background (Oh and Syn, 2015; Manikonda et al., 2016). In America, L. Li et al. (2013) found that well-educated users with a good salary were high contributors on Twitter and Flickr. Pew Research Centre identified younger, suburban and Hispanic segments of the population were more likely to use LBSN services (Zickuhr, 2013). In the tourism academic community, Munar and Jacobsen (2014) pointed out that a large proportion of users passively search for information on tourism service websites but never contribute. Therefore, the LBSN users of one platform may comprise only part of the population with a specific interest, although this issue could partly be mitigated by bringing in different LBSNs for the same study area as aforementioned in Section 2.3. Meanwhile, LBSN data can be skewed in the spatial and temporal distribution, especially in study areas with a low intensity of LBSN data coverage (Lovelace et al., 2016). A small number of users may generate the majority of content, and spambots or organisation users can heavily skew the raw data sets. Therefore, many studies make use of the density of users rather than their messages to mitigate the inherent bias and suspicious active users are also filtered out in many studies before analysis.

Another major debating point around these big data is the lack of demographic characteristics in many LBSN data sources, as only a small fraction of the users report attribute information in their user profiles. In other words, little is known about the characteristics of the persons tweeting or posting photos, although some researchers have attempted to estimate these characteristics based on forenames and surnames (i.e. Longley et al., 2015; Luo et al., 2016; Lansley and Longley, 2016a). Here perhaps smaller scale surveys do offer an advantage, especially purpose-built questionnaires which can include more material on individual attributes. If we could break down LBSN data by ethnicity, age or other socioeconomic status, then we would have more idea on how activity and mobility patterns vary by different demographic cohorts (Huang and Wong, 2016; Luo et al., 2016; Chua et al., 2016). Moreover, many studies use the most frequent night-time message location, or the most intensive grid cell of individual user's rasterised geolocated records, to identify the origin country, activity centre or even the home location of users (Malleson and Birkin, 2014; Zhong et al., 2015; Longley and Adnan, 2016; Huang and Wong, 2016; Stock, 2018). Sometimes text analysis also has acted as a support role to identify domestic activity (Birkin et al., 2013; Malleson and Birkin, 2014; Steiger et al., 2015). This so-called location inferences process enables urban geographers to link socioeconomic status (or the geodemographic classification in the UK case) according to the livelihood of a person (Longley and Adnan, 2016; Huang and Wong, 2016; Luo et al., 2016). In tourism studies, such identification of user's origin even at a coarse resolution help scholars to distinguish tourists from different source markets (Girardin, Calabrese, et al., 2008; Su et al., 2016; Vu, Li, et al., 2019). However, a problem related to the location inference of more serious concern is the ethical issue, mainly regarding privacy. Although the LBSN platforms feature robust security mechanisms for user data safety, the high accuracy data analytics and automatic data mining techniques potentially lead to the risks of privacy breaches (Zook et al., 2015; Yao et al., 2019). The research of Vu, Law, et al. (2019) offers an example of the possibility of LBSN data to breach traveller privacy. They revealed that the sensitive social relationship among tourists and locals could be mined via constructing joint Foursquare check-in records according to the coincided spatiotemporal traces of different users. Therefore researchers have advocated the privacy-preserving data analytic when applying LBSN data and to be cautious as to what extent may the research impinge on the LBSN users (Sui and Goodchild, 2011; Shoval and Ahas, 2016; J. Li et al., 2018; Yao et al., 2019). Such ethical considerations are also taken into account in this thesis during the analysis of LBSN data in Chapter 4, 5, and 6.A third concern when using LBSN data is validation – in other words, do the data at least give us similar aggregate insights or patterns as survey data (and therefore gives more confidence in the use of the data), and in what ways can they provide greater insights? A few studies have employed a validation step in their research to verify whether the LBSN data can be used to help understand specific human activities. Steiger et al. (2015) validated the distribution of work-related geotweet clusters against the workplace population census data and argued that the geotweets are a good proxy of workplace-based activities. But the same experiment on home-clustered tweets failed to fully represent residential populations. In relation to tourism, Comito et al. (2016) compared identified popular travel routes from geotweets in London against the top-recommended travel itineraries of the tourism service websites and most of the identified routes achieved a very high level of accuracy. Kádár (2014) verified that Flickr is a more precise indicator of visitor attendance at a singular attraction than the statistical data solely based on ticket selling. Hawelka et al. (2014) and Barchiesi et al. (2015) argued geolocated Twitter and Flickr data are good indicators in global mobility behaviour studies by validating their results against the official statistics of international tourism.

These examples support the argument that the fine granularity of LBSN data can be a supplementary dataset to inform tourist activity and movement patterns. But there is also some negative evidence exhibited. In the same study of Barchiesi et al. (2015) the Flickr-based model failed to capture the annual difference in visitor numbers to the UK from an individual country. Also, the research of Girardin, Calabrese, et al. (2008) and Lovelace et al. (2016) showed mobile roaming data, although not always easy to access, are better proxies for estimating the visitor numbers and travel flows than LBSN data due to the intensive data coverage. They suggested that LBSN data is more promising to reveal new dimensions of tourist travel behaviour rather than reinforce existing tourism knowledge. More importantly, Lovelace et al. (2016) further stressed that veracity is even more important than analysis, and hence the critical verification and systematic evaluation of the data sets and corresponding applications should not be overlooked. However, in reality the lack of fine-scale datasets regarding human activity and movement usually limits this task.

There have been long criticisms of Big Data analytics in empirical social science research. The main criticisms are usually around the bias and noisy nature of raw data and that analytics only finding answers to predefined questions rather than developing new understandings. The application of LBSN data in human geography research also faces similar challenges. However, increasing applications bringing in LBSN data to assist the investigation of a wide spectrum of research questions can be seen in the literature. This is because many research outcomes have supported the view that ‘the trends can be revealed clearly regardless of outliers’, albeit a large proportion of the work is around data cleaning and pre-processing (Mayer-Schönberger and Cukier, 2013). Although these findings cannot lead to general laws or reveal causality, LBSN data analytics allow deeper focus on what is happening at specific places and times (Miller and Goodchild, 2015). There are also some major debates around whether Big Data means an end to theory and the likely replacement, or redundancy, of small data

(Kitchin, 2017). In line with these discourses, the investigation of the LBSN data quality and its potential impacts on the analytical results has become an indispensable part of LBSN data applications, which is also included as an important component in the discussion of Chapter 7 (see Section 7.3).

2.5 Conclusions

This chapter has presented a comprehensive summary of the state-of-the-art of location-based social media services for exploring the activity and mobility patterns of urban tourists. In sum, LBSN data can be considered as a valuable data source to generate knowledge for planning and operational purposes in tourism destination management. The information extracted from LBSN data could shed light on how tourist providers are meeting the demand of different tourism markets. Although, as seen above, LBSN data faces criticism of skew and bias, it provides a much larger volume of readily accessible data about the digital footprints of individual tourists. These ever-growing novel datasets can be used to understand the activity spaces and dynamics of those tourists, especially in those domains where no better data sources are available, such as visitor space usage, tourist movement patterns, tourist interaction study and tourism impacts estimation. Text analysis uncovers the underlying semantic information within LBSN data, which can be used as a crucial complement to place contextual information on the spatial and temporal footprints of users.

The following chapter will closely examine some ongoing changes within patterns of tourism in many large cities and look at the potential that spatial interaction modelling might contribute to future research in tourism movements, especially in relation to grocery shopping.

Chapter 3 Linking urban tourism with retail location modelling

3.1 Introduction

Many approaches have been leveraged in retail location analysis for turnover forecasting and store performance benchmarking, from simple methods such as the analogy approach or regression analysis to more complex methods including radiation models and discrete choice models. Since the first gravity-type Huff model (Huff, 1963), spatial interaction models (SIMs) have been widely utilised in retail applications and significantly developed through collaborations between academia and industry. They have been shown to be very accurate with many applications in the grocery retailing sector in particular (Birkin et al., 2017; Birkin and Clarke, 2019). To adapt to the increasingly complicated retail demand side, recent attempts to disaggregate SIMs have advanced the models to increase the diversity of the grocery demand side: from persons making trips from home to the inclusion of commuters and workplace demand, coastal holiday tourists, school pupils and students (Newing, 2013; Waddington, 2017; Birkin, 2019).

Meanwhile, as shown in Chapter 2, the growth of tourism mobility in cities and towns has accentuated the intricacies of the urban population landscape. The variegated tourist population has penetrate areas which extend beyond traditional tourism areas into residential neighbourhoods, blurring the mobilities, behaviour and geographical boundaries between tourism and daily life (Maitland, 2013; Novy, 2018). Taking the under-examined tourist grocery shopping demand as a unique angle, this chapter demonstrates why and how tourists may heighten local grocery demand in urban areas and illustrates the necessity and challenges of adding urban tourist demand into the SIM for retail location analysis.

3.2 The extension of spatial interaction models

SIMs have been used as an essential tool for analysing, explaining and simulating flows (of people, information, capital, traffic, trade etc.) across space (Batty, 2008; Birkin et al., 2010; Fotheringham, 2017). The early Huff model and other SIMs were given greater analytical rigour through entropy-maximising techniques introduced by Wilson (1971, 1974). The entropy-maximising SIMs, derived from statistical mechanics, have a sounder theoretical basis and thus became widely accepted by human geographers and regional scientists and became one of the most successful geographical models used in applied research (Birkin et al., 2017). From the 1980s, SIMs started to act as a core technique to support location-based decision making in retail location planning, particularly in the grocery sector (Wood and Reynolds, 2011, 2012). Equation 3.1 is the most basic form of the SIM, which is built up of three components relating to supply, demand and interaction (Birkin et al., 2017). The inherent concept of the model is that the expenditure O_i from the demand side within any given small area i is shared by competing retailers j based on their relative 'attractiveness' W_j and accessibility. The accessibility is a function of the relative 'cost' in terms of distance d_{ij}, calibrated using a distance decay parameter β which reflects the propensity of consumer to travel to interact with supply.

$$S_{ij} = A_i O_i W_j exp^{(-\beta d_{ij})} \tag{3.1}$$

Where, S_{ij} is the predicted flow of expenditure (or customer) from residential area i to retail store j, O_i represents the amount of expenditure available in area i, W_j measures of attractiveness of retail store j and d_{ij} account for the cost incurred (i.e. distance or travel time) between i and j, β is the distance decay parameter, and A_i is an internal balancing factor written as

$$A_i = \frac{1}{\sum_j W_j \, exp^{(-\beta d_{ij})}} \tag{3.2}$$

to ensure that all demand is allocated to retail outlets within the region:

$$O_i = \sum_j S_{ij} \tag{3.3}$$

SIMs are typically used to estimate customer expenditure flows between local residential neighbourhoods and the retail supply side, which in turn allows the estimation of store revenue and local market penetration rates and delineation of store catchment areas. Thus, many retailers have found them especially useful as a basis for their store location strategic decision making (Birkin et al., 2017; Clarke, 2020). Moreover, they have also been adapted for other uses within retail geography. For instance, Clarke et al. (2002) used the predicted flows from their SIM to create an interaction-based accessibility index to evaluate the coverage of food provision in residential zones and thus identified potential food deserts in major urban areas. Moreover, SIMs have been extended to adapt to more complex contemporary business purposes, particularly for retail patronage applications. For example, Davies et al. (2019) and Beckers et al. (2021) developed their SIM to estimate the patronage of grocery Click & Collect services in the UK. Dolega et al. (2016) extended the SIM from traditional grocery store applications to delineate the catchment areas of retail centres (or agglomerations). Sevtsuk and Kalvo (2018) modified the demand side of the model from residential zones to individual buildings, delimiting the supply side as commercial clusters to produce a disaggregate estimation of retail agglomeration patronage. De Beule et al. (2014) extended the SIM by incorporating more influencing factors into the store choice probability analysis of the same brand, in order to improve the predictive power of the model in retail network performance.

Apart from the SIMs which model aggregate customer flows, there are also other disaggregated models to simulate mobile consumer travel behaviour and capture the temporal changes in accessibility on the supply side. Agent-based models have been used in location analysis modelling because of their ability to capture store choice behaviour at the individual customer level. Sturley et al (2019) provide a good illustration of progress relating to ABMs used in retail location modelling. Although they were not able to model every individual separately they used loyalty cards to identify different types of shoppers and then used the ABMs to decide on most likely retail store choice for each group. Birkin and Heppenstall (2011) provided an example of a hybrid agent-based model by integrating agent-based modelling in SIMs to simulate equilibrium-seeking behaviour in dynamic retail environments. Kowalski (2019) built a hybrid agent-based model to simulate trips to local swimming pools. Compared to the SIM approach, the agent-based model can better reflect the space-time variance of accessibility and interactions between the agents, although it often requires much higher computational capacity. ABMs also require some methodology for choosing different locations, which may involve the same trade off assumed in SIMs – accessibility v store attractiveness. Other disaggregated models such as utility theory and discrete choice models have also been used in location choice modelling to simulate complex consumer choice behaviour (Piovani et al., 2016). Moreover, the geometric patterns of urban morphology have also been considered in location analysis and space syntax analysis has been used in the measure of location accessibility (Morales et al., 2019). Also, spatial structure has been incorporate in SIM to reflect its impact on the distance-decay parameter estimates. Oshan (2020) reviewed four methods that are able to account for the spatial structure of the origins and destinations in SIMs using space syntax analysis of the local urban morphology. These technical enhancements have shown that it is possible to capture more complex consumer behaviour and therefore improve the accuracy of SIMs. But, such developments are usually computational intensive and require large samples of individual-level consumer data associated with specific retail activities. This research leverage the large amounts of individual-level data from LBSN, however, it is still insufficient sample data related to grocery shopping to support the more disaggregated geocomputation models as exemplified above. Therefore, the model used in this research follows the typical form of classic production-constrained entropy maximising SIM and the model calibration relies on both the individual level travel data from LBSN data and also reliable travel survey sources. It is also worth emphasising that retail organisations continue to prefer SIMs to alternative methodologies and hence it is important to continually improve methods in wide usage.

To meet the increasingly complex retail scene, the classic production-constrained entropy maximising SIM has been disaggregated over time. It can be expressed as:

$$S_{ij}^{kn} = A_i^k O_i^k W_j^{\alpha^{kn}} exp^{(-\beta^k d_{ij})} \quad (3.4)$$

where the flow S_{ij}^{kn} represents the predicted expenditure flow between origin demand zone i and store j disaggregated by brand n and household type k. The demand side of the model is disaggregated here to measure the available expenditure in demand zone i by household type k, while the W_j term in the supply side is disaggregated by a brand preference parameter α^{kn} to reflect the additional relative attractiveness of store j of brand n for household type k. The distance decay parameter β is also disaggregated here as β^k to represent the willingness or capacity to travel of household type k, and the internal balancing factor A_i is accordingly changed into A_i^k and written as:

$$A_i^k = \frac{1}{\sum_j W_j^{\alpha^{kn}} exp^{(-\beta^k d_{ij})}} \quad (3.5)$$

Thompson et al. (2012) and Newing et al. (2015) have built customised SIMs by embedding the disaggregated α^{kn} using relative attractiveness measurements generated from industry survey data. The model outputs have been validated against empirical market share and revenue data, which showed the effectiveness of SIMs in grocery retailing practices when disaggregated by brand and person type.

The distance deterrence term expresses the attenuating effect of distance, reflecting the propensity of consumers to travel (usually) from residential locations to specific stores. It enables the predicted consumer flows to replicate any known consumer trips (Birkin et al., 2010). The distance decay parameter β is often disaggregated by consumer type, to account for the propensity of certain consumer groups to travel further to shop at the store of choice. Thus accessibility (or ease of travel) can be varied for different household types, which may be restricted in movement by car availability or general cost of transport. Similarly, the retail trips of the professional daytime work population from workplace to the stores of choice usually shows more stronger spatial constraints than is usual for trips made from home (caused by limited time available at lunchtime or minimal opportunities to shop around work-based locations), and thus it is important to have higher β when modelling workplace demand. In the context of urban tourism, tourist shopping flows may originate from either the temporary accommodation locations or the attractions they visit in the daytime. The distance decay variable can also vary for different segments of overnight tourists. For example, the tourist staying with friends or relatives is more likely to generate expenditure flows in accordance with the host family, whereas the self-catered travellers of non-serviced accommodation may only shop in proximate areas. Therefore, as explored fully in Chapter 6, β is separately calibrated for local residents, workers and different tourist segments.

The outputs S_{ij}^{kn} of the model predict the expenditure flow from each fixed demand zone to each of the stores in the region. The zones that contribute the customer inflows of the store are considered within the catchment area of the store. One of the main research focuses has been on disaggregating the demand side variable to include non-residential demand. Models which rely solely on residential demand have been reported to underestimate the sales in many areas with complex population compositions or fail to reflect the temporal fluctuations of the stores therein (Birkin et al., 2017). For example, Newing et al. (2013a) showed a pronounced seasonal trade uplift patterns in major Cornish coastal resorts in the UK driven by tourists originating from outside the local catchment areas. Berry et al. (2016) and Waddington et al. (2017) also identified the sales fluctuation throughout the day, especially at stores close to workplaces and transport interchanges. Waddington et al. (2019) reported a considerable underestimation of revenues for convenience stores whose catchment areas contained workplaces or universities and schools. These studies suggested that as any region consists of intricate population subgroups, a simple estimation of solely residential demand is insufficient. Under such circumstances, scholars have attempted to extend the models on the demand side, to incorporate varied types of non-residential demand.

A highly disaggregated SIM was developed by Waddington et al. (2017) to incorporate demand from residents, workers, schools and universities. The new model was validated against the empirical store turnover of a selection of partner retailer stores. Their research showed that the extended SIM could effectively capture the daily fluctuations of each nuanced demand group in an urban context. Similarly, Newing et al. (2013) incorporated coastal tourist demand into SIMs, using a major tourist destination in the UK as a case study. Again using partner data from a major leading UK grocery firm, they estimated the seasonal variations within the tourist population and integrated tourist demand to the original residential SIM, enabling the model to forecast fluctuations in grocery store sales on a monthly basis. The research highlighted the surge of grocery demand in peak seasons and the distinctive store revenue uplifts from this. The methodological details of the non-residential demand estimation in these two research papers will be further discussed in Section 3.4.1. Despite the success of these works, little research has built on this to explore grocery demand in urban destinations that regularly attract a large influx of tourists. For the urban destinations hosting significant tourist visits, there is potentially an interdependence of the grocery demand and supply side. Whilst the seasonal peaks and troughs may not be so pronounced, the overall volume of tourists (and thus their expenditures) may actually be greater. The significant role of tourism within urban economies has been well articulated, but the estimation of additional demand driven by tourism has yet to be considered in retail location analysis.

3.3 Urban tourism as a driver of local grocery demand

3.3.1 Tourist shopping activity in urban destinations

Tourist shopping behaviour has seen increasing attention from both retail and tourism academic communities (Timothy, 2005; 2014). The literature has pointed out the importance of shopping activities for tourist experiences and the impacts of tourism on the destination economy and employment by diversifying the structure of the local economy, creating employment opportunities and generating foreign exchange (UNWTO, 2014).

The current literature regards tourist shopping as a recreational activity in which tourists browse, select and purchase goods to take home during their travel (Choi et al., 2016). A majority of these studies focus on the motivation, experience and perception of tourist shoppers, and a plethora of studies have investigated the determinant factors of tourist expenditure at destinations (Wang and Davidson, 2010; Sainaghi, 2012). However, grocery shopping has rarely been considered in broader research. This may be because traditionally tourists spend their nights in serviced accommodation like hotels and B&Bs, which usually have no cooking facilitates and thus guests are more likely to dine out rather than shop extensively at grocery stores. However, changing tourist accommodation patterns have reshaped the tourism landscape in many destinations and expanded tourist demand to include shopping for groceries at nearby convenience stores or supermarkets.

3.3.2 Changing tourism demand trends and their impacts on residential neighbourhoods

Larger cities and towns which function as major tourist destinations traditionally aim to attract visitors by creating planned tourist zones. These areas are known as ‘tourism precincts’ and usually contain a series of typical and often predictable attractions (alongside a high concentration of leisure and shopping facilities and accommodation services for tourists) and tend to take on certain spatial, social, cultural and economic identities (Law, 2002; Edwards et al., 2008; Hallyar et al., 2008).

However, increasing evidence shows that urban visitors may seek accommodation and activities beyond traditional tourist areas, even extending into suburban areas (Maitland, 2007; Novy and Huning, 2009; Maitland, 2019). Location-based technology such as navigation and geolocated tourism apps have also facilitated tourist specialised information about places, helping them access and explore new areas (Jansson, 2019). The local tourism policies in some major cities have started to encourage the spread of tourists outside the city centre (Smith, 2019a) to mitigate the crowds and other consequences of ‘overtourism’ on the one hand and offer tourists more diverse and distinctive destination experiences on the other hand (Maitland and Newman, 2009; Guttentag, 2015; Pappalepore et al., 2014).

Growing academic discourses have paid attention to the latent drivers of such tourism expansion and explored how the geographically re-distributed tourist demand impacts urban space (Novy, 2018). Airbnb, among other online rental platforms, is the best-known driver of this transformation, which has also gained the most attention from the academic community. This may be because the listing and occupancy data are readily available from AirDNA and other publicly accessible websites like Inside Airbnb (insideairbnb.com). The sharing economy accommodation such as Airbnb draws on houses and apartments from the existing housing stock, complementing the tourist accommodation provision within city centres. This may instigate ‘unplanned expansion’ of pre-designed tourism bubbles to residential neighbourhoods (Inkson, 2019; Ioannides et al., 2019). The expansion certainly eases pressure on the populated tourism areas in city centres and satisfies the desire of tourists to experience cities more ‘like a local’ (Maitland, 2013), but can also ignite new tensions between visitors and local residents (Novy and Colomb, 2019).

Recent studies have emphasised the significance of the economic, social and socio-cultural impacts of Airbnb on the host communities (Novy and Colomb, 2019). Airbnb has been criticised for driving negative effects on the housing market, hotel industry and the social relationships in neighbourhoods. The more lucrative tourist short-term rentals stimulate more development capital in the residential markets of destinations and thereby impose housing pressure on local inhabitants and students (Lee, 2016; Wachsmuth and Weisler, 2018). For example, the growth of short-term rentals is believed to affect the residential housing system and urban governance in London, since the weekly rates of short-term rentals are three times higher than long-term rentals according to the Valuation Office (Inkson, 2019; Shabrina, 2020). Although the Deregulation Act 2015 in London restricts to a 90-day limit the entire property short-term rental within a calendar year, Ferreri and Sanyal (2018) argued that the potentially illegal lettings circumvent urban and tourism planning and limits the effectiveness of urban governance.

Meanwhile, the economic loss of the hotel industry due to the disruptive impact of Airbnb has also been estimated by Zervas et al. (2017) and Dogru et al. (2019). They demonstrated the detriment of Airbnb supply on all the key hotel performance metrics. Another crucial critique of short-term rentals like Airbnb is that they accelerate tourism gentrification and displacement processes in the residential areas they occupy (Cocola-gant, 2018). Tourists may not only intensify the usage of urban facilities and services in the vicinity of their accommodation but also tend to transform mundane neighbourhoods into tourism spaces and induce supporting services such as restaurants, nightclubs and shops to become more oriented towards tourism catering (Ashworth and Page, 2011; Ioannides et al., 2019). Although these threats have been recognised, Airbnb and other similar accommodation services have developed into a mass-market from initially being a niche sector. Some global hotel suppliers and many mainstream online travel agencies have started to incorporate sharing economy accommodation into their offers (Inkson, 2019). In the case of London, the London Plan 2021 highlights the role of short-term rentals in providing more choices for visitors and expanding accommodation provision outside the central areas. As 72% of Airbnb listings are located in the outer boroughs of London, the trend of tourists moving into 'back regions' of the city is very likely to continue.

Apart from the self-rental accommodation trends, other shifting tourist trends have also been identified in the literature to advance the spatial expansion of tourism in urban destinations. The 'experience economy' creates new event-scapes to host day visitor crowds, resulting in the temporary incursion of visitors into public venues (Brown et al., 2015). The work of Smith (2019), for example, showed how hiring out park spaces as commercial event venues have transformed public parks from simple urban amenities into tourist attractions as a part of the visitor economy. Freytag and Bauder (2018) argued that the emerging touristification on-going in Paris is mainly driven by the growth of sharing economy platforms including house- and bike-sharing, enabling tourists to access non-central areas cheaply and easily. In the study of Buning and Lulla (2020), GPS tracking data of the Pacers bikeshare service suggested that visitors heavily used the bikeshare system in Indianapolis, accounting for a third of the users and a fifth of the trips. In contrast to the locals, who use the bikes as a mobility supplement to connect with public transportation, visitors used the bikeshare service for recreational urban exploration in a more relaxed and wider way, which allow them better access to peripheral areas and to extend their travel time (Hardy, 2003; Dickinson et al., 2011; Buning and Lulla, 2020). Additionally, tourism invested urban second home ownership has also been noted to affect local communities in a wide range of social, economic and political aspects in some cities of Spain, Portugal and Sweden (Hall, 2014; Cocola-gant, 2018; Back, 2020).

These studies have confirmed the expansion of tourists into suburban residential districts and highlighted the potential conflicts with local dwellers, although whether the latter have primacy in the claim to the city remains a matter of debate (Ashworth, 2009; Novy, 2018; Novy and Colomb, 2019). At the same time, growing evidence has also recognised that middle- and upper- class urban dwellers increasingly use urban spaces and resources in ways 'as if tourists' when exploring neighbourhoods outside of their daily life communities (Novy, 2018). While research has started to investigate the social and economic impacts of these new tourism trends, the quantification of such impacts has not been scrutinised on a city-wide basis, particularly at the small-area level. The spatiotemporal patterns of tourist population at fine granularity are essential to understand how tourist demand overlaps and intertwines within the local neighbourhoods. The thesis develops this research area but with specific reference to tourist impacts on the local grocery retailing market.

3.3.3 The under-researched tourist grocery demand in urban destinations

Even though major urban areas may attract considerable tourists for much of the year, traditionally these visitors are not regarded as significant drivers of local grocery retailing demand, except within the core tourism districts, and therefore they have not been comprehensively considered in retail location modelling. However, the aforementioned new trends have dispersed tourists beyond the conventional tourism districts to mix with locals even in peripheral suburban areas. In these areas, customer demand is generated by variegated population subgroups and spatial modelling of a disaggregated demand side has been proven to increase the predictive power of retail models (Waddington et al., 2019). Newing (2013) for the first time shows the considerable seasonal uplifts of grocery store sales driven by the holiday tourists in the study area of a coastal tourism region. The temporal (by season or by month) fluctuation of tourism inflow may not be that noticeable in many urban destinations, however, for the major urban destinations that host a large number of tourists and visitors, retail location analysis should also consider the retail demand uplifts caused by tourist concentrations.

Urban tourist consumers for grocery shopping can be divided into overnight tourists and day trip visitors in compliance with their needs. According to prior literature, overnight tourists may spend a large proportion of their time in the immediate vicinity of their accommodation sites (Shoval et al., 2011), and therefore the location of tourist accommodation is a critical influential factor of their consumption patterns in urban destinations. Airbnb (2018) claimed that 43% of their international guests' expenditure in the UK is spent in the neighbourhoods where they stay, among which an average figure of around £10 is typically spent on groceries for self-catering. It is also expected that guests staying with friends and relatives will increase grocery shopping demand in local neighbourhoods. Even tourists staying at hotels and other serviced accommodation may have occasional usage of the nearby grocery stores and supermarkets, even though that may be limited.

On the other hand, key attractions, public green spaces and sports venues also host a mass of leisure daytime visits and temporary events. These day trip visitors are also reported to purchase food and snacks at the grocery stores close to the localities and attractions that they are visiting (GBDVS, 2018).

The potential grocery expenditure from individual urban tourists may still be much less compared to expenditure by local residents, given the short length of stay of the tourists and the amount of money they might spend. However, in many major urban destinations, the volume of tourist visits can be substantial. Meanwhile, local convenience retail planning usually considers grocery demand only from the residential population, thus overlooking the emerging tourist demand and resulting in areas with insufficient grocery provision as discussed in Chapters 5 and 6. Therefore, this thesis argues that the examination of tourist demand in urban destinations is important. Such an investigation of tourist demand patterns may help to understand what challenges and opportunities the local retailing market might face after the influx of tourists and visitors and provide extra knowledge for destination management operations related to tourist food shopping. The following section will discuss two major challenges when adding urban tourist demand into retail location modelling: demand estimation and distance decay calibration.

3.4 Challenges of adding urban tourist demand into retail location modelling

3.4.1 Demand estimation: spatial modelling of tourist population distributions

Retail demand estimation is usually a bottom-up process, measured by the population in a geographical zone alongside the expenditure rates on different products. Complete and robust demand estimation is the primary foundation for spatial modelling of retail analysis. As noted above, demand estimation conventionally is orientated around local habitants, benefiting from the robust and well-developed census for small-area residential populations. More recently, the tailored subgroup population surveys have also helped demand estimation modelling by adding other regular population information, for example commuters at workplaces, students in schools and universities and visitors at holiday resorts and major attraction sites (Newing et al., 2013b; Waddington et al., 2017). Thus, to incorporate tourist demand into retail models the first task is to model tourist population distributions at the small area level.

Newing et al. (2013a) is the first relevant study to demonstrate the notable out-of-catchment demand - seasonal visitor demand in their case study of Cornwall, UK. They identified the proportion of external trading in a series of example stores located in the coastal tourism area in Cornwall. They demonstrated that store level sales, and the proportion of trade drawn from non-residents, varies from a peak in August (school holiday period) to a low in January due to the influx of tourist from outside the region. Newing et al. (2013) stated that the tourist accommodation stock is a key driver of visitor numbers and also determines their spatial distribution within a tourist area such as Cornwall.

To identify the spatiotemporal distribution of visitor demand in their study area, Newing et al. (2013b) captured the provision of four types of self-catering accommodation – tourist campsites, holiday centres and villages including sites with static caravans, rented cottage/apartment and second home owners. They also identified the corresponding occupancy rates for each type of accommodation (as published by regional or local destination marketing organisations) to estimate the small-area visitor population in coastal resort areas.

Meanwhile, by comparing customers' regular and in-trip transaction records from the loyalty card data from a major grocery retailer in the UK, Newing et al. (2014) argued that visitor grocery expenditure is complex and varies by store, destination and type of customer. Therefore, different surveys were referenced or conducted to estimate the grocery expenditure of the visitors utilising these four types of self-catering accommodation in the study area (Newing et al., 2013b). Consequently, Newing et al. (2013b) showed that it is possible to estimate the small-area visitor monthly expenditure on groceries based on the provision of self-catering accommodation, inferred occupancy rates and associated expenditure estimates.

The methodology of generating small-area demand estimates not only contributes to quantifying the seasonal and spatial variation of visitor expenditure, but also offers important inputs to spatial modelling within retail location analysis (Newing, 2013). Waddington et al. (2017) distinguished the daytime population from the census-based night-time residential population, revealing that daytime grocery demand fluctuates temporally due to varied consumer behaviours generated at workplace, school, university and leisure visitor attractions. Specific education and workplace population statistics, along with annual tourism surveys, were used to estimate these daytime population distributions and build them into a SIM as separate disaggregated layers, based on the spatiotemporal distribution of each of these population groups. Both Newing and Waddington's research benefited from unprecedented access to empirical store sales data and loyalty card scheme data from their collaborating grocery retailers to make their analyses possible and contribute in a way that no other research has done.

In broader research, the literature has recognised the significance of spatiotemporal tourist distributions as a component of the transient daytime population in urban areas (Bhaduri, 2008; Smith and Fairburn, 2008; Batista e Silva et al., 2020). In major urban destinations, tourists and visitors mix with local dwellers, commuters and students to interweave a complex mosaic of population demand. Nevertheless, the fine-scaled population distribution of tourists is still a thorny problem. Unlike other subgroups who have regular surveys and statistics to understand their geographical patterns, even at the small-scale level, there is little spatial detail on tourist numbers across a city or region (Martin et al., 2010). This inevitably hinders the understanding of local tourist demand and its impact on the neighbourhoods and localities of urban destinations.

To capture the spatiotemporal fluctuations of the tourist population, researchers have attempted to leverage diverse novel big data sets as seen through some of the examples reviewed in Section 2.2.3. The capacity and occupancy of tourism accommodation establishments have been used as reliable and effective geolocated data sources to estimate overnight tourist numbers. For example, Batista e Silva et al. (2018) produced monthly tourist density grids across the EU by integrating Eurostat data with geolocated tourist accommodation datasets retrieved from Booking.com and TripAdvisor. In contrast, Sánchez-Galiano et al. (2017), whose study area hosted a large proportion of unregistered day trip visitors, sought to use utility consumption data sets (such as the utility of water, wastes and electricity) along with housing occupancy to estimate the temporal fluctuation of the tourist population at the intra-municipal level. In a similar fashion, Spalding et al. (2017) attempted to quantify the complete picture of tourist variability in global coral reef areas by combining day visitor numbers generated through the Flickr geo-photos in conjunction with the overnight tourist numbers staying in global coral reef jurisdictions via a commercial accommodation database.

Increasingly, studies have started to employ a data fusion approach in spatiotemporal population estimation to compile high-resolution population surfaces and demographic data in the region (Panczak et al., 2020). Official statistics (census and surveys) can be integrated with geospatial data from both conventional (mapping agencies like Ordnance Survey) and emerging data sources acting as proxies of urban dynamics (i.e. social media, mobile phone, footfall and sensor) to generate multi-layered dasymetric maps (Batista e Silva et al., 2020). The analysis in later chapters of this thesis takes account of both overnight tourists and day visitors, but is limited by the suitable official data sources which could be used to estimate small-area tourist population and their associated grocery expenditure. Therefore, the later chapters adopt a data fusion method by collating multisource LBSN data sources with conventional datasets to map and model the diurnal and nocturnal tourist population distribution patterns based on tourist overnight stay and location of activities.

3.4.2 Distance decay calibration: spatial patterns of tourist travel behaviour

Another primary challenge of adding urban tourist demand into retail location modelling is the calibration of the distance decay parameter. As addressed in Section 3.2, the SIM built for the location analysis of grocery retailing aims to reproduce the customer travel behaviour between origin locations and the destinations – the grocery outlets. The origin locations usually are the home address for the residential consumers, workplace buildings for the commuters, and universities and schools for the students. Additionally, in the context of urban tourism, they are the temporary accommodation establishments for the overnight tourists and attraction venues for the day visitors. However, tourist travel behaviours are commonly known as diverse, individualised and non-habitual, hence difficult to survey in the same way as other population groups.

Recent studies have benefited from a variety of big geospatial datasets in spatial model calibration, especially mobility research where there is a lack of bespoke movement surveys (Birkin and Clarke, 2019). Trajectory data traced by tracking technology like GPS has been used in the calibration of SIMs. Siła-Nowicka and Fotheringham (2019) utilised a sample of GPS data to calibrate production-constrained SIMs for individual home-based shopping trips. The high-resolution temporal and spatial GPS-derived movement data enables the comparison of resident shopping behaviour on weekdays and weekends and under different weather conditions. GPS-based daily intra-urban logistics big data has also been used to fit models of urban freight movements and to understand the distance decay of different types of goods (Zhao et al., 2020). However, GPS tracking technologies used for tourist travel at the city scale are rare and have more commonly been seen in the research of tourist movements within a single attraction (Huang and Wu, 2012; Md Khairi et al., 2019). This could partly be due to the high cost of both devices and volunteer recruitment to support sizeable samples for this data-intensive modelling process in a large region, and is also limited by the accessibility of other sorts of device data related to tourists such as mobile roaming data (J. Li et al., 2018).

In light of this, free and ready accessible social media data generated by tourists has become an important alternative of tourist travel data and used for tourist activity and mobility studies as introduced fully in Chapter 2. For the three types of LBSN data reviewed in Chapter 2, the geotagged social media data such as geotweet is mainly people-based, which have been broadly used to understand human mobilities, whereas location-based check-in data and tourism service website data are place-based and thus are apt to simulate the interaction between places or human activities. In relation to tourism-related spatial interactions, Jin and Xu (2018) extracted tourist movements between hotels and attractions from an online travel diary website to calibrate the distance decay coefficients between the hotel and the attractions. By both text and spatial filtering, Lovelace et al. (2014) identified the geotweets of visiting local museum to inform SIM and compared with the baseline model. Liu et al. (2014) used Weibo check-in data to calibrate inter-urban interactions in 370 cities in China. As reviewed in Chapter 2, among all three types of LBSN data location-based check-in data is the most suitable one to investigate trajectories of human activities (to recap, geotagged social media data is deficient to inform human activities in an explicit way while tourism service website data normally fails to offer abundant individual trajectories). Therefore, this research extracts tourist grocery shopping trips from the Foursquare movement data in London to calibrate the various distance decay parameters required. The grocery shopping trips are selected by the origin-destination pairs between accommodation or attractions to the grocery stores. The dataset and method used will be elaborated in Chapter 6 and Appendix C.

3.5 Conclusions

SIMs have acted as essential tools in retail location analysis, particularly for location modelling within grocery retailing. The model has been disaggregated to adapt to the increasing complexity of the demand side, but has not yet included the additional demand from urban tourism. The incorporation of urban tourist population into spatial modelling has become important in the urban areas hosting significant tourist visits.

Urban tourism has witnessed a geographical spread away from urban core central areas. Online rental platforms such as Airbnb, amongst other new tourism trends, re-distribute tourists outside conventional tourist areas and into more suburban areas, producing increasingly intricate microgeographies within urban neighbourhoods. New patterns of grocery shopping demand will inevitably be associated with this spatial expansion of tourism. This chapter has argued that although challenges exist, it is important to add urban tourist demand into retail location models to understand the impact of urban tourism on local neighbourhoods and increase the effectiveness and accuracy of retail location analysis in urban destinations.

The estimation of urban tourist demand and the modelling of the tourist SIM in London with the help of LBSN data sets will be reported in Chapters 5 and 6 respectively. Before that, Chapter 4 will first present the opportunities and challenges of harnessing LBSN data in exploring tourist activities, dynamics, and consumption-related behaviours in the research area of this thesis – London, by the detailed investigation of an under-researched (outside China) but typical LBSN data source - Weibo for the unique source market – China.

Chapter 4 Understanding Chinese tourist mobility and consumption-related behaviours in London using Sina Weibo check-ins (Paper I)

Abstract: In this paper we detail an individual-level analysis of under-exploited location-based social network (LBSN) data extracted from Sina Weibo, a comprehensive source for data-driven research focussed on Chinese populations. The richness of the Sina Weibo data, coupled with high-quality venue and attraction information from Foursquare, enables us to track Chinese tourists visiting London and understand behaviours and mobility patterns revealed by their activities and venue-based 'check-ins'. We use these check-ins to derive a series of indicators of mobility which reveal aggregate and individual level behaviours associated with Chinese tourists in London, and which act as a tool to segment tourists based on those behaviours. Our data-driven tourist segmentation indicates that different groups of Chinese tourists have distinctive activity preferences and travel patterns. Our primary interest is in tourists' consumption behaviours and we reveal that tourists with similar activity preferences still exhibit individualised behaviours with regards to the nature and location of key consumption activities such as shopping and dining out. We aim to understand more about Chinese tourist shopping behaviours as a secondary activity associated with multi-purpose trips, demonstrating that these data could permit insights into tourist behaviours and mobility patterns which are not well-captured by official tourism statistics, especially at a localised level. This analysis could be up-scaled to incorporate additional LBSN data sources and broader population subgroups in order to support data-driven urban analytics related to tourist mobilities and consumption behaviours.

Keywords: Location-based social networks, Sina Weibo, Chinese tourists, London, Tourist segmentation, Retail behaviour

4.1 Inferring tourist behaviours from LBSN data

Tourism is an important driver of urban mobility within major cities. A micro-level understanding of tourist characteristics, mobility trajectories and consumption-related behaviours are essential for urban planning and urban service analysis (McKercher and Lau, 2008). Headline statistics and survey-derived insights, such as those drawn from the UK International Passenger Survey (IPS), provide aggregate level overviews of inbound tourist magnitudes, attitudes and self-reported behaviours. Headline estimates of tourist numbers and associated expenditures act as a barometer of tourism activity yet disaggregating these across space and by tourist origin or activity/expenditure type is notoriously tricky (Ashworth and Page, 2011; UNWTO, 2014; Song and Li, 2008).

Location-Based Social Networks (LBSN) generate spatiotemporal data which could enable novel insights into these localised tourist behaviours (Vu et al., 2019; Chua et al., 2016; Comito et al., 2016). In a comprehensive review of the literature, Li et al. (2018) note that user-generated data for tourism research have grown rapidly, predominantly drawn from geo-located photos, microblogs or location-based check-ins. These insights can be broadly thought of as first aggregate level indicators of tourist activity preferences captured by ‘hot spots’ of tourism activity at a destination (Salas-Olmedo et al., 2018; Vu et al., 2015); and second as more individualised insights into tourist itineraries and activity patterns, which is the focus of our discussion.

One difficulty in using LBSN data to infer these activity patterns is in classifying social network users into different groups (e.g. based on country of origin or individual demographics), especially when self-reported information (such as ‘place of residence’) in their user profile may be unreliable. We address some of these challenges by drawing on the social networking service Sina Weibo (referred to hereafter as Weibo). Weibo offers an opportunity to identify behaviours and mobility trajectories specifically associated with one important sub-group of UK tourists; Chinese inbound visitors. We use London as a case study city and exploit the under-utilised value in Weibo check-in data. Our analysis of Weibo check-in data also covers the two-step approach to discover the ‘hot spots’ areas and movement patterns of Chinese tourist in London, but in contrast to many previous published studies, we distinguish the similarity as well as the differences of the multipurpose travel behaviour of our Chinese tourists in London. Meanwhile, this tourist group has the highest per capita expenditure when in the UK (VisitBritain, 2018). Therefore, we also attempt to explore their consumption-related activities during their multipurpose trips in London.

4.2 Introducing the Weibo check-in data

Weibo is the most comprehensive LBSN source for data-driven research focussed on Chinese populations with approximately 210 million active users (Weibo, 2019). Recent examples of its application include studies of Chinese population mobility (Liu and Wang, 2015), cultural ties (Wu et al., 2016), urban planning (Zhen et al., 2017) and domestic tourism destination research (Shao et al., 2017). Weibo offers the potential to undertake a comprehensive assessment of activity preferences and travel trajectories associated with inbound Chinese visitors in major destinations which attract tourists of Chinese origin, such as London. London received over 19.83 million international tourist visitors in 2017 (VisitBritain, 2019). Visitors of Chinese origin represented the 8th largest group of international inbound tourists in London by spending in 2018 (ONS, 2019). Headline statistics suggest that Chinese visitors to London have a longer length of stay and higher expenditures than other groups of international inbound tourists (VisitBritain, 2018).

Weibo allows users to generate microblogs which can be associated with specific points of interest (POIs) at which the user 'checks-in' in a similar fashion to geo-located Tweets or Foursquare check-ins. Whilst sources such as Twitter are typically available free of charge for a sample of only 1 – 2% of all tweets, we have access to a near-complete set of Weibo check-ins for a given time period. Our data thus enables a very comprehensive insight into Weibo-derived LBSN check-ins among Chinese tourists. The collecting and pre-processing of the Weibo check-in dataset used in this study is detailed as follows:

(1) Data collection

An iterative program was set up to collect Weibo check-in data via the Application Programming Interface (API) (https://api.weibo.com/2/place/pois/users.json) within Greater London. Queries to the API return the latest 1,500 user check-ins at any specific POI within this defined study area. Providing that the 1,500 user check-in threshold isn't reached, every user and all check ins related to the POIs are returned by the free API. Each retrieved Weibo check-in contains information including user ID, check-in ID, check-in time, POI name, category, location and ID, alongside the user-generated textual message attached to the check-in. There are 2,665 points of interest (POIs) being checked in at by Weibo users during our study period of 1st Jan 2016 to 28th August 2018. Only 31 POIs (0.01%) returned the maximum 1,500 user check-ins and we therefore have a near-complete sample of all user check-ins at almost 100% of the relevant POIs. Data collection began in 2016 to coincide with increased inbound Chinese tourism resulting from new visa regulations (GOV.UK, 2016).

(2) Tourist identification

Each check-in is associated with a uniquely identifiable user ID, enabling individual Weibo users to be tracked across multiple time and location stamped check-ins. Social bots (generated as ‘fake’ accounts used primarily for advertising) are first filtered out from our check-in dataset, identified by intensive check-ins at multiple locations over a very short time span. Tourists are further distinguished from local residents and long-stay non-tourist visitors (such as those studying or working in London). We used check-in frequency and check-in timespan to infer short-stay tourist visitors, applying a length of stay threshold of 20 consecutive days to distinguish tourists from other Weibo users. This threshold is based on surveyed data from the IPS and VisitBritain insight (ONS, 2018; VisitBritain 2018c). Our Weibo tourists have an average length of stay of just over 6 days and reveal a propensity for users’ first and/or last Weibo check-ins to take place at London Heathrow Airport, a major international airport which accounts for over 90% of aircraft seat capacity between China and the UK (VisitBritain, 2018).

(3) POI categorisation

A set of 20,233 geolocated check-ins from 6,465 unique Weibo users were identified as tourism activity, approximately 20% of our raw dataset (the remainder attributed to Weibo users who are resident or on a long stay [e.g. student] visit). To avoid potential mis-categorisation of Weibo POIs and to generate data that are comparable with other LBSN datasets, we associated each Weibo POI with a named venue derived from Foursquare, which provides a high-quality set of categorised venues which we use as a consistent set of points of interest (POIs). This results in a total of 962 venues which are frequently visited (minimum of 10 unique user check-ins) by Weibo tourist users. We acknowledge that Weibo users will not check-in at all POIs visited and that there may be a higher propensity to check-in at major attractions. Nevertheless, these cleaned and processed data present a novel and unique opportunity to identify key mobility behaviours associated with this group of tourists. After data cleaning, the structure and format of an individual Weibo check-in is as shown in Table 4.1.

Table 4.1 Data structure of an individual Weibo check-in after pre-processing and assignment to a specific foursquare venue.

	Data field	Example
Weibo Check-in	Check-in ID	151
	User ID	1006657733
	Check-in time	2018-07-31 10:05
Foursquare venue	Venue ID	4ac518cdf964a520eea520e3
	Venue title	Westminster Abbey
	Venue detailed category	Church
	Venue main category	Professional & Other Places
	Venue subcategory	Spiritual Centre
	Venue popularity	4751
	Latitude/longitude	-0.127356648 / 51.49936
Attraction	Attraction name	Westminster Abbey
Tourism activity	Activity type	Visiting landmarks & buildings

4.3 Extracting insight from our Weibo dataset

Our study employs the following methodology to investigate the activity and mobility patterns of Chinese tourists and to explore their consumption-related activities in London:

4.3.1 Identify the spatial distribution of Chinese tourist activities

Density maps provided an initial overview of the spatial distribution of our LBSN point data set, with each point representing an individual timestamped and geo-located check in. We use Kernel Density Estimation (KDE) to transform the check-in data (by activity type) into a series of smoothed density surfaces, presenting hot spots of tourism activity. KDE is a commonly used methodology to identify the intensity of the spatial distribution of georeferenced point data, with application in assessing human activity distribution from user-generated check-in data (e.g. see Li et al, 2013; Lansley and Longley, 2016). The same density maps representing the venue distribution in London are also generated based on the Foursquare dataset, as a comparison to present the distinctive attraction choices of the tourism activities of Chinese tourist users.

4.3.2 Extract Weibo user's check-in trajectories to understand Chinese tourist mobility behaviour

Network analysis (based on graph theory) has been applied in tourism research to understand the spatial structure of tourist behaviours in relation to the network of tourist attractions and other venues visited (Liu et al., 2017; Lee et al., 2013). In our study, Weibo tourist user's daily check-in trajectories at the individual level have been aggregated to build a core attraction network specifically related to Chinese tourists observed behaviours. We employ eigenvector centrality, a measure of centrality drawn from graph theory which measures the importance of a node (in this case a venue or attraction) based on the number and relative importance of adjacent nodes (Prell, 2012; Bonacich, 1972). The concept has been widely adopted in urban network analysis to describe positions within a given system (Agryzcov et al., 2019) and evidenced in tourism research to understand the spatial hierarchical structure of tourist attractions (Kang et al., 2018), with implications for tourism planning (Asero et al., 2016; Lue et al., 1993). We use the revealed network of core attractions visited by Chinese tourists to infer aggregate level sightseeing and consumption behaviours, whilst also assessing the extent to which individual users deviated from these ‘typical’ behaviours as a tool to help segment Chinese tourists based on their observed behaviours.

4.3.3 Segmentation of Chinese tourists based on their multipurpose travel behaviours

We explore specific attraction choices and mobility patterns exhibited by aggregate level Weibo tourist check-ins, but at the individual level tourist travel motivation and interest preferences are varied. Therefore we segment tourists based on their observed individualised multipurpose travel behaviours using a set of derived indicators capturing a diverse range of indicators derived from our Weibo check ins including trip characteristics (length of stay and number of different attractions visited), activity preferences (relative frequency and diversity of each activity type) and mobility patterns (the dimension, shape and structure) of travel trajectories at the individual level. In total 41 variables are used as detailed in Section 6.

Our segmentation employs K-means, a widely applied clustering algorithm which partitions observations into a set of k groups, where k is pre-specified and represents the number of groups. It is highly efficient for a large volume of data and has been successfully applied in tourism research to segment tourists based on their characteristics (Grinberger et al., 2014; Huang and Wu, 2012). It works in an iterative way to classify objects into multiple clusters so that the intra-cluster variation is minimised, whereas the inter-cluster variation is maximised (Gan et al., 2014). The only prior knowledge of K-means is the specification of k – the number of clusters. In the research, we use NbClust package (Charrad et al., 2014) and the ‘elbow’ method to determine the optimal number of clusters. The NbClust package provides 30 indices for determining the number of clusters and proposes to the user the best clustering scheme from the different results obtained by varying all combinations of number of clusters, distance measures, and clustering methods. The K-means clustering is conducted within the R package ‘cluster’. We evaluate the goodness-of-fit of the clustering result by the average total within-cluster sum of square (WSS) and average total between-cluster sum of square (BSS). The selection of k and the process of k-means clustering is detailed in Section 6.

4.3.4 Understanding Chinese tourist multi-purpose trips and their shopping-related activities

Whilst our segmentation captures key observable travel behaviours of each cluster, the underlying multipurpose trip patterns of Chinese tourists requires further investigation. The topic modelling technique latent Dirichlet allocation (LDA) is employed on a cluster-by-cluster basis to extract more detailed activity patterns exhibited by tourists within each segment, drawing on the within-cluster heterogeneity in terms of venue choice and activity patterns. We benefit from the individual-level richness of the Weibo data, along with the depth of venue category information within our check-in dataset (as shown in Table 4.1). LDA is a generative statistical model within natural language processing used to calculate the probability distributions of topics and associated words in a large collection of documents (Blei et al., 2003). It has been used in wider contexts to infer behavioural and lifestyle characteristics from foursquare check-ins (Vu et al., 2019; Hasan and Ukkusuri, 2015; Qu and Zhang, 2013). In the following sections we present and discuss our insights into these tourists’ destination-level behaviours.

4.4 Spatial distribution of Chinese tourist activities

Weibo-derived check-in data from those users inferred to represent tourists suggests they are predominantly associated with check-ins related to visiting and sightseeing activities (museums, historical sites, art galleries, castles, monuments etc.), with these venues representing 49% of check-ins in our sample. More than 70% of tourist users also checked in at ‘Travel and Transport’ venues, with these contributing almost a quarter of our total check-in activity, highlighting the importance of urban public transport infrastructure in enabling tourist mobility. Figure 4.1 illustrates the spatial distribution of 5 main Chinese tourist activities derived from our Weibo check-in data using KDE. As a comparison, similar KDE maps based only on the POI venue distribution of the respective categories are also constructed as in Figure 4.2. It is clear that hot spots of Chinese tourist activities within these categories are more spatially targeted on specific, localised, bounded and identifiable hot spots.

Although tourism attractions are widely spread over Inner London (Figure 4.2), Figure 4.1 suggests two significant and several secondary hot spots of Chinese tourist ‘Visiting and sightseeing’ activity. The two key hot spots are centred on Westminster (London Eye and Big Ben) and the Leicester Square/Covent Garden areas, both of which are major attractors for tourists. Secondary hotspots are centred on locations which correspond with key tourist attractions such as the British Museum, Hyde Park, Baker Street, Exhibition Road (home to many popular museums) and the Tower of London. In contrast to the concentration of overnight accommodation services evident around Hyde Park in Figure 4.2, our check-in data highlights a propensity for Chinese tourists to use accommodation which is spread across two large spatial clusters, one centred on the Southbank and one around Green Park. In both Figure 4.1 and Figure 4.2, ‘Shopping’ and ‘Dining out’ activities have quite similar spatial patterns. For shopping (Figure 4.1) Chinese tourists focus on Knightsbridge (Harrods) and Oxford Street. There are clear spatial overlaps between different activity types in Figure 4.1, suggesting that tourists combine multiple activities within complex itineraries which include sightseeing/visits to attractions and shopping, alongside dining out, facilitated by the transport network and overnight accommodation.

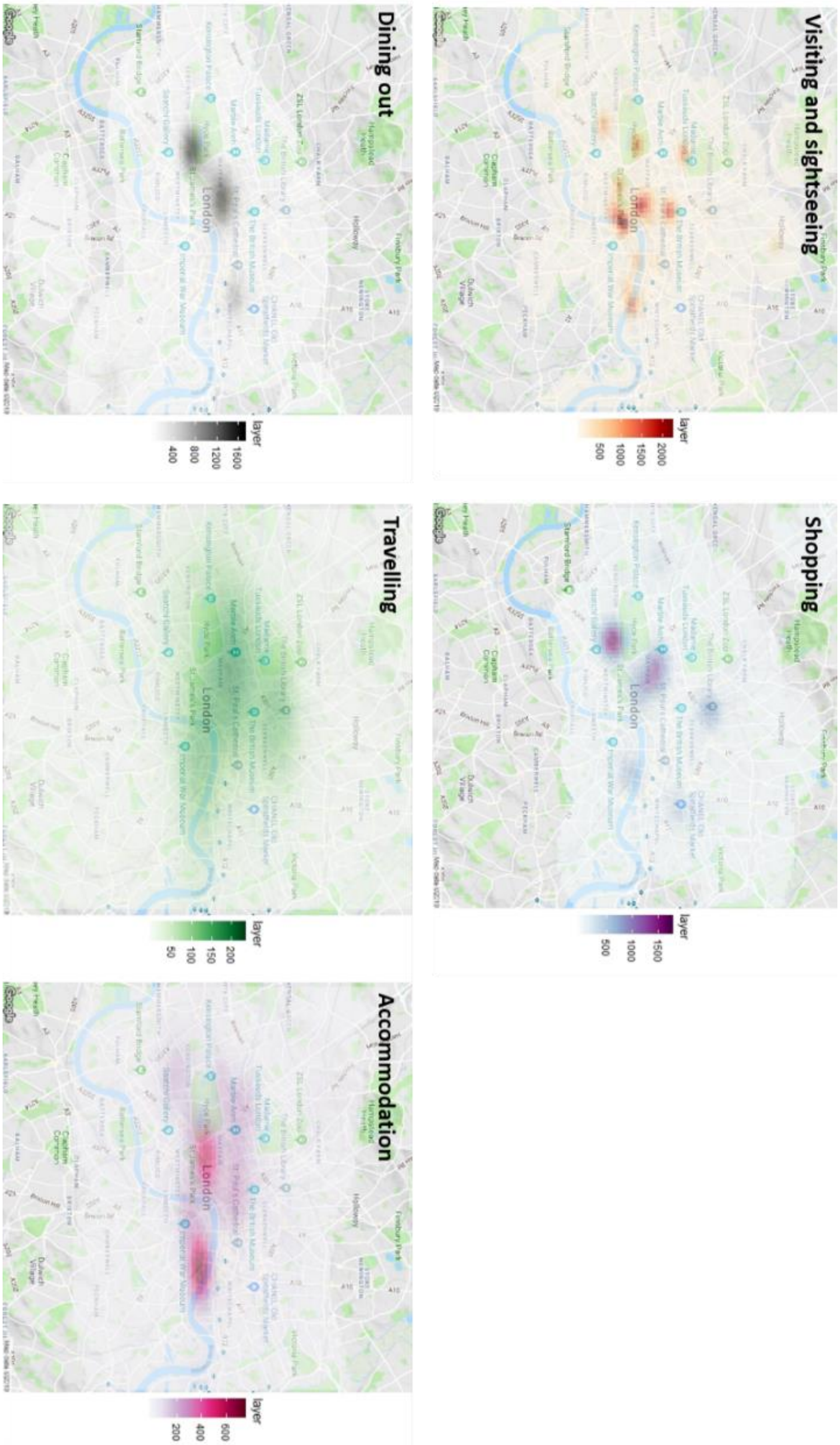


Figure 4.1 Kernel density estimation (KDE) of different tourist activities based on Weibo tourist users' check-in associated with Foursquare venues: (a) visiting and sightseeing, (b) shopping, (c) dining out, (d) travelling and (e) accommodation.

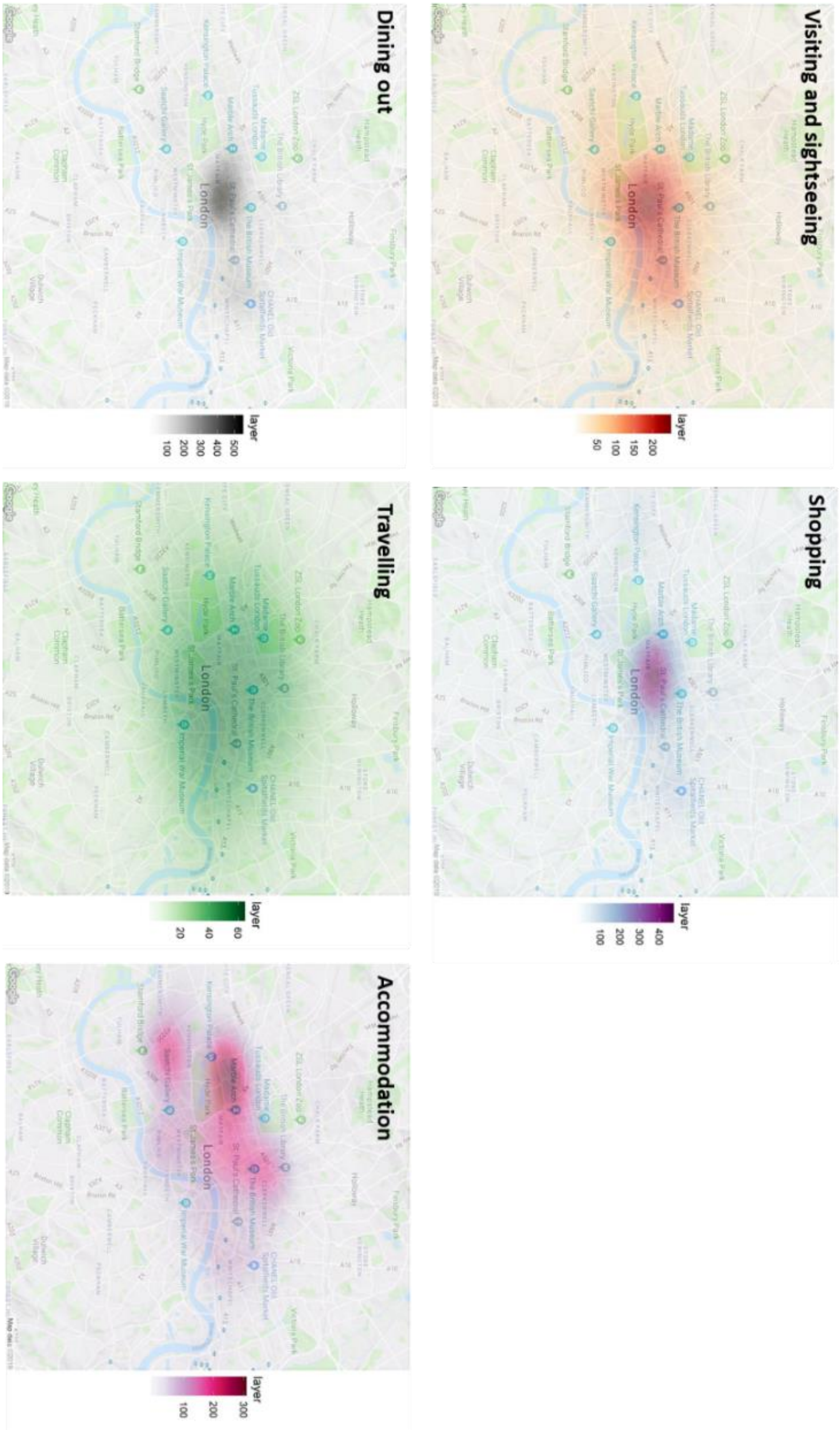


Figure 4.2 Kernel density estimation (KDE) of different tourist activities based on distribution of Foursquare venues by category: (a) visiting and sightseeing, (b) shopping, (c) dining out, (d) travelling and (e) accommodation.

In the following section we consider the trajectories or visit sequences of individual Weibo users during their visits to London.

4.5 Understanding Chinese tourist mobility behaviours

Although the hot spots of tourist activity highlighted in Figure 4.1 enable us to draw interesting observations about the spatial distribution and density of different activity types, the activity and location-based behaviours from which these are built permit a far richer set of insights into tourist mobility patterns. Figure 4.3 illustrates the trajectory for one tourist Weibo user during a trip to London. Check-ins associated with each user have been sorted according to their timestamp and grouped into discrete days in order to extract their daily trajectories. The example illustrated has been chosen to demonstrate the richness of these data in uncovering the mobility and venue preferences associated with individual tourists. Whilst it is acknowledged that users may visit a broader range of attractions and venues than those at which they choose to check-in, their revealed mobility trajectories could provide a valuable and previously under-exploited insight into the characteristics of tourist behaviour in London at the individual level.

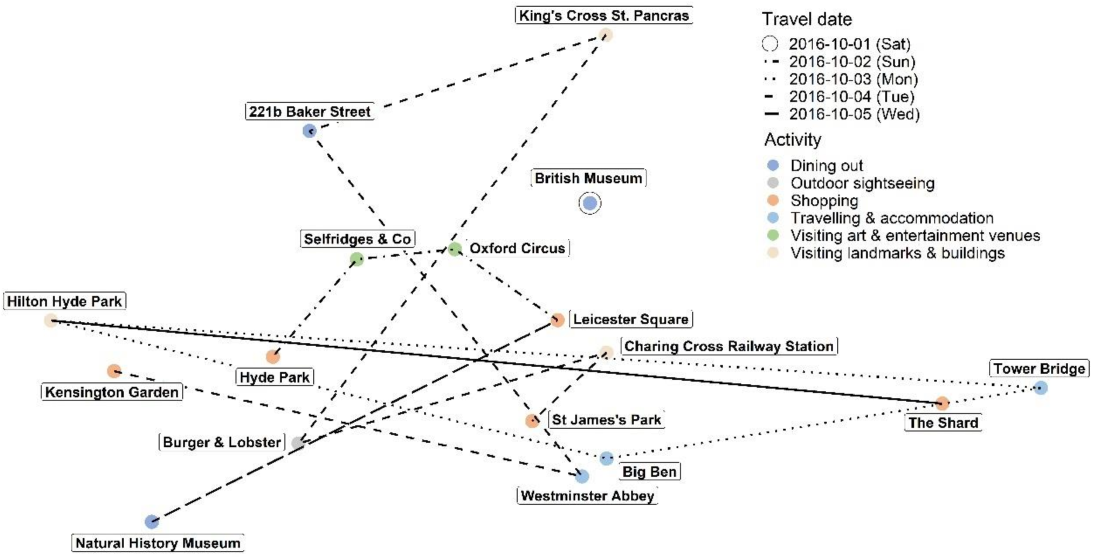


Figure 4.3 Indicative individual trajectory for a specific tourist Weibo user, capturing five separate days' worth of activity during a single visit to London.

By aggregating all individual user-level trajectories (such as those shown in Figure 4.3 and derived for each of our 6,465 individual tourist users) we can understand more about the core attraction networks in London. Our analysis reveals that 22 attractions act as key nodes based on their eigenvector centrality within the network, shown in Figure 4.4, which depicts the eigenvector centrality scores using graduated colours (whilst the size of the node reflects the number of check-ins according to our Weibo dataset). The results reveal that among all attractions in the network, The London Eye and Hyde Park are the most influential and vital attractions for Chinese tourists in London. Figure 4.4 shows an important trade-off between centrality and the number of check-ins, as highlighted by key transport interchanges. London King's Cross Railway Station does not have as many check-ins as London Heathrow Airport, but exhibits much higher importance within this local network, forming an important node in Chinese tourists' daily trip making behaviours whilst visiting London.

The thickness of the 'edges' (lines) connecting nodes is indicative of the volume of tourist flows between these two attractions, with 57 key edges identified. The strong links between central London attractions are clear, as is the importance of specific pairs of attractions such as The London Eye visited with Big Ben, Buckingham Palace and Westminster Abbey, and the British Museum with Trafalgar Square. These findings are comparable with Comito et al. (2016) who use geotagged Twitter data to extract key connections between London attractions. We find that the Chinese Weibo tourist users share the same top connections as the Twitter users but that Chinese tourists show much stronger movements connected to the British Museum, Chinatown and the Sherlock Holmes Museum, which are not so popular in the Twitter data.

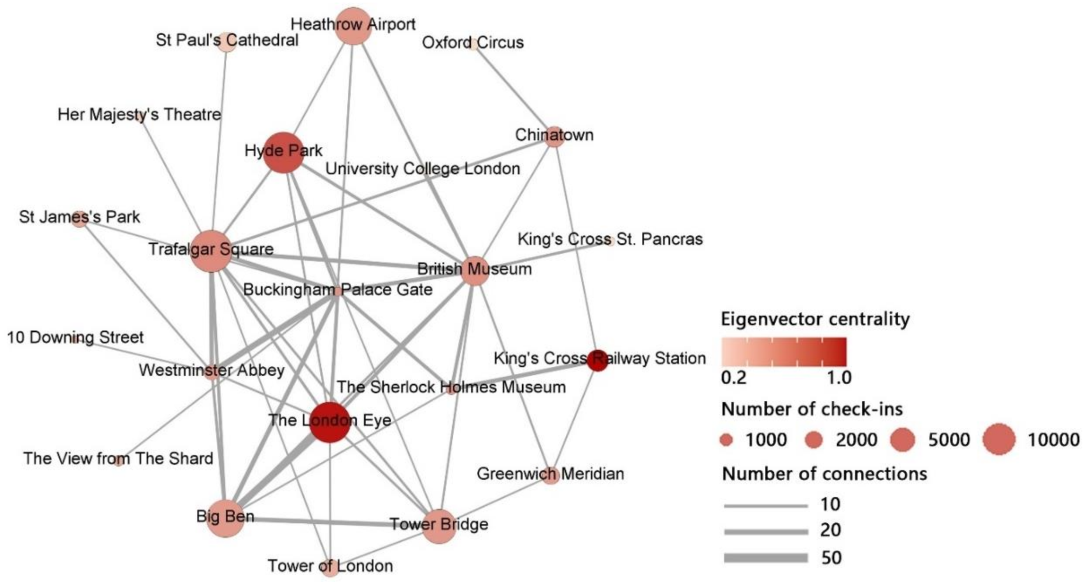


Figure 4.4 Core attraction network based on Chinese tourists' daily Weibo check-in trajectories.

The ability to derive attraction networks from LBSN data, and thus construct aggregate level indicators of centrality and connectivity between attractions can be useful for tourism destination management. The highlighted networks identify the importance of attractions along tourism routes as regard to both the visitation and their core–periphery position. The key origin-destination pairs also could offer benefits for urban planning in tourism destinations, include the evaluation and provision of visitor-oriented public transport, promotion of ancillary tourist services along principal routes utilised by tourists, support to enhance tourist safety and wellbeing along these routes, marketing opportunities or the effective provision of supplementary retail, leisure and catering opportunities.

4.6 Segmentation of Chinese tourists based on travel characteristics, activity preference and mobility patterns

4.6.1 Travel behaviour variables

In this section, we attempt to investigate more on the individualised multipurpose trips of our Chinese Weibo tourists. A set of derived indicators (Table 4.2) are used to capture trip-related behaviours of each tourist user utilising their individual trajectories (such as the example shown in Figure 4.3).

Table 4.2 Outline of 41 variables capturing tourist Weibo users travel characteristics, activity preference and mobility patterns for use in segmentation.

	Domain	Variable	Description
1	Travel characteristics	Length of stay in London	Number of days between first and last check-in
2		Number of trips	Number of days having check-ins
3		Number of stops	Number of check-ins
4		Number of different attractions	Number of visited distinct attractions
5-11	Activity preferences	Activity frequency (calculated separately for each of 7 different activities)	Proportion of check-ins associated with activity of interest (e.g. dining out) relative to all activities
12-18		Daily main purpose frequency (calculated separately for each of 7 different activities)	Proportion of days that given activity (e.g. Visiting landmarks/buildings) represents the main activity
19-25		Venue diversity (calculated separately for each	Count of unique venues visited for each activity type

		of 7 different activities)	
26-32		Popularity (calculated separately for each of 7 different activities)	Sum of the popularity of visited places of each activity type
33		Multi degree	Average number of activities
34	Mobility patterns	Mean daily attractions	Mean count of attractions visited per day
35		Return probability	Probability of returning to the same venue
36		Mean distance	Mean travel distance per trip
37		Mean placement	Mean distance between stops
38		SDE size	The area of the standard deviation ellipses for all check-in points
39		SDE shape	The eccentricity of the standard deviation ellipses for all check-in points
40		Total weight	The total weight of a tourist travel route according to the whole attraction networks
41		Total centrality	The total centrality of the attractions along a tourist travel route according to the whole attraction networks

Travel characteristics capture the basic check-in behaviour of Weibo tourists during their stay in London, in common with approaches used to identify indicators such as length of stay from LBSN data (Preis et al. 2019, Chua et al., 2016). The activity preference indicators cover four dimensions of activity across 7 key activity types and reflect the frequency of visit and relative importance of different activity types. The mobility pattern indicators capture the dimension, shape and structure of each tourists' individual mobility trajectory. These include measures of the probability of a user returning to a previously visited attraction, their mean daily travel distance, and the mean distance between attractions visited. We measure the size and shape of their spatial 'footprint' using Standard Deviation Ellipses (SDE), a centrographic measure to summarize the distributional trend for a set of point locations by reference to the centre of the ellipse and the lengths of the two orthogonal axes (e.g. Huang and Wong, 2016). We also assess the extent to which an individual user exhibits mobility patterns which are consistent with the typical behaviours within our dataset by considering the relative total weight and total centrality when compared to the aggregated typical attraction networks shown in Figure 4.4.

In total, we have 41 derived variables of individual tourist Weibo user behaviours, which we use in the following section to segment tourists based on those observed behaviours. Since these variables are derived from individual Weibo users' observed behaviours and trajectories, we benefit from a rich and multi-dimensional dataset from which to classify our Chinese tourists into distinct groups or segments based on their observed behaviours.

4.6.2 Data clustering

K-means is used to segment tourist Weibo users into distinct groups by the 41 variables outlined in Table 4.2, as described in section 4.3. Before clustering, all the input data has been standardised by z-scores. The optimum number of clusters was determined via R package 'NbClust'. Among the clustering solution from 2-10, the result of NbClust suggest five as the best number of clusters. When k=5 the ratio of total between sum of squares to the total sum of squares is 78.3%, the best goodness-of-fit. The results are shown in Table 4.3. Our largest cluster 'traditional tourists' represents almost half of our tourist Weibo users with key attractions, landmarks and tourist amenities (such as accommodation) featuring prominently within their check-ins (Figure 4.5). Tourists in the 'traditional tourists' and the similar 'outdoor sightseeing' segments exhibit fairly homogeneous travel behaviours, linking together key attractions within our network utilising common routes.

Table 4.3 Chinese tourists' trip-related behaviours and activity preferences by segment.

Cluster	Prop.	Dominant activity	Trip patterns	Shopping venue choices
Traditional tourists	48.4%	Visiting and sightseeing	Low complexity, homogenous patterns – check-ins predominantly at museums, parks, landmarks hotels and transport hubs	Rare shopping, occasionally at gift and souvenir shops
Shopping enthusiasts	20.5%	Shopping	Short mean travel distance – combine	High diversity choices of popular shopping

			shopping with visits to museums, performance venues and outdoor sightseeing attractions	venues: department stores, souvenir shops, markets, and major shopping centres
Gourmets	11.6%	Dining out	Travel within the core attraction network with a focus on dining out	Shopping not the main trip purpose but department stores, markets, souvenir shops, and clothing stores feature prominently
Education	9.8%	University-related	Less focus on core attraction network, travel mainly related to university venues	Some evidence of shopping activity across all shopping venue types.
Outdoor sightseeing	9.6%	Outdoor sightseeing	Diverse outdoor sightseeing activity, high travel distance per day.	Few shopping trips but greater diversity of shopping venues when incorporated: department store, market, souvenir shop

We are also interested in the remaining clusters which capture smaller groups of Chinese visitors who exhibit a preference for a broader range of activity and venue types, including shopping or dining out. These groups exhibit greater heterogeneity between tourists, with more dispersed spatial footprints and show less reliance on key nodes and common routes. Figure 4.5 illustrates key differences between ‘traditional tourists’ and ‘shopping enthusiasts’ with the latter (which comprise around one-fifth of our tourist Weibo users), exhibiting a higher propensity for their trips to be dominated by regular, prolonged and repeat visits to key shopping venues (major department stores and principal retail centres within Greater London). Whilst these ‘shopping enthusiasts’ may typically exhibit a shorter average travel distance and less likelihood for return visits to individual attractions than ‘traditional tourists’, their longer length of stay and the greater diversity of the individualised trajectories have important implications for the planning and management of tourist infrastructure. To illustrate the importance of any one activity group we consider these shopping activity behaviours in more detail in the following section.

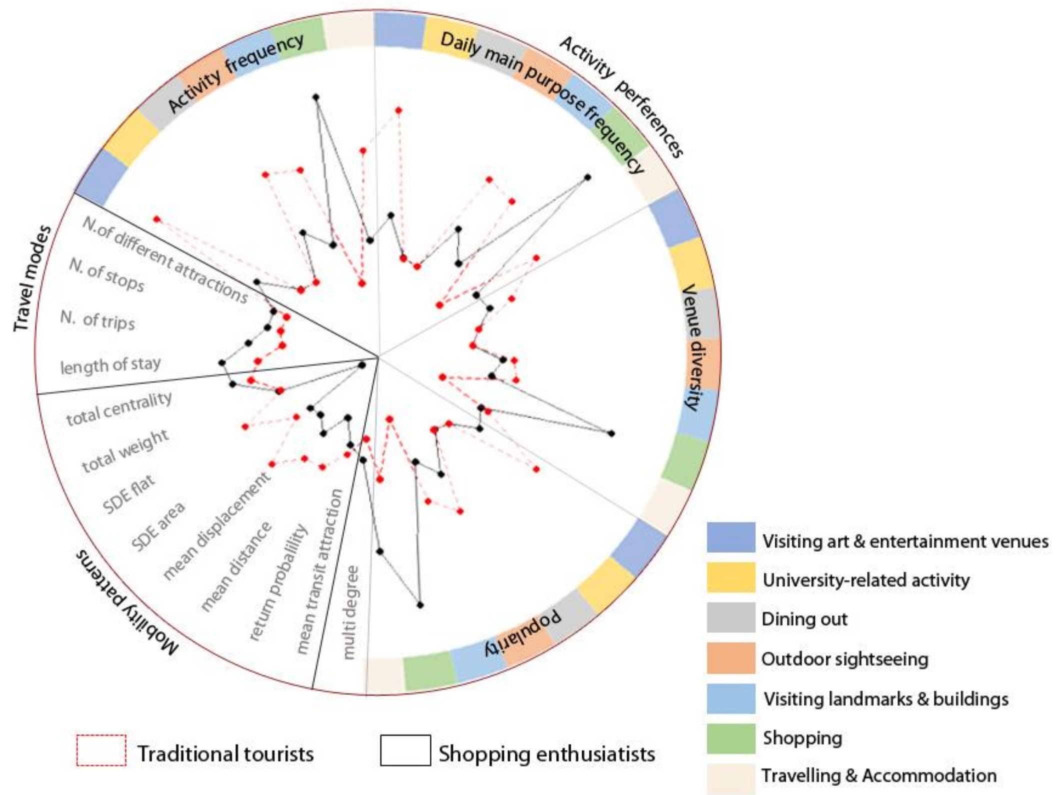


Figure 4.5 Radar charts to illustrate tourist Weibo users trip characteristics and behaviours on 41 key variables for a) 'traditional tourists' and b) 'shopping enthusiasts'.

4.7 Shopping related activities

Shopping related activity is a key driver of Chinese tourist expenditures in London (VisitBritain, 2018b; China Tourism Academy, 2014). Chinese tourists have a greater propensity to undertake high-value shopping as a core activity whilst in London (China Tourism Academy, 2014), spending an average of £2,059 per visit to Britain in 2017, more than three times the average for all other international inbound visitor groups (VisitBritain, 2018). While shopping does not feature as a key driver of behaviour for all segments revealed from our clustering, all segments do exhibit some propensity to visit venues associated with shopping, highlighting the importance of major retail venues as a driver of Chinese tourist intra-destination mobility. Our Foursquare-derived venue level POIs enable us to drill down further and exploit the depth and richness of venue level information in order to understand the shopping activities undertaken as a secondary or tertiary activity by tourists with an alternative primary trip purpose, employing the topic modelling method LDA.

LDA enables us to mine the full richness of the venue information to understand more about activity preferences and how these activity choices connect as multipurpose travel patterns within each segment. Our LDA-driven insights are shown in Table 4.3 as part of the 'multipurpose patterns' and 'shopping venue choice' columns. These reveal that groups such as 'Gourmets' and 'Outdoor sightseeing' exhibit far more heterogeneous activity patterns than other clusters, revealing more individualised preferences at the venue sub-category level in relation to the types of cuisine favoured, the type of stores frequented or the less popular attraction types visited.

It is useful to explore one category in more detail: the shopping behaviours associated with tourists that fall into segments dominated by non-shopping activity types. We can use these to understand the shopping venue choices and the role of shopping for each segment. As the largest segment, 'Traditional Tourists' have a relatively low complexity and homogeneous multipurpose travel patterns, in which shopping activities are rare and only gift shops and souvenir shops are occasionally included. In contrast, shopping enthusiasts visit a more diverse types of shopping venues: department stores, souvenir shops, markets, electronic stores, bookshops, shopping malls, flea markets and gift shops. They may intersperse these activities with visits to museums or other forms of sightseeing and their activity choices are likely to favour more popular venues. The 'Gourmets' also enjoyed varied shopping activities at department stores, markets, souvenir shops and clothing stores but seldom as their daily main purpose.

The LDA-driven enhanced venue-specific insight by cluster highlights the importance of mining the depth and richness of venue or POI information associated with LBSN check-ins, in this case enabling us to reveal venue-level specific insights for each cluster. Specifically, it has enabled the identification of shopping and consumption-related behaviours associated with tourists who have been classified as having different primary drivers of trip-related behaviours.

4.8 Conclusions

Weibo-derived user-generated LBSN data could offer spatial and temporal insights into the behaviours of this sizeable and important sub-group of tourists. We have cleaned and pre-processed these data, adding value by attaching high-quality venue-level information and distinguishing tourist users from other forms of temporary visitors or semi-permanent residents. The spatial characteristics of Chinese tourist check-ins are in keeping with our expectations of a tourist user base, with spatial clusters of Chinese tourists around core London attractions, museums, landmarks and major transport interchanges, in line with key tourist hot spots uncovered by previous research into tourist attraction preferences in London derived from LBSN sources (Comito et al., 2016; García-Palomares et al., 2015). However, we find that key hot spots of Chinese tourist activity are far more discretely clustered and delineated than the underlying distribution of those attractions themselves. Thus, clear spatial clusters of Chinese tourist activity are evident, and these are not entirely driven by the underlying distributions of core attractions, accommodation provision or opportunities for consumption activities. Analysis of this nature affords insights into the micro-level spatial distribution of subgroups of tourists.

Weibo thus presents an opportunity to focus on the behaviours associated with inbound tourists from a single country of origin. Given that Chinese outbound tourism is increasing rapidly, it is unsurprising that China has become the most important global outbound market in terms of expenditures (UNWTO, 2018). Our insights into the spatiotemporal dimensions of these tourists' activities within a major receiving destination such as London are useful in their own right in order to understand the behaviours of this subset of tourists. Moreover, we demonstrate the potential which Sina Weibo affords researchers in uncovering observed behaviours associated with Chinese international tourists including the range and spatial extent of destinations visited and their trip purpose – extending far beyond the scale and volume of data that could be collected via survey mechanisms.

Our findings could also support wider interest in deriving new proxies that can be used to infer tourist flows and behaviours. There is considerable interest in deriving official tourism statistics from LBSN data streams, such as to supplement IPS sample survey estimates of inbound tourism magnitudes. For example, geotweet from twitter and geotagged photo sharing from the Flickr platform has been found to offer great potential in inferring counts of visitors for a variety of inbound markets in the UK (see Steiger et al. 2015, Barchiesi et al. 2015) and other major tourist receiving countries (see Preis et al., 2019). Weibo-derived check-in data could offer similar headline insights along with within-destination mobility and consumption patterns which are not well-captured by official tourism statistics. The availability of a near-complete record of all check-ins (rather than a sample of check-ins as afforded by the free Twitter API) and the clearly identifiable user group (predominantly tourists with a Chinese residential origin) heightens the potential usefulness of these data and warrants further study, especially in relation to their potential to support the production of official population statistics.

Construction of individual tourist mobility trajectories reveals that many Chinese tourist Weibo users restrict their activities to a core network of attractions, with distinct groups of tourists exhibiting behaviours associated with consumption activities (e.g. shopping) or a tendency to explore more peripheral locations and attractions. This level of insight is typically omitted from surveys or official statistics. We offer new perspectives on observed multipurpose tourist activity patterns, derived from those users' digital footprints. This fills a gap in the existing literature and highlights the effectiveness of these data to generate insights into tourist destination choices alongside the value of these data within a data-driven segmentation of tourists based on destination-level behaviours and mobility patterns.

Our segmentation captures these spatial and attribute dimensions of users' activities and indicates that each Chinese tourist segment has distinct multi-purpose travel behaviours and activity venue choices. These findings help to shed further light on Chinese tourist travel mobility and consumption-related behaviours in London. These insights could be utilised to understand Chinese tourist demand and support tourism package design (e.g. see Majid et al. 2013), in administering within-destination surveys (see Abbasi et al. 2015) and destination management. We are particularly interested in these tourist shopping behaviours and our ongoing wider research focusses on extracting those behaviours to support retail demand estimation and store location planning.

Acknowledgement

This study is supported and funded by the University of Leeds and the Chinese Scholarship Council (201709370048).

References

Abbasi A, Rashidi T, Maghrebi, & Waller, S. (2015). Utilising Location Based Social Media in Travel Survey Methods. In Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks (pp. 1–9). New York: ACM Press.

Agryzcov T, Tortosa L, Vicent JF, et al. (2019) A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science 46(4*): 668–689.

Asero V, Gozzo S and Tomaselli V (2016) Building Tourism Networks through Tourist Mobility. *Journal of Travel Research 55(6).* 751–763.

Ashworth G and Page SJ (2011) Urban tourism research: Recent progress and current paradoxes. *Tourism Management 32(1).* 1–15.

Barchiesi D, Moat H, Alis C, Bishop S and Preis T (2015). Quantifying International Travel Flows Using Flickr. *PLoS ONE, 10*(7), e0128470. doi:10.1371/journal.pone.0128470.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research, 3,* 993–1022.

Bonacich P (1972) Factoring and weighting approaches to status scores and clique identification. The *Journal of Mathematical Sociology 2(1).* 113–120.

Charrad M, Ghazzali N, Boiteau V, et al. (2014) NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set Malika. *Journal of Statistical Software 61(6):* 1–36.

China Tourism Academy (2014) Annual report of China outbound tourism development 2014. Beijing.

Chua A, Servillo L, Marcheggiani E, et al. (2016) Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management 57,* 295–310.

Comito C, Falcone D and Talia D (2016) Mining human mobility patterns from social geo-tagged data. *Pervasive and Mobile Computing 33*. 91–107.

Gan G, Ma C and Wu J (2014) *Data Clustering Theory, Algorithms, and Applications*. New York, USA: ASA-SIAM Series on Statistics and Applied Probability.

García-Palomares JC, Gutiérrez J and Mínguez C (2015) Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography* 63: 408-417.

GOV.UK (2016) The Home Office launches new two-year Chinese visa pilot. Available at: https://www.gov.uk/government/news/the-home-office-launches-new-two-year-chinese-visa-pilot.

Grinberger AY, Shoval N and McKercher B (2014) Typologies of tourists' time-space consumption: a new approach using GPS data and GIS tools. *Tourism Geographies 16(1*). 105–123.

Hasan S and Ukkusuri S V. (2015) Location contexts of user check-ins to model urban geo life-style patterns. PLoS ONE 10(5). DOI: 10.1371/journal.pone.0124819.

Huang X-T and Wu B-H (2012) Intra-attraction Tourist Spatial-Temporal Behaviour Patterns. *Tourism Geographies 14(4)*: 625–645.

Huang Q and Wong DWS (2016) Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science 30(9*). 1873–1898.

Kang S, Lee G, Kim J, et al. (2018) Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: An application of anchor-point theory*. Journal of Destination Marketing and Management 9.* 358–370.

Lansley G, & Longley, P (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems, 58*, 85–96.

Lee SH, Choi JY, Yoo SH, et al. (2013) Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis. *Tourism Management 34*. 14–24.

Li L, Goodchild MF and Xu B (2013) Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science* 40(2): 61-77.

Li J, Xu L, Tang L, et al. (2018) Big data in tourism research: A literature review. *Tourism Management 68.* 301–323.

Liu B, Huang S (Sam) and Fu H (2017) An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management 58.* 132–141.

Liu X and Wang J (2015) The geography of Weibo. *Environment and Planning A 47(6):* 1231–1234.

Lue CC, Crompton JL and Fesenmaier DR (1993) Conceptualization of multi-destination pleasure trips. *Annals of Tourism Research 20(2).* 289–301.

McKercher B and Lau G (2008) Movement patterns of tourists within a destination. *Tourism Geographies 10(3)*: 355–374.

Majid A, Chen L, Chen G, Mirza H, Hussain I & Woodward J (2013). A contextaware personalized travel recommendation system based on geotagged social media data mining. *International Journal of Geographical Information Science, 27(4),* 662– 684.

ONS (2018) Travelpac. Newport: Office for National Statistics.

ONS (2019) Number of International Visitors to London. Available at: https://data.london.gov.uk/dataset/number-international-visitors-london.

Prell, C. (2012). *Social network analysis: History, theory and methodology.* Los Angeles, CA: Sage.

Preis T, Botta F and Moat HS (2019) Sensing global tourism numbers with millions of publicly shared online photographs. *Environment and Planning A:* 1–7.

Qu Y and Zhang J (2013) Trade area analysis using User Generated Mobile Location Data. In: Proceedings of the 22nd International Conference on World Wide Web, 2013, pp. 1053–1063.

Salas-Olmedo MH, Moya-Gómez B, García-Palomares JC, et al. (2018) Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management 66*: 13–25.

Shao H, Zhang Y and Li W (2017) Extraction and analysis of city's tourism districts based on social media data. *Computers, Environment and Urban Systems 65.* 66–78.

Song H and Li G (2008) Tourism demand modelling and forecasting-A review of recent research. *Tourism Management 29(2):* 203–220.

Steiger E, Westerholt R, Resch B, & Zipf A (2015). Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems, 54*, 255–265.

UNWTO (2014) Global Report on Shopping Tourism. 8: 1–66. DOI: 10.1007/s13398-014-0173-7.2.

UNWTO (2018) UNWTO Tourism Highlights 2018 Edition. Madrid. DOI: 10.18111/9789284419876.

VisitBritain (2018) China - Inbound Tourism Market Snapshot.

VisitBritain (2019) 2018 snapshot. Available at: https://www.visitbritain.org/2018-snapshot.

Vu HQ, Li G, Law R, et al. (2015) Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management 46.* 222–232.

Vu HQ, Li G and Law R (2019) Discovering implicit activity preferences in travel itineraries by topic modeling. *Tourism Management 75*. 435–446

Weibo (2019) Investor Relations. Available at: https://ir.weibo.com.

Wu W, Wang J and Dai T (2016) The Geography of Cultural Ties and Human Mobility: Big Data in Urban Contexts. *Annals of the American Association of Geographers* 0(0): 1–19.

Ye Z, Clarke G and Newing A (2020) A review of the potential of Social Media Data for tourism research. Working paper, School of Geography, University of Leeds.

Zhen F, Cao Y, Qin X, et al. (2017) Delineation of an urban agglomeration boundary based on Sina Weibo microblog ‘check-in’ data: A case study of the Yangtze River Delta. *Cities 60*. 180–191.

Chapter 5 Estimating small-area demands of urban tourist for groceries: The case of Greater London (Paper II)

Abstract Tourist retail demand within urban areas brings both opportunities and challenges to the local economy. Taking Greater London as the study area, this paper integrates conventional statistics and survey datasets with novel crowdsourcing big data sources to identify and estimate four types of tourist grocery demand at the small-area scale: travellers staying with Airbnb, tourists using traditional commercial accommodation, guests staying with relatives or friends and day trip visitors. Based on this combined tourist retail demand layer we show the spatial variations at the small area level and, as an illustration of the demand uplift, we estimate additional grocery expenditure that is associated with this tourist demand. Thus, the paper indicates the neighbourhoods with significant grocery demand uplift from tourist stays. We argue that the new retail demand layer has tremendous potential to be used as an additional input to retail location modelling tools to support new store revenue estimation and store performance evaluation within the grocery retail sector.

Keywords: Small-area demand estimation; Grocery retail; Urban tourist; Airbnb; Greater London

5.1 Introduction

Spatial models are important tools in retail location planning, especially within the UK grocery sector. A major step in the modelling process is the estimation of retail demand within small-areas in order to then predict individual store revenues and trading potential. The estimation of small-area retail demand has traditionally been residence-based, using population data from censuses coupled with expenditure data from surveys such as the UK Office for National Statistics (ONS) Living Costs and Food Survey. Usually, that census data can be disaggregated to estimate demand by age, social class, ethnicity etc. (Birkin et al., 2017). However, for areas that have a more complex population composition, demand estimation may not be adequately captured using residence-based demand alone. Recent work has started to shed light on other essential drivers of retail demand, covering populations of work-based consumers in major cities (Berry et al., 2016), school and university students (Waddington et al., 2019), and tourists at coastal resorts (Newing et al., 2015). However, for towns and cities also hosting large volumes of tourists, it is rare to see any consideration of the potential uplift in retail demand from these consumers. One of the reasons why this has been under-researched to date is the lack of effective fine-granularity datasets related to tourist travel behaviour and expenditures, especially at the small area level.

As one of the world’s most visited destinations, London attracted 19.1 million international tourists in 2018 alongside 11.9 million domestic visitors (VisitBritain, 2019; VisitEngland, 2018). These visitors generated spending estimated at £12.33 billion and £2.98 billion respectively (VisitBritain, 2019; VisitEngland, 2018). Whilst the UK International Passenger Survey and Great Britain Tourism Survey provide monthly data on visitor numbers and spending, little is known about the distribution of these tourists or their economic impact at the local level.

Tourists and day visitors make up over 10% of the daytime population in London, which brings both tremendous opportunities and challenges for the local economy (Greater London Authority (GLA), 2014). Also, for tourists staying within London overnight, previous studies have recognised that a large share of their time is spent in the immediate vicinity around their accommodation locations (Shoval et al., 2011). Grocery stores are not usually located purely to serve tourists but many individual stores may be impacted by tourism. However, tourist or visitor expenditures on groceries, much of which take place at a local level, are rarely recalled in sample-based surveys of tourist expenditures. Therefore, this paper aims to estimate the demand from tourists and day visitors for groceries in London at the small-area scale. Aligning with the International Passenger Survey and Great Britain Tourism Survey definitions, we identify four groups of prominent tourists in London: visitors using the sharing-economy accommodation platform Airbnb; tourists staying at hotels and other serviced accommodation (i.e. B&Bs, guest houses, motels and hostels); ‘free’ guests who spend their nights with relatives or friends; and day trip visitors who only undertake leisure activity-based trips during the day/evening without using overnight accommodation.

Since there is little information from existing surveys concerning the spatial distribution of these four groups, we explore the potential offered by novel big datasets (principally from Airbnb and Twitter) when combined with conventional official statistics and survey datasets. The Airbnb listing and reservation records (available at the property level) offer new insights into the distribution and occupancy rates of their guests. In addition, user-generated crowdsourced data (such as those extracted from geotweets) could enable detailed space-time modelling at the individual level to help capture the distribution of day trip visitors, who are not captured by accommodation-based measures of tourism.

Thus our aim in this paper is to estimate the small-area grocery demand of each of the four tourist groups at a neighbourhood level and to incorporate them into a final tourist grocery demand layer for London. It is argued that the additional demand from these types of tourism is substantial and produces a significant uplift of expenditure (and hence revenues) across many localities in London, some of which are not traditionally associated with tourism activity. We argue that the incorporation of tourist retail demand into the spatial modelling process has the potential to enhance retailers' managerial decision-making in two major ways. First, it is important retailers understand more about the type of shoppers who frequent different stores. This can have implications for local marketing, store promotions and local stock provision. Second, it is especially important in estimating revenues for future stores. Forecasts of future revenues based purely on residential demand may not provide the necessary profit to make a store economically viable and hence the retailer may miss an important opportunity for growth. The inclusion of tourist demand may take revenues beyond the threshold needed for long-term store viability.

The rest of the paper is structured as follows. In Section 2, we review the existing literature on small-area retail demand estimation in location planning, particularly for non-residential populations. We argue that small-area tourist demand estimation, in the context of urban tourism, is far under-researched and that location-based online big data can effectively act as an important data source for tourist activities and consumer behaviour. Section 3 details the datasets used in our research. In Section 4 we estimate the potential grocery expenditure of each of the 4 groups. The final, combined, tourist retail demand layer in London is illustrated and discussed in Section 5. Finally, Section 6 offers some concluding comments. Throughout the paper we use the UK grocery market, and specifically London, as our case study.

5.2 Literature review

5.2.1 Retail demand in location planning

The estimation of small-area retail demand is important in its own right but is also crucial when constructing retail location planning models. The standard approach is to combine small area population census statistics with market survey data (in the UK from sources such as the Living Costs and Food Survey) which gives average weekly household expenditures for a variety of retail goods and services. For the production of more realistic spatial models, retail demand is usually disaggregated by person type (to take account of different types of consumer behaviour) including gender, age, ethnicity, social class/ income. Based on the effective simulation of residential demand, it is possible to apply these demand side estimations within spatial models to facilitate retail location planning (Birkin et al., 2017).

Due to the ongoing collaboration between academia and a growing number of major UK retailers, valuable commercial big datasets, including store trading records and loyalty card scheme data, are becoming more accessible to the academic community for research purposes. These corporate datasets have already been used to validate the success of store revenue forecasts in traditional models which focus substantially on residential demand (Birkin, 2019). Newing et al. (2013a) demonstrated the possibility of adding non-residential retail demand into spatial models (in this case adding seasonal visitor demand around coastal tourist resorts in Cornwall, a coastal region in south west England that receives significant tourism) using newly available customer-level loyalty card sales data provided by a partner retailer. This provided valuable new information for store location planning. They identified that the proportion of trade accounted for by non-residents in one case study coastal store varied from a peak of 50% in August to only 12.5% in January. Without such detailed estimations, the prediction of store revenues within such tourist areas in peak seasons has often been underestimated by retailers. It is very difficult to simply upscale residential demand consistently to obtain accurate demand-side or individual store revenue estimates (Newing et al., 2014). Subsequently, Berry et al. (2016) identified workplace and commuter-derived trade from retail partner transaction records at a number of convenience stores in London. They noted that these stores exhibit huge temporal spikes in sales of ‘food to go’ products, especially when in close proximity to major employment centres or transport interchanges. Waddington et al. (2017) demonstrated that a conventional residential-based retail model works poorly when attempting to estimate the revenues for convenience stores located in catchments containing workplaces or which are in proximity to universities, colleges or large secondary schools. These areas have a more diverse demand than stores serving a predominantly residential demand and these stores thus exhibit a sales pattern which may reveal noticeable demand uplift associated with these different drivers of demand (often at different times of the day/week).

In the above-mentioned papers, the distribution of the subgroups is still mainly based on static census-derived statistics and bespoke survey datasets. For example, Berry et al. (2016) used census-derived workplace population statistics to understand the micro geographies of workplace demand. Waddington et al. (2017) disaggregated non-residential retail demand into four different drivers: workers, school children, university students and daytime visitors. In common with Berry et al. (2016) they made use of workplace population statistics to capture the workplace population, supplementing these estimations with data from administrative and survey sources to capture demand associated with schools, universities and principal tourist attractions. In the case of coastal tourist demand, Newing et al. (2013a) explored the linkage between tourist grocery demand and the distribution of self-catering accommodation at the small area level. They used listings of self-catering tourist accommodation, in conjunction with surveyed occupancy rates, to estimate the small area spatiotemporal distribution of tourists as a demand-side input to a retail location model (Newing et al., 2013b).

Estimating the expenditure rates of the non-residential population is also a major challenge, since rarely can such expenditure rates be directly sourced. For example, after discussion with industry representatives, Waddington et al. (2019) incorporated an expenditure of £5 per person per week as the workplace-derived expenditure on groceries, while the school- and university-based expenditure were allocated £3.50 and £1.50 per person per week respectively, according to available consumer research in ad-hoc surveys. For the estimation of visitor demand in Cornwall, Newing et al. (2013b) collected data from a variety of organisation reports, industry surveys and academic research related to the self-accommodation hospitality sectors in the UK to estimate the different spending rates of visitors staying at different types of accommodation.

These studies highlight that store trading records and customer-level loyalty card data can provide evidence of the links between store sales and spatiotemporal demand fluctuations in a catchment area, driven by the ebbs and flows of transient populations within the region. They also highlight that little is known about the expenditure patterns of non-residential demand types and that both academic and industry research have to rely on (often incomplete) ad-hoc surveys and insight to infer expenditures associated with these demand sub-groups. We argue in this paper, in relation to urban tourist demand, that a bottom-up approach is useful in order to construct a comprehensive tourist-based demand surface. This involves aggregating the estimated expenditure of each demand subgroup, accounting additionally for their spatial and temporal distributions.

5.2.2 Urban tourism and big data

As noted above, there is a deficiency of official tourism statistics at a disaggregate spatial scale. This presents a number of challenges for tourist population modelling. However, big data is beginning to help fill that gap. There are two major types of big data which can be valuable in tourist research. The first is user-generated content from location-based social media (i.e. geotagged tweets, geo-photos, and reviews of point-of-interests (POIs) etc.). Such data are beginning to enable space-time modelling at the individual level and have been used to assess international mobility patterns (Barchiesi et al., 2015), monitor visitor behaviour around tourist attractions (Tenkanen et al., 2017), identify tourist hotspots in cities (Kádár, 2014) and to characterise tourist flows within destinations (Chua et al., 2016). The academic community has started to leverage these emerging big data sources as important supplements to conventional statistical sources. This enables fine-scaled spatiotemporal patterns of tourist distributions to be uncovered (comprehensive reviews appear in Li et al., 2018, Ye et al., 2020). Such studies are shedding new light on the role that emerging location-based big data could offer in helping to outline the spatial distribution of tourists, but none has focused so far on the use of this data to provide more nuanced local retail expenditure estimates. The second big data source covers tourist accommodation. The location of tourist accommodation has a profound impact on tourist movements and can impact greatly on service providers that are in the catchment area of these accommodations. These include data sources such as AirDNA which provides data on the growing use of Airbnb. This dataset is beginning to be used in tourist studies more widely (Gutiérrez et al., 2017). We shall examine both these types of data more fully in the next section before using them to estimate retail demand in later sections.

5.3 Data and methodology

To fully understand tourist grocery demand at the small-area level, we need to capture tourists who stay in London using a range of different accommodation types. These first include the rapidly growing sharing economy platform Airbnb. Primarily drawn from the existing housing stock, Airbnb has become a popular alternative to traditional accommodation, enabling tourists to have distinctive and more local experiences at their destination (Greater London Authority, 2017). This research uses a big dataset supplied by AirDNA and accessed via the Consumer Data Research Centre (CDRC) at the University of Leeds. The AirDNA dataset enables us to retrieve all the Airbnb listing and reservation records in London for a 12-month period: here we access data between June 2017 and June 2018. The accommodation capacity of the Airbnb properties, coupled with reservation records, is used to delineate the spatial distribution of Airbnb rental units, identify typical occupancy rates and thus calculate the number of local Airbnb guests during the one-year time span.

In addition, it is important to look at the more traditional accommodation sectors. According to the International Passenger Survey (2019) and the Great Britain Tourism Survey (2019) (Table 5.1), for both overseas and domestic tourists, the two main accommodation types are serviced commercial accommodation (hotel/guest houses, bed & breakfast, hostel/university/school) and free guest/ home tourists staying with relatives or friends in their own houses. Since individual establishment level data related to the provision of serviced commercial accommodation is hard to access from any official channel, we have generated a dataset derived from the combination of Ordnance Survey Point of Interest (POI) data and OpenStreetMap. As a result, a total of 2,042 geo-located serviced accommodation establishments in London have been identified and extracted from those sources.

Table 5.1 % of nights by accommodation type in London, 2018.

		Visits (000)		**Nights (000)**		**Spend (£m)**	
Inbound tourists	**All staying visits**	**19,090**		**110,932**		**12,329**	
	Serviced	12,514	65.56 %	57,590	51.92 %	8,959	72.67 %
	Non-serviced	914	4.79 %	8,185	7.38%	873	7.08%
	Free guest / Own	5,189	27.18 %	42,717	38.51 %	2,312	18.76 %

	home						
	Others	567	2.97 %	2,441	2.20%	186	1.51%
Domestic visitors	**All staying visits**	**11,852**		**27,878**		**2,983**	
	Serviced	6,631	55.90 %	12,293	44.10 %	2321	77.80 %
	Non-serviced	334	2.80 %	745	2.70%	102	3.40%
	Free guest / Own home	4,725	39.90 %	13,580	48.70 %	515	17.30 %
	Others	325	2.80 %	975	3.50%	70	2.40%

Source: International Passenger Survey (2019) and Great Britain Tourism Survey (2019)

Whilst Table 5.1 shows that we have estimates of the total number of people staying with friends, there is no definitive source which records the location of these visits. The simplest approach to generate this data would be to distribute known numbers of these tourists/guests across all residential areas of London pro rata to their population size. For domestic visitors that makes sense. However, it is more likely that international visitors would stay with relatives or friends who share a similar country of birth or ethnic origin: for example, witness the great concentration of Australian tourists and visitors around Earl's Court in London. Although this will not always be the case we feel this method of allocation to residential areas makes more sense than an even distribution across all areas of London. We can plot the residential population by country of origin (derived from 2011 Census-based statistics), and link this to data on international tourists by country of origin/ethnicity provided by the UK International Passenger Survey (see Section 4 below).

Another overlooked tourist group who may also impact upon local retail demand is day visitors. In a daytime population survey by the Greater London Authority in 2014, they categorise day visitors as '3 Hour + Leisure Day Visitors'. In addition, the Great Britain Day Visits Survey (2018) notes that these visits are distinguished by 'visitors not undertaking activities that would regularly constitute part of their work or would be a regular leisure activity'. Day visitor statistics at the London borough level are available via the Greater London Authority daytime population survey (unfortunately no fine-scale statistics are available). Thus we need to spatially redistribute those borough level day visitor estimates across the Lower Super Output Areas (LSOAs) within each respective borough. To do this we can estimate the levels of tourist/leisure activity in different LSOAs using another big data source – geotagged Tweets (geotweets) from the popular social networking service Twitter. We collect geotweets within our London study area from Oct. 2018 to Sept. 2019 which occurred between 9 am and 5 pm (daytime geotweets). By tracing back the historical geotweets of these Twitter users, we can usually infer their countries of origin. The modelling of day visitor distributions only leverages geotweets from UK domestic Twitter users. Day visitor activity is distinguished from activity around the usual residence of a user by examining the frequency of tweets – more than 10 geotweets per user at a particular location during our data collecting timespan would suggest a home location. The day visitor count in each borough is distributed across its constituent LSOAs according to the proportion of geotweets in each of those LSOAs.

For each of the four types of tourists outlined above, we plot the spatial pattern of their distributions across the LSOAs in London in Section 5.4. Then we estimate the average personal expenditure rate per week of each tourist type (for grocery shopping). The resultant small-area tourist grocery demand layer across London is thus generated by the combination of these four demand groups. The datasets employed in the research are summarised in Table 5.2 and the appendix gives a web link for accessing these data sets directly.

Table 5.2 Data sources and dates of availability.

Visitor Type	**Data source**	**Date**
Airbnb	AirDNA	May 2017 – Jun. 2018
Serviced accommodation	Ordnance Survey POI	Dec. 2019
	OpenStreetMap	Dec. 2019
Free guest/own home	International Passenger Survey	2018
	Great Britain Tourism Survey	2018
	Census 2011	2011
Day visitors	Daytime population Survey	2014
	Twitter	Oct. 2018 – Sept. 2019

5.4 Estimating tourist grocery demand

5.4.1 Airbnb guest

As a new form of commercial self-catering accommodation, Airbnb guests can dine out or purchase and prepare their own food in-house. Airbnb (2018) reported an average expenditure of £100 per guest per night from their international guests in the UK, with an average 10% spent on groceries and 33% spent on food and drink eaten in other establishments. Additionally, 43% of the £100 is spent in the neighbourhoods in which they stay Airbnb (2018). Therefore, we may expect a substantial uplift of revenue for the grocery stores in the localities that host many Airbnb guests. In the absence of any other survey data for now we use the £10 per guest per day reported by Airbnb (2018) as the average grocery spending rate of an Airbnb guest. It is plausible that the actual rate in London would be typically higher on a per-person per night-basis than comparable stays elsewhere in the UK, given that London accounts for 41.7% of all overnights visit but 53.8% of inbound tourist expenditure in the UK (VisitBritain, 2019).

Figure 5.1 shows the accommodation capacity of Airbnb listings aggregated into neighbourhoods (LSOAs) across London. In common with all forms of accommodation, it shows a centralisation of accommodation stock within central London where many of the main tourist attractions are located. However, compared to the hotel distribution across LSOAs in London (Figure 5.3), Airbnb accommodation is far more widely dispersed, including within many residential areas that are not traditionally associated with tourism activity.

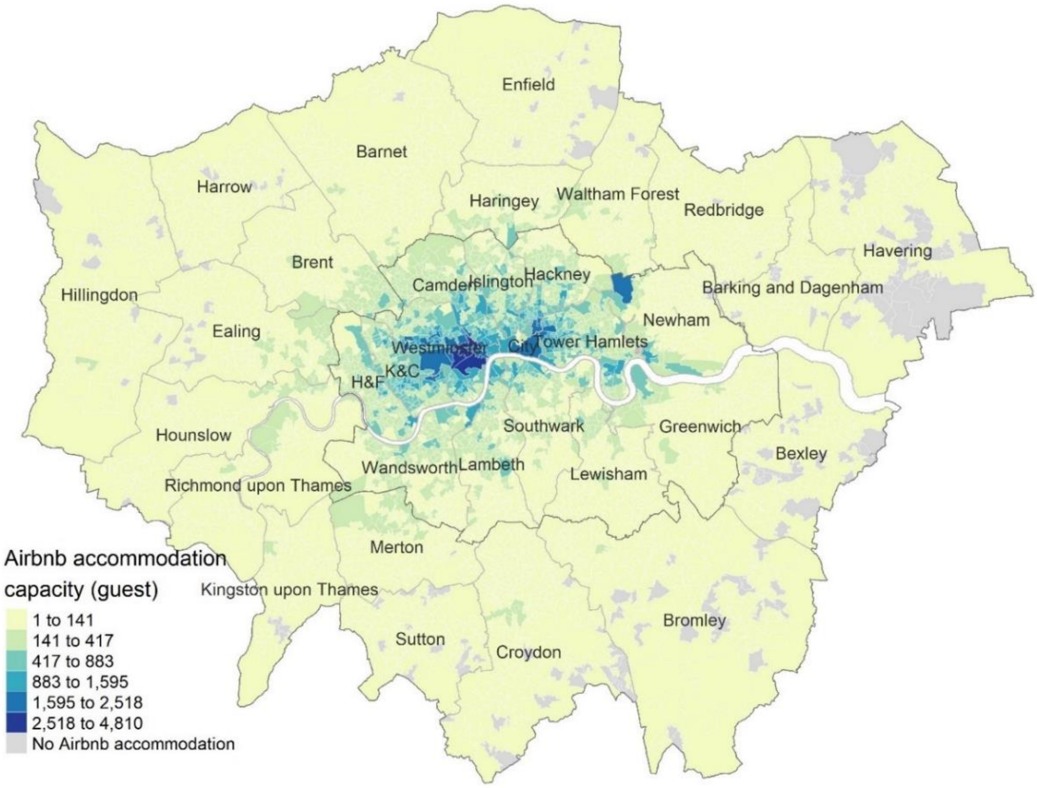


Figure 5.1 Spatial distribution of Airbnb accommodation by LSOA.

To depict the spatial distribution of occupied Airbnb premises (and therefore the presence of tourists), we use the reservation records of each Airbnb listing during the 12-month period to obtain the actual utilisation of the Airbnb properties in London. For all the 207,117 properties listed on AirDNA, only 51.65% (106,974) have online reservation records between June 2017 to May 2018, accounting for 24.6m guest nights. The remainder represents listings that appear not to have received paying guests, at least not via Airbnb. Combining accommodation capacity and the occupancy rate for these properties, alongside an average grocery expenditure of £10 per guest per day, the estimated Airbnb guest grocery expenditure per week across the LSOAs in London totals a considerable annual demand uplift of £246.2m, or £4.72m per week. The spatial distribution of this additional tourist-driven spend is presented in Figure 5.2. The spatial pattern of Airbnb grocery demand in inner London presents a similar pattern to the distribution of capacity in Figure 5.1, but some Airbnb properties in outer London generate low demand due to low recorded occupancy. Nevertheless, Figure 5.2 illustrates that Airbnb properties generate tourist grocery expenditures which are distributed across Greater London and not solely associated with core central London tourist hotspots.

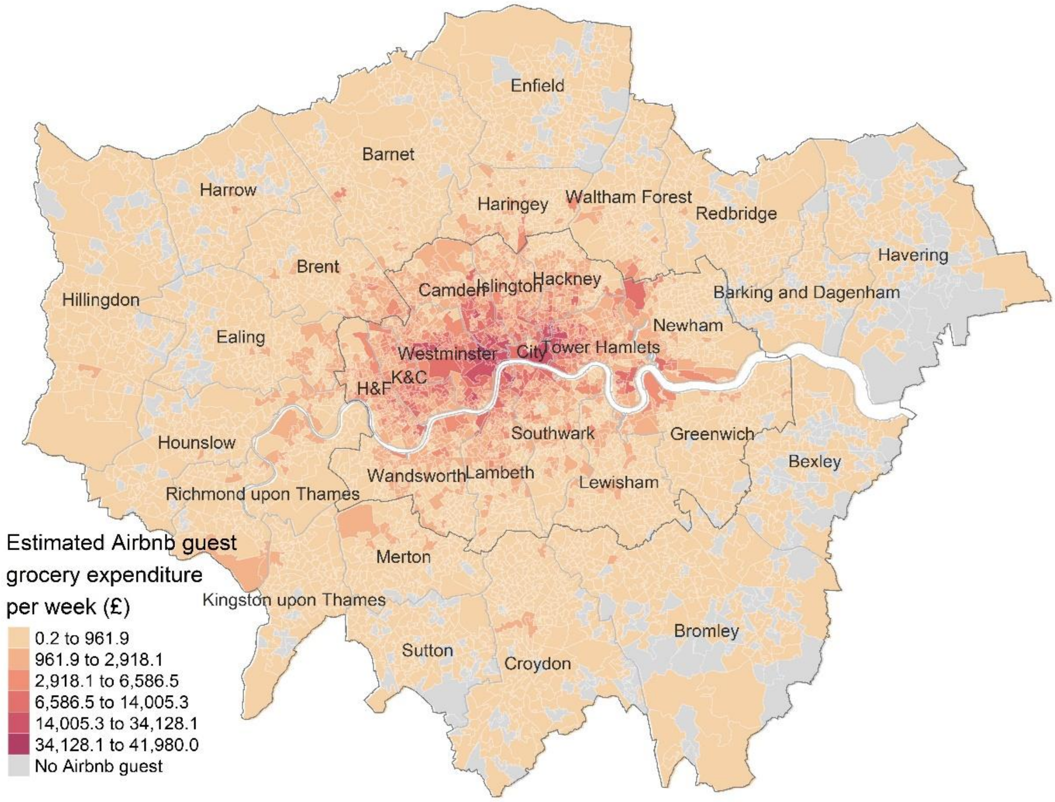


Figure 5.2 Estimated Airbnb guest grocery expenditure using Airbnb utilisation and grocery expenditure rates per week.

5.4.2 Hotel and commercial accommodation traveller

The Accommodation Stock Audit conducted in 2016 reported serviced and non-serviced accommodation provision at the borough level across London (VisitEngland, 2016), which is the finest spatial resolution source of published data relating to traditional commercial accommodation provision (Table 5.3). We assume that most of the non-serviced accommodation are listed as Airbnb properties and have therefore already been captured by the Airbnb dataset discussed above. Therefore, in this section we specifically consider expenditures associated with tourists staying within serviced commercial accommodation. By combining data from the Ordnance Survey POI and OpenStreetMap, we obtained the locations of 2,314 serviced accommodation establishments in London. The Accommodation Stock Audit gives the number of rooms or bed space for each London borough which we disaggregate evenly across LSOAs in each borough. Figure 5.3 maps the estimated bed space counts for serviced accommodation across the LSOAs in London. Most are centrally located as expected, but the patterns clearly show a cluster of accommodation near Heathrow Airport to the far west of London.

Table 5.3 Accommodation stock in London.

	Establishments	Rooms	Bedspaces
Serviced Accommodation (Hotels and similar)	2,582	197,624	448,160
Non-serviced Accommodation	319	36,266	40,182

Source: VisitEngland (2016)

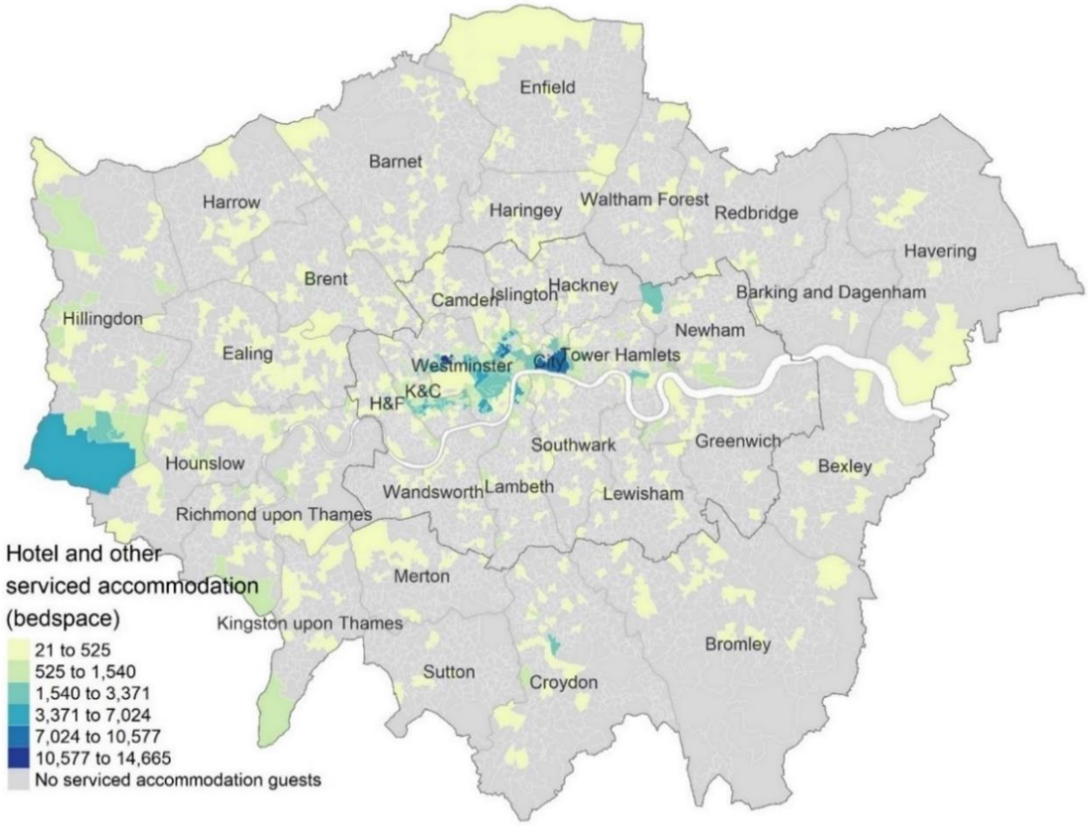


Figure 5.3 Serviced commercial accommodation stock (bedspace) across the LSOA.

The actual utilisation of the serviced commercial accommodation stock can be estimated by multiplying the number of bed spaces by headline published occupancy rates for serviced accommodation. According to VisitEngland (2019), the annual average bed space occupancy of serviced accommodation in 2019 in London is 60.3%, i.e. on any given night approximately 60% of bed spaces are occupied and thus have potential to generate grocery expenditures. Clearly this will be higher in the summer months but since London is subject to fewer seasonal peaks in tourism than other destinations in the UK, we do not consider seasonal pattern analysis in this paper. Tourists staying in serviced accommodation rarely have cooking facilities provided and therefore will not be major purchasers of groceries. Their grocery shopping behaviour is thus more likely to be similar to workplace populations, who typically purchase snack food items to go (Berry et al., 2016). Therefore, in the absence of any more comprehensive survey data or insight into these tourist expenditures, we deem it sensible to use the same £5 per person as the average groceries expenditure (per week) associated with a typical non-residential worker (Waddington et al., 2017). Using this method, the grocery expenditure of serviced accommodation tourists in London is estimated at £1.35m per week. The spatial distribution of this additional grocery expenditure is shown in Figure 5.4.

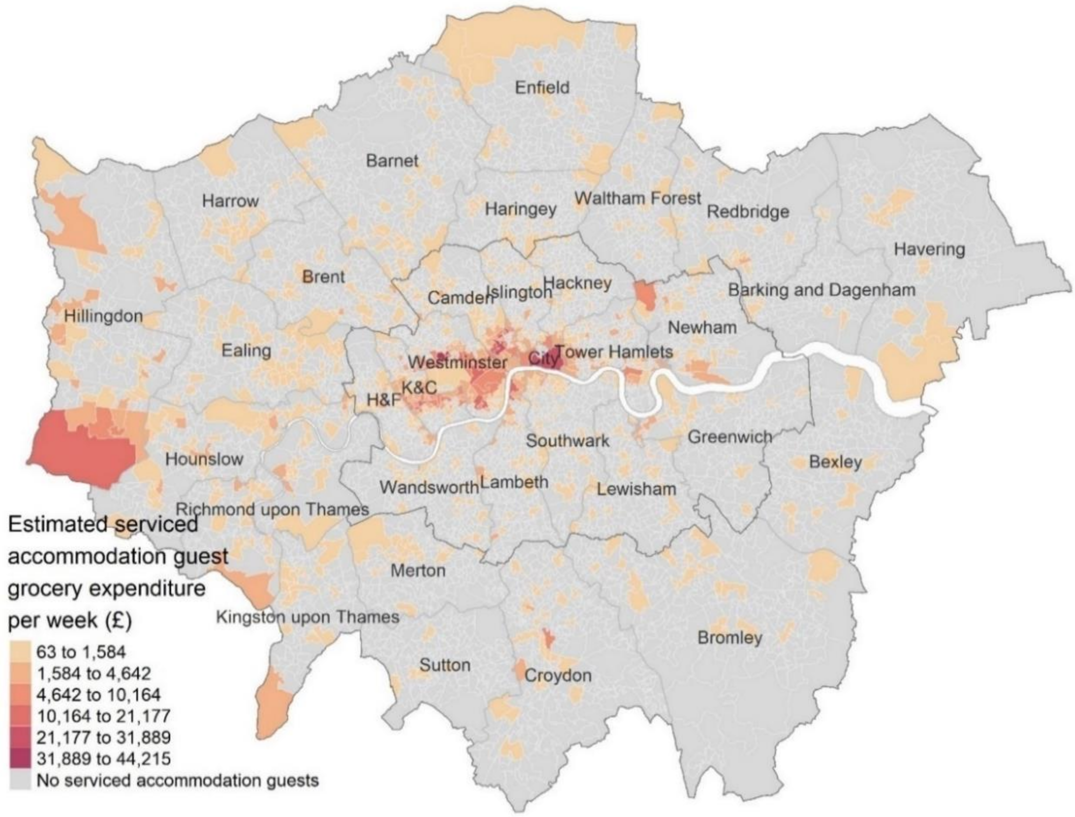


Figure 5.4 Estimated serviced commercial accommodation traveller grocery expenditure.

5.4.3 Free guest/own home

Guests staying with relatives or friends is another major form of accommodation in London, whose economic impact has been largely overlooked (Shani and Uriely, 2012; Backer, 2007). As Table 5.1 illustrates, 27.2% of the inbound tourist visits in London are made by guests staying with their relatives or friends, making up for 38.5% of all inbound tourist nights. Domestic visitor numbers in London are even higher – around 40% of the trips are registered as staying at a friend or relative's home, accounting for 48.7% of domestic tourist nights in London. Inbound tourists and domestic visitors generate around 19% and 17% respectively of the total spend in London, but how these visitors impact the local economy at the small-area level is an issue that is rarely examined (International

Passenger Survey, 2019; Great Britain Tourism Survey, 2019). Guests staying with local residents are likely to generate a further uplift to local grocery demand.

Table 5.4 lists the total night stay of international and domestic guests in London. Using this International Passenger Survey data we now allocate the total nights in London of each country across the LSOAs according to the distribution of the usual residents in the corresponding ethnic group (as described above). Thus, we first assign each country of origin to the corresponding ethnic group.

All 38 countries listed can be grouped into 20 ethnic census groups, as shown in the last column of Table 5.4. The 14 European countries (France, Spain, Germany, Netherlands, Switzerland, Hungary, Portugal, Belgium, Sweden, Norway, Czech Republic, Austria, Finland and Luxembourg) can be categorised as “European Mixed”; Polish, Irish, Italian and Greeks can be kept as unique ethnic groups; USA and Canada can be combined as “North American”; UAE and Saudi Arabia as “Arab”, and mainland China and Hong Kong are taken together as “Chinese”. Most of the other countries can be directly associated with one specific ethnic group, except South Africa, Singapore and India. In this study, South African is combined with African, White African and White and Black African; Singapore is added to Chinese and Malaysian; Indian can be aggregated with Anglo Indian and Indian or British Indian; finally, Russian is represented by CIS as a whole since no separate Russian group is provided.

The population distributions of these 20 ethnic groups for the LSOAs in London can be retrieved from the 2011 Census and are illustrated in Figure 5.5. By distributing the total visitor nights in London by each ethnic group according to Table 5.5, we have been able to use the spatial distribution of these usual residents by ethnic group to estimate the spatial distribution of guests staying with friends or relatives across LSOAs in London. By summing the tourist nights staying with different ethnic groups in each LSOA, Figure 5.6(a) visualises the spatial distribution of the inbound guests staying with family or friends. Domestic visitors are equally important as the international tourists in bringing more demand into local residential areas in London. However, for domestic visitors it is difficult to meaningfully allocate these visits in relation to any underlying ethnic indicator. Therefore we have allocated these 13.58m domestic free guests evenly across households in the LSOAs in London, and Figure 5.6(b) presents the spatial pattern of the domestic guest staying with family or friends.

Table 5.4 Inbound free guest/own home tourist nights in London and their corresponding ethnic group.

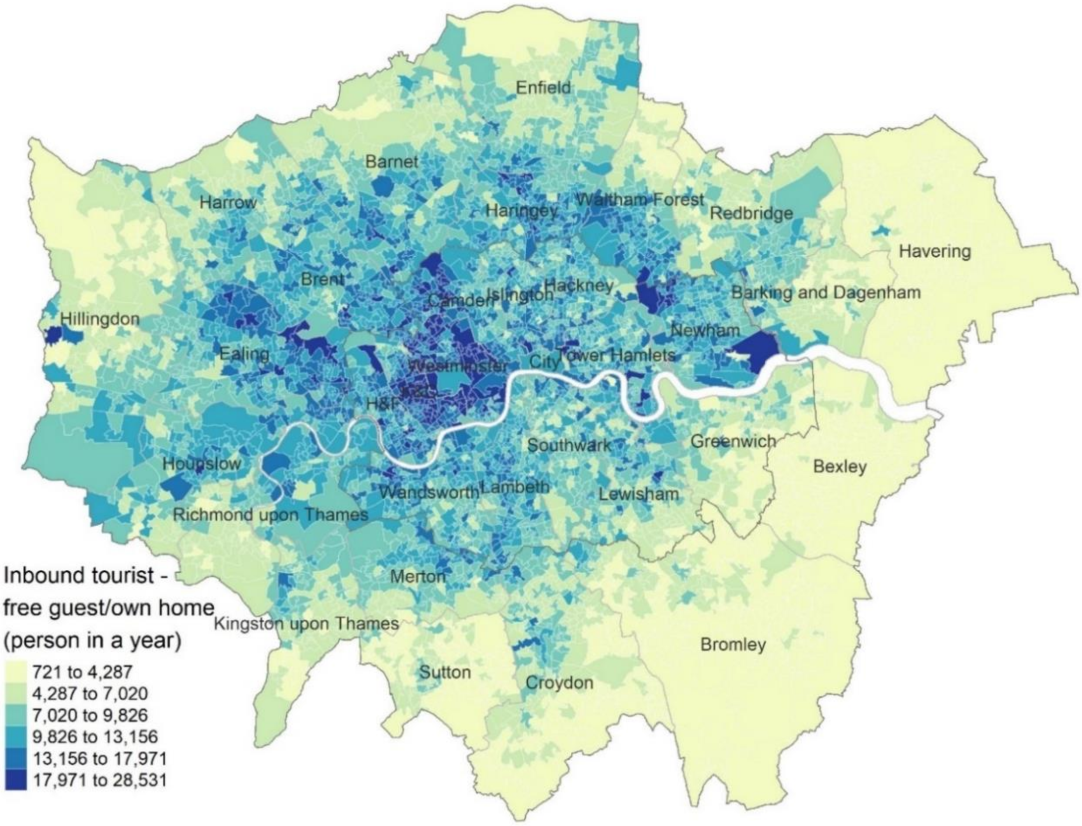
Country	Total nights (Free guest)	Total nights (Own home)	Total nights	Ethnic group (detailed) in 2011 Census
Total	41,523,000	11,945,235	42,717,524	
USA	3,505,271	83,860	3,589,132	North American
Poland	3,102,747	11,852	3,114,599	Polish
France	2,800,603	93,267	2,893,870	European Mixed
Spain	2,769,689	56,302	2,825,991	European Mixed
Australia	2,482,524	23,328	2,505,852	Australian / New Zealander
India	2,469,332	25,290	2,494,622	Anglo Indian Indian or British Indian
Germany	1,617,935	61,175	1,679,110	European Mixed
Ireland	1,412,140	31,667	1,443,808	Irish
Canada	1,408,221	50,194	1,458,415	North American
Italy	1,273,430	30,731	1,304,162	Italian
Netherlands	831,570	64,512	896,082	European Mixed
New Zealand	686,738	3,682	690,420	Australian / New Zealander
Switzerland	628,188	20,802	648,990	European Mixed
South Africa	621,811	-	621,811	African White African White and Black African
UAE	569,399	152,078	721,477	Arab African/Arab White and Arab
Hungary	520,186	10,033	530,219	European Mixed
China	480,583	4,476	485,059	Chinese
Portugal	478,526	8,428	486,954	European Mixed
Belgium	451,961	42,995	494,957	European Mixed
Sweden	446,313	2,169	448,483	European Mixed
Argentina	442,619	-	442,619	Argentinian
Denmark	392,517	9,933	402,450	European Mixed
Greece	376,008	1,489	377,496	Greek Greek Cypriot
Norway	348,393	1,656	350,050	European Mixed
Czech Republic	314,921	10,425	325,347	European Mixed
Singapore	310,734	9,667	320,401	Chinese Malaysian
Hong Kong	269,635	22,864	292,499	Chinese
Brazil	233,855	3,952	237,807	Brazilian
Malaysia	226,342	5,314	231,656	Malaysian
Saudi	167,179	33,691	200,870	Arab

Arabia				African/Arab White and Arab
Russia	159,499	1,622	161,121	Commonwealth of (Russian) Independent States
Austria	146,052	669	146,721	European Mixed
Thailand	141,522	4,232	145,754	Thai
Japan	123,613	288	123,901	Japanese
Finland	103,995	2,160	106,155	European Mixed
Mexico	103,845	-	103,845	Mexican
South Korea	91,933	-	91,933	Korean
Luxembourg	58,674	1,633	60,307	European Mixed
Other	8,954,497	308,085	9,262,582	Other

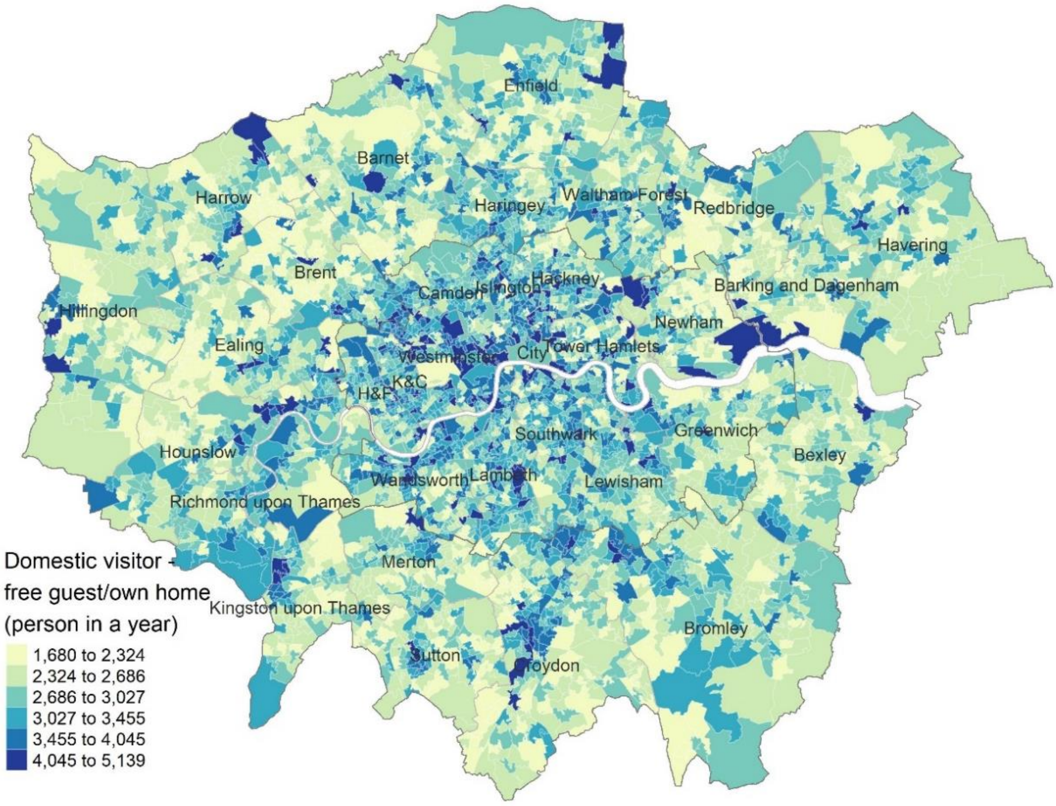
Table 5.5 Free guest/own home tourist nights by ethnic group.

Ethnic group of the usual resident	Total free guest/own home nights in London
European Mixed	12,295,685
North American	5,047,547
Polish	3,114,599
Australian/New Zealander	3,196,272
Indian	2,494,622
Irish	1,443,808
Italian	1,304,162
South African	621,811
Arab	922,347
Chinese	777,558
Argentina	442,619
Greek	377,496
Singapore	320,401
Brazil	237,807
Malaysian	231,656
Russian	161,121
Thai	145,754
Japanese	123,901
Mexican	103,845
South Korean	91,933
Other	9,262,582

Figure 5.5 Spatial distribution of the usual residents of 20 ethnic groups at the LSOA level.



(a)



(b)

Figure 5.6 Free guest/ own home visitor distribution across LSOAs in London: (a) inbound tourist; (b) domestic visitor.

Having estimated the spatial distribution of guests across London we need now to consider additional household expenditure generated by these visitors. Here we can extract information from the Living Costs and Food Survey (2019). The average weekly household expenditure on “Food and non-alcoholic drinks” in London is £62.40, with the average number of persons per household being 2.6 people. Therefore, on average, expenditure per person on “Food and non-alcoholic drinks” for a week in London can be inferred to be approximately £24. Thus, we assume that the additional household expenditure associated with hosting a guest is £24 per guest per week. Figure 5.7 shows the spatial distribution of additional household expenditure associated with these visitors. Altogether, we estimate a sum of over £3.7m extra grocery spend per week is induced by hosting guests staying in London, which means the overseas and domestic guests staying with family or friends in London generate almost £193m grocery expenditure per year.

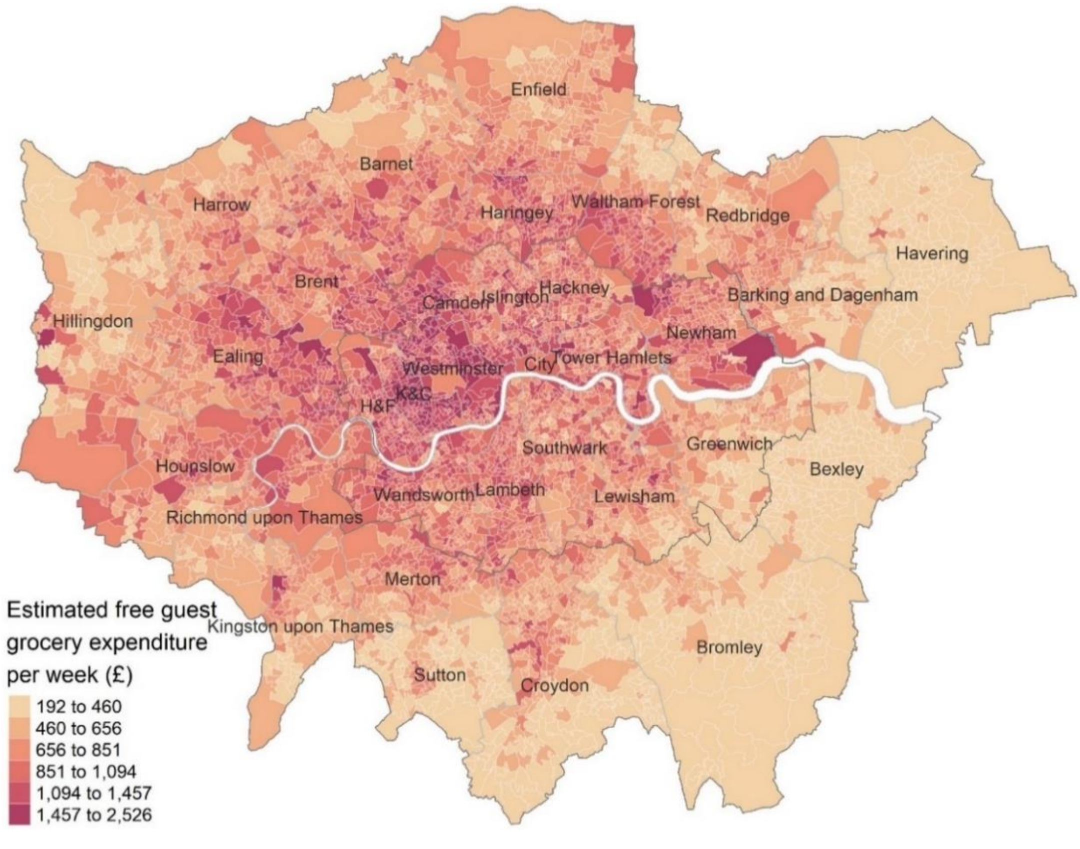


Figure 5.7 Estimated free guest visitor grocery expenditure across LSOAs in London.

5.4.4 Day trip visitor

London hosts 319.2 million day trips and receives more than £13.96 billion in expenditure from day visitors. At the aggregate level this is higher than the spend of overnight tourists from both overseas and domestic markets (Table 5.1). In fact, according to the Greater London Authority (2014), day trip visitors make up 7.33% of London's daytime population, whereas overnight visitors only account for 3.61%. However, the travel behaviour of day visitors is less-well understood, as well as their contributions to the local economy such as expenditures on groceries. The Great Britain Day Visits Survey (2018) reports that day visits to London are dominated by two source regions: London itself (79.0%) and the South East (7.5%). Since these visitors are staying at home overnight, their propensity to purchase groceries to prepare meals (as opposed to snacks or 'food on the go') is limited, and they are unlikely to exhibit a high grocery spend during their visit. Nevertheless, on aggregate, they could generate considerable additional expenditure on incidental purchases due to the large size of this group.

Day trip visitors are defined in line with the daytime population survey (Greater London Authority, 2014) as the 3 Hour + Leisure Day Visitors, whose counts at the borough level are available via the survey. However, day trip visitor statistics in London at the finer spatial scale are unavailable. To help make progress in this respect we harness the percentage of daytime geotweets in each LSOA. Since the day visitors are consisted by a majority of local residents in London as well as day-trippers from outside London, only the geotweets from domestic Twitter users are used at here. The number of day visitors in each borough can be distributed over the LSOAs according to the proportions shown in Figure 5.8.

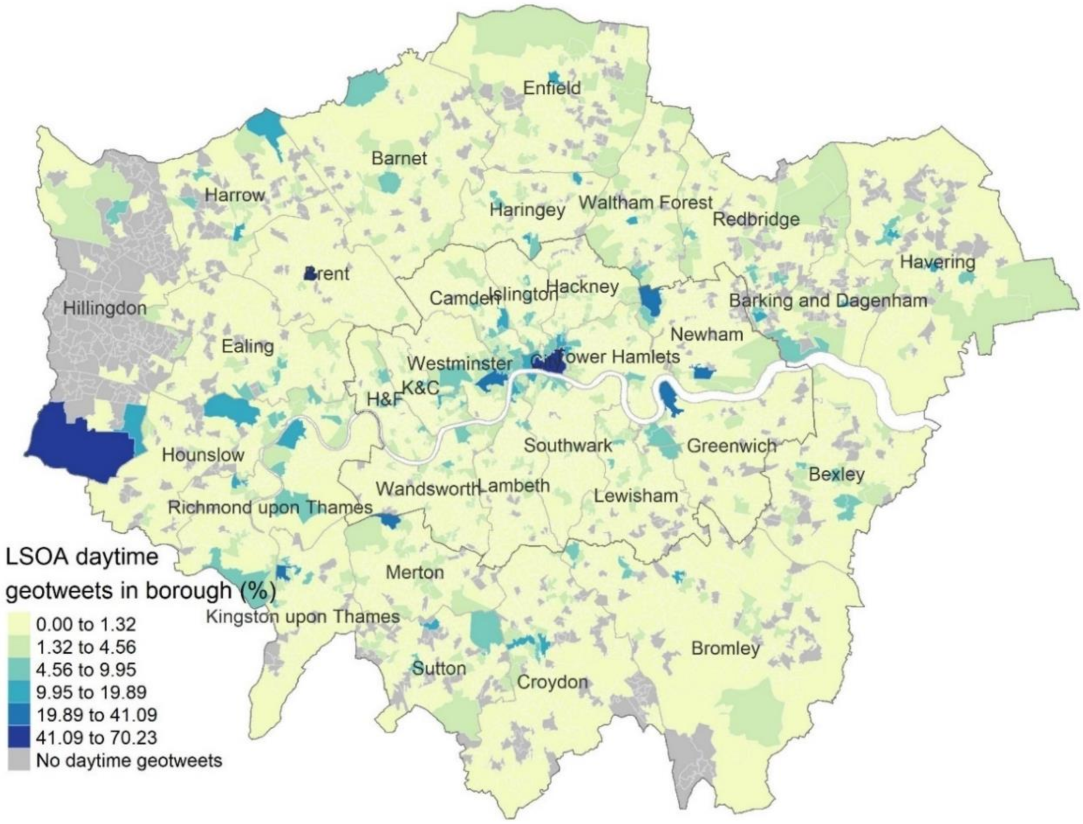


Figure 5.8 The proportion of daytime geotweets in the LSOA, nested within borough.

According to the Great Britain Day Visits Survey (2018), there is a historical average spend of £44 per 3hr+ visit in London, with purchases involving food in shops or takeaways accounting for 27% of the expenditure (£11.88) of these visits. In line with the suggestion of Newing et al. (2013b), half this expenditure is a good estimation of the amount spent in grocery stores, the remainder attracted to other takeaway food sources. This results in approximately £4.29m spent per week in grocery stores by day trip visitors, which accounts for an additional £223.7m grocery expenditure per annum. The spatial distribution of the estimated day visitor associated grocery expenditure across London is shown in Figure 5.9.

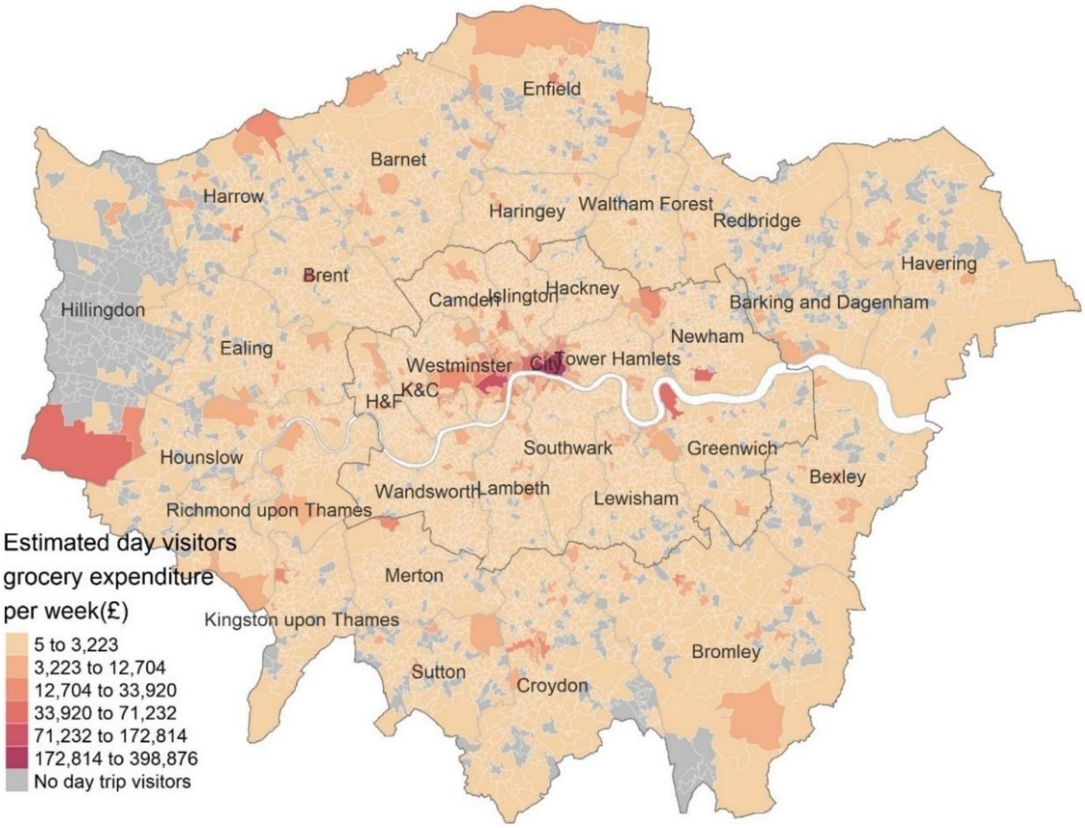


Figure 5.9 Estimated day trip visitor grocery expenditure across the LSOA in London.

5.4.5 Small-area tourist demand layer in London

In this paper we have attempted to build a new demand layer for tourists and visitors in London in relation to a variety of temporary accommodation types: Airbnb, traditional commercial accommodation, guests staying with relatives or friends and day trip visitors who do not stay in overnight accommodation. Table 5.6 exhibits a summary of the grocery demand from these four tourist groups. The total tourist expenditure in Table 5.6 is £14.06m. Whilst this is less than both the residential and workplace demand (estimated by the census population data coupled with expenditure rate as £199.8 m and £22.5 m respectively), it contributes a 6.32% increase in the total expenditure in London. What might this mean for individual grocery stores? The Geolytix Retail Points dataset (updated to Jan. 2020) provides individual grocery store locations and aggregated floorspace by LSOA in London (Figure 5.10). According to the Geolytix database, there are 1,759 grocery stores providing a sum of 14.8m square foot (sqft) of floorspace in London including 985 convenience stores (with a store size under 3,013 sqft), 647 supermarkets (store size between 3,013 and 30,138 sqft) and 127 hypermarkets (store size over 30,138 sqft). On average, the store revenue from residents and workers in London is £126,379 per week, while tourist adds a further £7,993 revenue to give £134,372 per week, resulting in a £16.0 per sqft per week sale density. This figure is in line with industry reports concerning the sales density of UK's top brands - for example Tesco is £17.11 per sqft in the UK and ROI (Tesco PLC, 2020). Of course, the spatial distribution of tourist activity we have shown above means that this additional expenditure won't be shared evenly among these stores, with some stores likely to attract considerable tourist expenditure whilst others very little. Our subsequent research will seek to allocate these additional tourist expenditures to stores in order to consider the spatial distribution of store level sales uplift.

Table 5.6 Summary of tourist grocery demand by accommodation type.

Tourist group		N. of tourists with grocery demand (average per day)	Grocery expenditure rate (per tourist per week)	Estimated grocery expenditure (per week)
Airbnb guest		67,448	£70	£4.72m
Hotel and other serviced accommodation tourist		270,691	£5	£1.35m
Free guest/own home tourist	Overseas	117,034	£24	£2.81m
	Domestic	37,204		£0.89m
Day trip visitor		103,096	£41.58	£4.29m

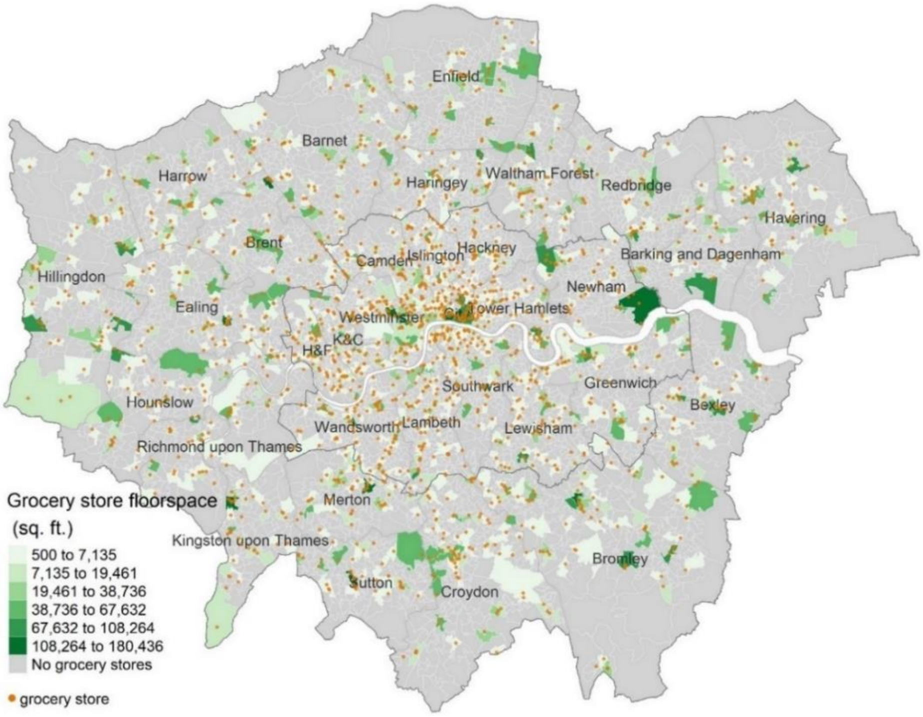


Figure 5.10 Grocery store location and aggregated floorspace by LSOA in London.

Figure 5.11 shows the combined spatial pattern of estimated tourist grocery expenditure at the LSOA level. By comparing this with the grocery expenditure of usual residents in London, Figure 5.12 presents the uplift of grocery expenditure by LSOA according to the addition of tourist demand. By overlapping with the tourist attractions recommended by VisitLondon (2020) we find that the LSOAs in inner London that are adjacent to major attractions unsurprisingly have substantial uplift. In addition, however, there are other neighbourhoods which have limited tourist attractions but also benefit from tourist demand, such as Heathrow airport, southeast Bromley, northwest Hillingdon and east Enfield. The demand within these areas can be further identified via the examination of demand type. For example, the outer London areas with sizeable uplift are usually due to day trip visits, whereas inner London shows more balanced demand sources from all the four tourist groups.

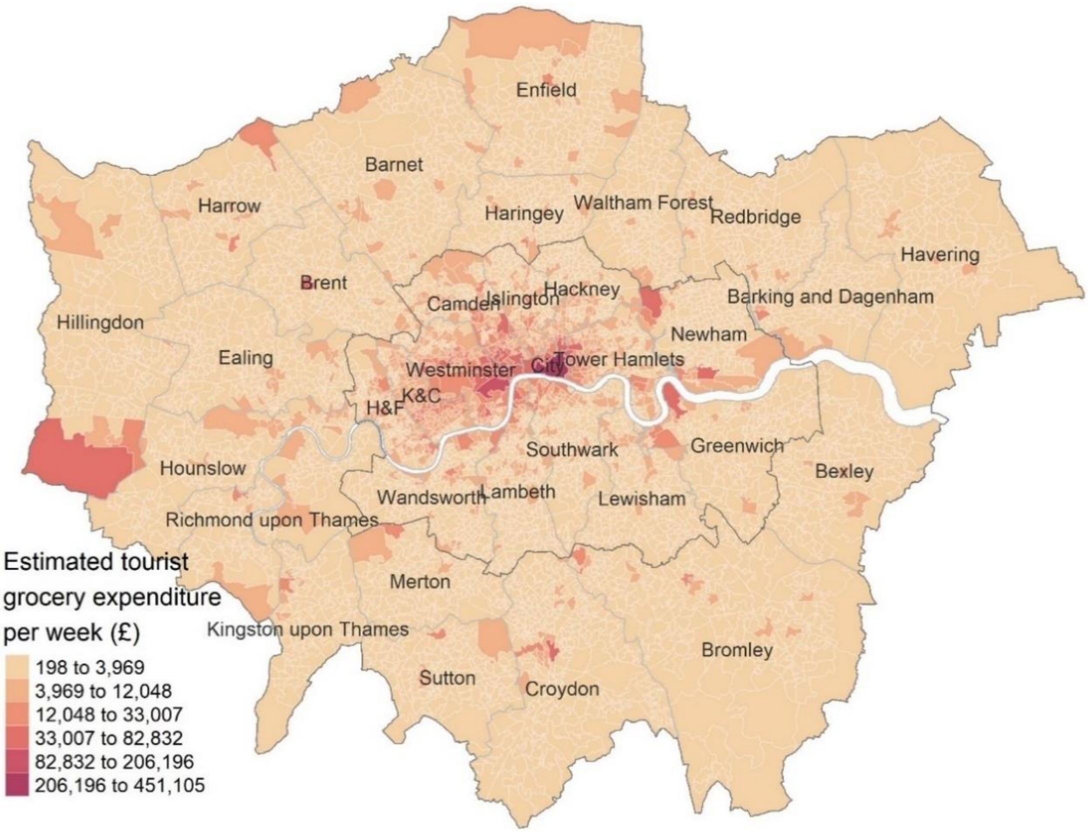


Figure 5.11 Estimated tourist grocery expenditure including Airbnb, serviced accommodation, free guest/own home and day visitor.

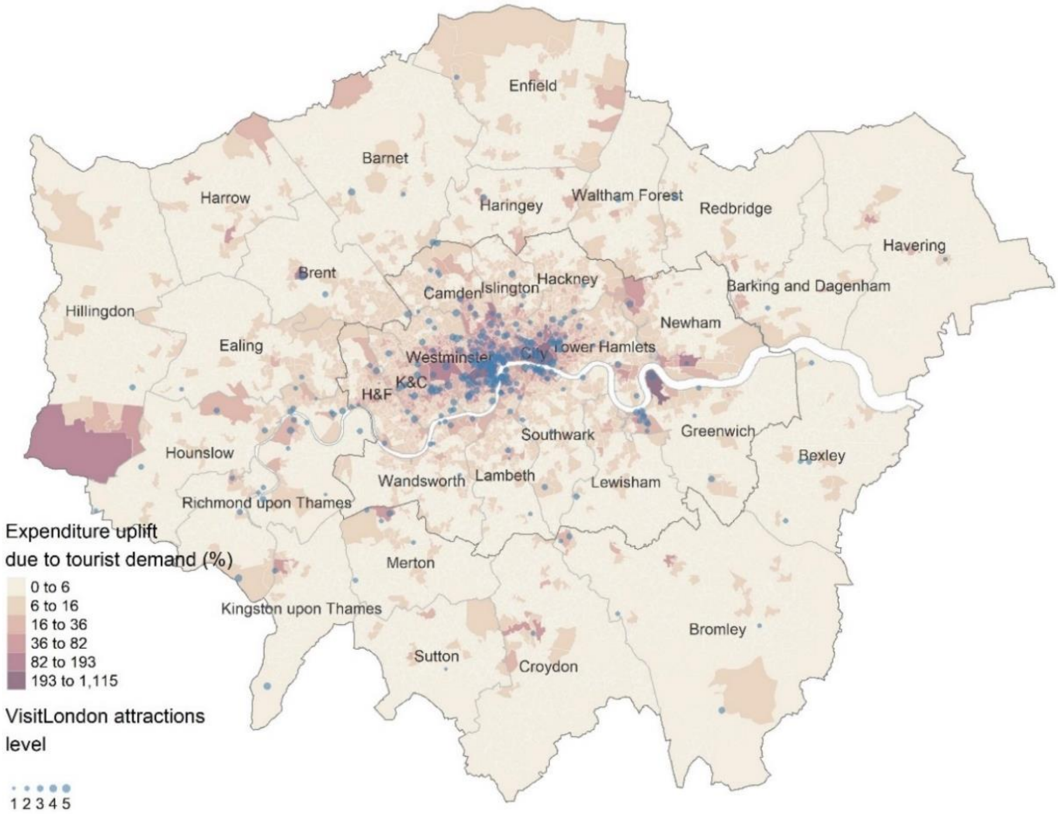


Figure 5.12 The uplift (%) in the usual resident demand due to estimated tourist expenditure.

5.5 Discussion

In this section we discuss how these results may be of benefit to individual retailers. Adding tourist demand into the estimation of local retail grocery demand first enables the retailers to understand the volume and composition of the customers in each individual store catchment area and tailor the product range to maximise sales opportunities. Any local marketing campaigns might thus be more focused around the accommodation stock in the area. In some areas of London we estimate that 20-30% of all revenue at individual stores may come from tourists, especially those that are likely to be self-catering. The catchment areas of stores in these areas may consider more fresh food which is easy to prepare (especially ready meals, 'food to go' etc). Store management teams might also be encouraged to think about different types of tourist. Mak et al. (2012) provides a good review of how tourists from different countries behave in relation to local food consumption. Some tourists value food products related to their own home countries. Pizam and Sussmann (1995) suggest French, Italian, Chinese and Japanese tourists are more likely to fall into this category for example, whilst Henderson (2016) discusses the importance of providing halal food for visiting Muslim tourists. Thus exploring the country of origin component to the tourist make-up included in our estimates could provide useful additional information for product shelfing. However, there are those tourists that might be searching for more typical UK products to enrich their experiences of new taste sensations. Everett and Aitchison (2008), Kim et al. (2011) and Getz et al. (2014) all discuss the growth of food tourism and show how different demographic groups may be more likely to indulge in the search for local, regional food. Even though this is likely to be more appropriate for restaurants than grocery stores these types of tourists might be inclined to try traditional UK products available in supermarkets. In the future, detecting the demographic groups more inclined to indulge in food tourism in different localities could be aided by the tourist Twitter activity data by person type (cf. Longley et al 2015).

Second, the incorporation of tourist demand within a spatial modelling framework also enables a more informed evidence base to support robust location planning, performance evaluation and impact assessment of local service development and delivery. For example, in relation to existing store performance, actual versus expected or forecasted store revenues might be reviewed in light of the extra demand available from tourists. We also argue that the new demand layer could be used as an input to modelling tools used to support retailers in estimating new store revenues which are so important to consider before a new store is actually constructed (Birkin et al., 2017). Forecasts of future revenues based purely on residential demand may underestimate the potential of a particular location. This may, in turn, mean the financial case for a new store would not stack up and be rejected by senior management. The addition of tourist demand may not simply give better forecasts – it might take estimated revenues beyond the threshold needed to ensure long-term store viability. In subsequent research we hope to show the value of this new demand layer when introduced into typical store location models.

Third, the estimated tourist demand also has the potential to investigate any latent "tourism food deserts" in our study area. In Figure 5.10, we showed how the distribution of grocery stores is widespread across both inner and outer London regardless of the locations of main tourist attractions, since these stores were predominantly located to serve local residential demand. The distribution of tourist demand plotted against existing stores can reveal a significant mismatch between tourist demand and grocery supply. The gaps, or tourist food deserts, may thus provide retailers with opportunities for (convenience) stores targeted more fully towards tourists. To illustrate this idea we employ some additional spatial analysis here. In Figure 5.13 we identify the areas that have the great differences between estimated tourist demand and grocery floorspace. The darkest shading in Figure 5.13 represents areas which have high estimated tourist demand but low grocery retail floorspace. These areas may provide substantial degrees of sales uplift to newly opened stores or increased store floorspace within existing stores.

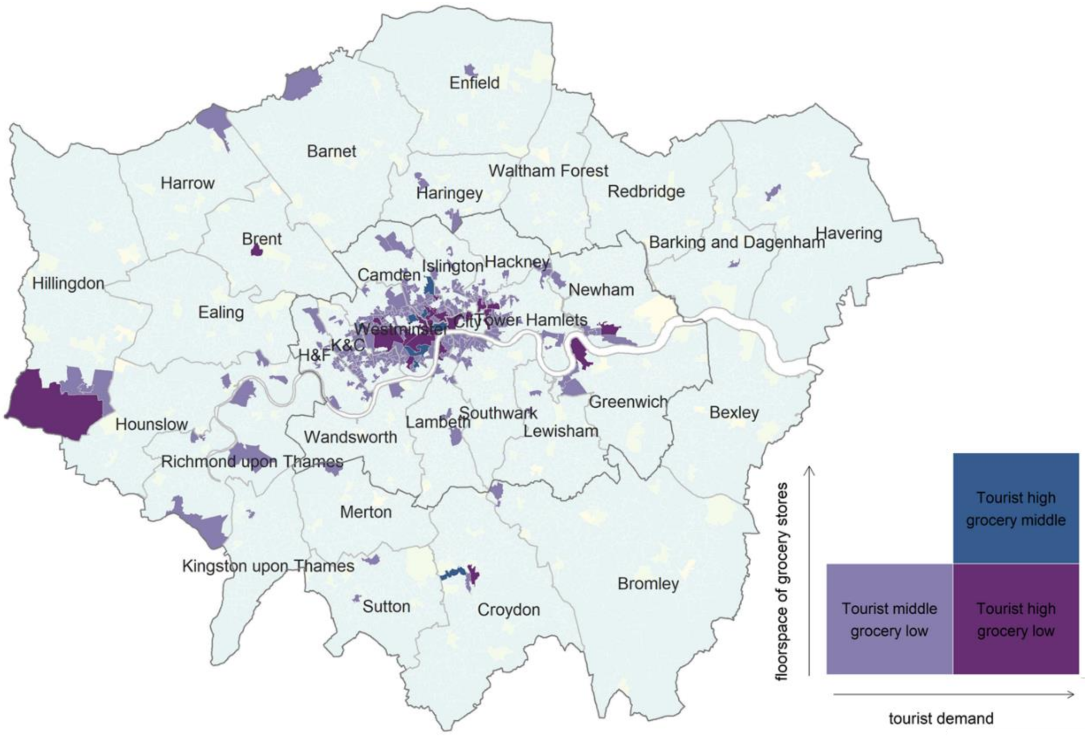


Figure 5.13 Estimated tourist demand versus grocery floorspace in London by LSOA.

In undertaking this research we realise that we have made some assumptions that are difficult to validate in the normal way, especially when trying to attach sensible spending patterns to each tourist group. For data which is 'missing', we have tried to use figures routinely used by internal store location teams (obtained via our retail partner links). We understand that a detailed and fine-tuned tourist spending survey conducted at the local area could be beneficial in future work. In the meantime, there are no existing studies which have reported any small-area demand estimates which can be used to validate the accuracy of some of the estimations we have needed. That said, we have shown that even when using cautious spending estimates from the retail industry, the uplift to grocery spend is very significant. However, we hope in the future to be able to work with a leading partner grocery retailer in London to see if these uplifts make sense when we compare sales between stores which are in tourist areas with those that are not. Exploring store revenue performances in this way will go a long way to helping us to produce even more robust sales uplift forecasts in the future.

5.6 Conclusions

Urban tourism plays an important role in stimulating local economies. Most of the existing research examines the economic impact of tourist shopping via questionnaires to survey the tourist shopping behaviours and expenditures at the macro-level of destinations or associated with specific shopping sites or particular temporary events (Sullivan et al., 2012; Murphy et al., 2011). These studies help shed light on tourist shopping experiences to assist destination management and operation, but contribute little to the understanding of the geographically varied distribution of tourist shopping or the magnitude of these economic drivers at the local, small-area level. The main problem traditionally has been the deficiency of fine-scale population statistics regarding tourists and their behaviours. To overcome this barrier, we have collated data from multiple data sources, for four different types of visitors: overnight tourists staying in Airbnb self-catering accommodation, traditional serviced accommodation, guests staying with relatives or friends and day trip visitors who do not use overnight accommodation. To achieve this task, we have combined conventional data from published industry surveys and statistics with emerging online (big data) sources to estimate small-area tourist numbers in urban areas. The enriched tourist population dataset shows the key patterns of tourist distributions at a finer spatial and temporal granularity, which can be incorporated with residential and workplace statistics to further understand the diurnal and nocturnal population composition in our Greater London study area.

Whilst our primary focus and interest is on tourists' grocery expenditures, our high spatial and temporal granularity estimates of small area tourist distribution, disaggregated by visitor type, could offer tremendous potential for other retail sectors and for tourism destination management and planning. It highlights a detailed example of the benefits and insights that can be gained by combining big data sources which capture sub-groups of tourists and their spatiotemporal distributions at the sub-destination level. Ye et al. (2020) present a broader suite of examples which demonstrate the type of destination management issues that indicators derived from these data could support. In relation to our specific focus on tourist grocery shopping expenditure, our study reveals that this expenditure is considerable and spatially varied according to the degree of tourist concentration.

Coupling the associated grocery spending of each of the four tourist subgroups with the small-area spatial distribution of tourist populations has allowed us to estimate an additional demand layer to the more traditional residential-based estimates used primarily in store location research. The additional tourist demand surface demonstrates a concentration of grocery expenditure uplift around London's Central Activity Zone with a far more dispersed pattern of tourist demand existing in the neighbourhoods of outer London.

Although we have taken London as our example, the devised methodology used here to produce the new small-area tourist retail demand estimates can also be leveraged to other urban destinations which host significant tourist activity. In addition, although we have focused on the grocery shopping market the approach could be utilised for all other retail segments providing that average expenditure rates are available for the appropriate retail sector. We hope other researchers may take up the challenge to explore other cities and retail sectors.

Acknowledgements

This study is supported and funded by the University of Leeds and Chinese Scholarship Council (201709370048).

Appendix

The produced tourist population dataset and the final tourist grocery demand layer are available at https://rpubs.com/ziye/touristdemandestimation.

Declaration of interest

The authors declare that they have no competing interests.

References

Airbnb, 2018. Airbnb UK Insights Report. URL https://www.airbnbcitizen.com/wp-content/uploads/2018/10/AirbnbUKInsightsReport_2018.pdf (accessed 9.13.19).

Backer, E., 2007. VFR travel: An examination of the expenditures of VFR travellers and their hosts. Curr. Issues Tour. 10, 366–377. https://doi.org/10.2167/cit277.0

Barchiesi, D., Moat, H.S., Alis, C., Bishop, S., Preis, T., 2015. Quantifying international travel flows using Flickr. PLoS One 10, 1–8. https://doi.org/10.1371/journal.pone.0128470

Berry, T., Newing, A., Davies, D., Branch, K., 2016. Using workplace population statistics to understand retail store performance. Int. Rev. Retail. Distrib. Consum. Res. 26, 375–395. https://doi.org/10.1080/09593969.2016.1170066

Birkin, M., Clarke, G., Clarke, M., 2017. Retail Location Planning in an Era of Multi-Channel Growth. Routledge, London. https://doi.org/10.4324/9781315605937

Chua, A., Servillo, L., Marcheggiani, E., Moere, A. Vande, 2016. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. Tour. Manag. 57, 295–310. https://doi.org/10.1016/j.tourman.2016.06.013

GBDVS, 2018. The Great Britain Day Visitor 2018 Annual Report [WWW Document]. URL https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/England-documents/260139488_-_kantar_tns_-_gbdvs_2017_annual_report_v5r.pdf (accessed 2.10.20).

GBTS, 2019. Online Data Browser: Great Britain domestic overnight trips (GBTS) [WWW Document]. URL https://gbtsenglandlightviewer.kantar.com/ViewTable.aspx (accessed 1.16.20).

Greater London Authority (GLA), 2017. Projections of demand and supply for visitor accommodation in London to 2050 [WWW Document]. URL https://www.london.gov.uk/sites/default/files/visitor_accommodation_-_working_paper_88.pdf (accessed 12.20.19).

Greater London Authority (GLA), 2016. Central activities zone supplementary planning guidance [WWW Document]. URL https://www.london.gov.uk/sites/default/files/caz_spg_final_v4.pdf (accessed 3.5.20).

Greater London Authority (GLA), 2014. Daytime Population of London 2014 [WWW Document]. Gt. London Auth. URL https://data.london.gov.uk/dataset/daytime-population-borough (accessed 7.19.19).

Gutiérrez, J., García-Palomares, J.C., Romanillos, G., Salas-Olmedo, M.H., 2017. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. Tour. Manag. 62, 278–291. https://doi.org/10.1016/j.tourman.2017.05.003

IPS, 2019. Inbound accommodation research [WWW Document]. URL https://www.visitbritain.org/inbound-accommodation-research (accessed 2.4.20).

Kádár, B., 2014. Measuring tourist activities in cities using geotagged photography. Tour. Geogr. 16, 88–104. https://doi.org/10.1080/14616688.2013.868029

Li, J., Xu, L., Tang, L., Wang, S., Li, L., 2018. Big data in tourism research: A literature review. Tour. Manag. 68, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Living Costs and Food Survey (LCF), 2019. Family spending workbook 3: expenditure by region [WWW Document]. URL https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/datasets/familyspendingworkbook3expenditurebyregion (accessed 2.7.20).

Lovelace, R., Birkin, M., Cross, P., Clarke, M., 2016. From Big Noise to Big Data: Toward the Verification of Large Data sets for Understanding Regional Retail Flows. Geogr. Anal. 48, 59–81. https://doi.org/10.1111/gean.12081

Murphy, L., Moscardo, G., Benckendorff, P., Pearce, P., 2011. Evaluating tourist satisfaction with the retail experience in a typical tourist shopping village. J. Retail. Consum. Serv. 18, 302–310. https://doi.org/10.1016/j.jretconser.2011.02.004

Newing, A., Clarke, G., Clarke, M., 2015. Developing and Applying a Disaggregated Retail Location Model with Extended Retail Demand Estimations. Geogr. Anal. 219–239. https://doi.org/10.1111/gean.12052

Newing, A., Clarke, G., Clarke, M., 2014. Exploring small area demand for grocery retailers in tourist areas. Tour. Econ. 20, 407–427. https://doi.org/10.5367/te.2013.0277

Newing, A., Clarke, G., Clarke, M., 2013a. Identifying seasonal variations in store-level visitor grocery demand. Int. J. Retail Distrib. Manag. 41, 477–492. https://doi.org/10.1108/09590551311330843

Newing, A., Clarke, G., Clarke, M., 2013b. Visitor expenditure estimation for grocery store location planning: a case study of Cornwall. Int. Rev. Retail. Distrib. Consum. Res. 23, 221–244. https://doi.org/10.1080/09593969.2012.759612

Shani, A., Uriely, N., 2012. VFR tourism. The Host Experience. Ann. Tour. Res. 39, 421–440. https://doi.org/10.1016/j.annals.2011.07.003

Shoval, N., McKercher, B., Ng, E., Birenboim, A., 2011. Hotel location and tourist activity in cities. Ann. Tour. Res. 38, 1594–1612. https://doi.org/10.1016/j.annals.2011.02.007

Sullivan, P., Bonn, M.A., Bhardwaj, V., DuPont, A., 2012. Mexican national cross-border shopping: Exploration of retail tourism. J. Retail. Consum. Serv. 19, 596–604. https://doi.org/10.1016/j.jretconser.2012.07.005

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci. Rep. 7, 1–11. https://doi.org/10.1038/s41598-017-18007-4

VisitBritain, 2019. Quarterly Inbound Update Full Year 2018 [WWW Document]. URL https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/2018_uk_and_regional_ips_summary.pdf (accessed 1.20.20).

VisitEngland, 2019. England Occupancy Survey [WWW Document]. URL https://www.visitbritain.org/accommodation-occupancy-archive (accessed 3.1.20).

VisitEngland, 2018. England Domestic Overnight Trips Summary-All Trip Purposes-2018 [WWW Document]. URL https://www.visitbritain.org/gb-tourism-survey-2018-overview (accessed 12.15.20).

VisitEngland, 2016. Accommodation Stock Audit [WWW Document]. URL https://www.visitbritain.org/accommodation-stock (accessed 2.1.20).

VisitLondon, 2020. Visit London Official Visitor Guide [WWW Document]. URL https://www.visitlondon.com/search?category=/things-to-do (accessed 1.5.20).

Waddington, T., Clarke, G., Clarke, M.C., Hood, N., Newing, A., 2019. Accounting for Temporal Demand Variations in Retail Location Models. Geogr. Anal. 426–447. https://doi.org/10.1111/gean.12179

Waddington, T.B.P., Clarke, G., Clarke, M., Newing, A., 2017. Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand. Int. Rev. Retail. Distrib. Consum. Res. 3969, 1–26. https://doi.org/10.1080/09593969.2017.1333966

Ye, Z., Clarke, G., Newing, A., 2020. A review of the potential of Social Media Data for tourism research.

Chapter 6 Adding urban tourism to retail location models (Paper III)

Abstract

The spatial interaction model has proved to be a robust and frequently implemented technique for retail location planning, notably in the grocery sector. Over the years, a number of important extensions to the models have been proposed related to both the demand and supply sides. This paper adds urban tourist demand into the models to improve revenue predictions for both existing and new grocery stores in tourist areas. The new model is built upon tourist grocery demand layers across London and calibrated by origin-destination flows extracted from large-scale Foursquare data. The research shows the benefit of incorporating this additional demand within the retail planning process, highlighting the estimated revenue uplift attributable to tourism in certain stores. It also shows how the model can be used to suggest locations for new investments, especially in areas with insufficient tourist food shopping provision. Investment in these areas may not only bring potential profit to local food retailers but could also improve the tourist experience by adding greater service provision. Finally, by operationalising a number of 'what-if' scenarios drawn from realistic development plans, the work examines the potential revenue of new stores recommended in these areas of insufficient tourist grocery provison.

6.1 Introduction

Although many different methodologies exist to support retail site location research, the spatial interaction model (SIM) has been widely developed in academia and used by many retail organisations across Europe, America, Japan and Australasia (in particular) (Birkin et al., 2017). Over the years, these models have been improved for applied use in the grocery retail sector through the adoption of many types of model disaggregation, especially by person type on the demand side and shop type on the supply side. A major development in relation to the former has been the addition of non-residential demand: in particular, work-based demand and demand from schools, colleges, hospitals etc. (Birkin et al., 2010, 2017; Waddington et al. 2019). Newing et al. (2018) incorporated tourist demand into the models, especially those related to coastal holiday resorts (using a case study of a major UK tourist region - Cornwall). This analysis showed the importance of tourism to store revenues in such regions and highlighted the considerable uplift to store revenues during the peak tourist season. Their analysis also revealed that an understanding of the spatial and temporal variation in underlying tourist demand across the study region was crucially important in predicting store-level impacts. Impacts on revenue were highlighted by the fact that uplift in store revenue, relative to a baseline outside the tourist season, could be as high as 100% for stores in some locations, whilst other stores were shown to be reliant almost solely on residential demand (Newing et al., 2013b). The incorporation of this additional demand directly into the models helped to improve model accuracy, especially when compared to the standard uplift used in the retail industry in such areas (typically around 30% across the entire study region) (Newing, 2013).

To date, there has been little analysis related to tourism uplift for grocery stores in non-coastal regions. This might be deemed especially important in larger urban areas where tourist visits contribute substantial additional local expenditures. If all tourists stayed in central hotels, then the impact on local grocery stores would be minimal, as this type of accommodation is unlikely to have cooking facilities and to encourage the purchases of groceries. However, the tourism landscape in cities is being reshaped by a variety of new tourist trends. Maitland (2019) notes the ongoing spatial expansion of tourism districts in London beyond city cores into more peripheral neighbourhoods, which helps to ease pressure on traditional city centre locations as well as allowing tourists to experience more local (and authentic) urban environments. Such tourism expansion has been enhanced greatly by a series of rapid developments including the growth of tourist rental markets such as Airbnb, to supplement the tourism accommodation stock, sharing bike schemes such as Mobike, and a diversity of geolocated online travel media platforms to help navigate and to provide specialised information about places (Novy, 2018). Among these developments, the short-term rental and sharing economy accommodation has enabled more tourists to stay within self-catering houses and apartments, especially in areas beyond central tourism districts. In addition, new urban marketing strategies have often created new ‘eventscapes’ (i.e. sports, cultural and food events) to encourage day visitor tourists, which themselves bring more retail demand into the city as a whole (Brown et al., 2015).

The aim of this paper is to add urban tourism into classic retail location models in the grocery sector. This will be important for a number of reasons. First, this will allow retailers to have a better understanding of the nature of trade in certain store catchment areas. If it is possible to estimate the percentage of customers who are more likely to be tourists, especially by type and origin of tourist, then product range could also be tailored to serve that market better. Second, it could help to identify the areas where tourists may find it difficult to obtain food products or neighbourhoods that experience significant decrease of grocery provision (per head) when tourist grocery demand is included. Retailers may be encouraged to invest in these areas and urban planners perhaps might be more encouraged to give planning permission for new stores. Third, the addition of tourist demand will lead to better predictions of the revenue of future stores in such areas. This might make a store that seems unprofitable when just examining residential and work-based demand to be economically viable when tourist demand is accounted for, and hence more likely to be approved for development by company executives (and perhaps local planning authorities).

The rest of the paper is structured as follows. After a broader literature review in Section 6.2, Section 6.3 outlines the spatial expansion of tourism from central tourism zones, using the study region of London, UK. It also outlines the estimation of grocery shopping demand disaggregated by customer groups. Section 6.4 details the development and calibration of our revised model. This section also highlights the importance of newly available big data sets to help calibrate the model, given the fact that there is little information on tourist movements from traditional published data sources. In Section 6.5 the outputs of the SIM are used in three ways: first, to estimate revenue uplifts to individual stores. Second, to evaluate tourist grocery accessibility and assess the changes of local grocery provision rates due to the influx of tourist demand; third, to show how the model can be used in a 'what-if' fashion to examine the impacts of new scenarios relating to store development. Finally, some concluding comments are given in Section 6.6.

6.2 Literature review

This section first looks briefly at the evolution of the disaggregated SIM and then sets out the changing nature of urban tourism. SIMs have a long tradition in geography and regional science as a tool to quantify the likely flows of persons (or money) between origins and destinations (Birkin et al., 2017). SIMs have been widely used to support location-based decision making in retail location planning, particularly in the grocery sector (Wood and Reynolds 2011, 2012). Of course, regression and discrete choice models have also been popular (Oppewal and Timmermans 2001; Wood and Browne 2007), but the SIM has been used in many collaborations between academia and business, with many extensions over time (Birkin et al., 2017; Clarke, 2020). It has been successfully applied in store sales estimation, future store performance evaluation and patronage prediction for individual grocery stores. Much recent attention has been paid to incorporating non-residential demand to better forecast the temporal variation of grocery store sales. Waddington et al. (2019) developed a highly disaggregated SIM comprising residential, workplace and demand from schools and universities to include daily fluctuations in each of these demand side drivers of expenditure. Newing et al. (2015) incorporated tourist spending around coastal resorts as additional demand into the residential-based SIM, which has seen success in areas where the localised impacts of short-term tourist-driven demand fluctuation are evident. Other recent studies have developed the model to delineate catchment areas for grocery Click & Collect services (Davies et al., 2019) and retail agglomerations at the UK national level (Dolega et al., 2016). Meanwhile, tools required for model development and more sophisticated model calibration are now available via open-source software such as R and Python, opening up the potential for better fitting and calibration of SIMs (Oshan 2016; Dennett 2018). However, major urban tourist areas may also attract significant visitor-driven expenditure for much of the year, although the dispersal of this expenditure across the urban area may mean that it is less immediately noticeable than in traditional tourist resorts.

As noted in the introduction, the tourism landscape in major destination cities is changing and providing a more complex spatial pattern of demand in urban areas. No longer are tourists concentrated solely in town or city centres. Researchers have examined the economic and socio-spatial impacts on residential neighbourhoods in a series of tourism cities. In particular, Airbnb has been responsible for helping to transform residential neighbourhoods, bringing visitors closer to the day-to-day experience of local residents (Freytag and Bauder 2018; Stors, 2020). Ioannides et al. (2019) suggested that Airbnb takes on a role as the instigator of urban tourism bubble expansion and intensifies the potential conflicts between visitors and locals. Novy (2018) also observed the changing commercial demand in the residential neighbourhoods in Berlin fostered by heightened tourist demand related to Airbnb. Gutiérrez et al (2017) showed the tourist pressure on services in residential areas in Barcelona by analysing the spatial association of Airbnb, hotels and tourist attractions with service provision. Shabrina (2020) argued for greater food service provision for

Airbnb guests as more Airbnb hotspots appear in residential areas. Additionally, Buning and Lulla (2020) and Freytag and Bauder (2018) found that bikeshare services have been increasingly used by visitors to access and explore urban peripheral areas, further mixing visitors and local habitants. While research has started to investigate the social and economic impacts of these new tourism trends, particularly those associated with Airbnb, the quantification of tourist impacts on urban neighbourhoods at the small-area level has not been analysed on a city-wide basis. This paper addresses this issue with specific reference to tourist-driven demand for groceries.

The link between retail location modelling and tourism is currently under-researched. Food is a daily need for tourists during their visits. Traditionally hotel guests, with no cooking facilities, are more likely to eat out, thus increasing revenues of cafes and restaurants rather than grocery stores. However, many Airbnb guests have the opportunity to self-cater and therefore to add food shopping consumption in the areas they reside. Airbnb (2018) reported that their international Airbnb guests in the UK spend on average £10 per person per night (PPPN) on groceries when self-catering. Such grocery demand is also noted by Freytag and Bauder (2018) and Stors (2020), who both evidenced that the ready access to daily shopping at grocery stores is among the top mentioned facilities in Airbnb listing descriptions.

Day visitors are also reported to purchase at food stores around the venues they visit. The Great Britain Day Visitor Survey (GBDVS) (2018) surveyed the expenditure of day trip visitors to discover that on average day trippers spend £11.88 per person on food in shops or takeaways when they travel around London. If as little as half of this expenditure is at grocery stores then the potential expenditure uplift of such stores in residential neighbourhoods could be significant. Meanwhile, visitors staying with friends and relatives also increase the grocery consumption of their host families, and even hotel residents may contribute a small amount of expenditure in local grocery stores (Newing et al., 2013b).

Current research on tourist food shopping consumption and expenditure is largely targeted at understanding the influencing factors of tourist food shopping motivation or its impact on destination experiences (Mak et al. 2012; Mynttinen et al. 2015; Björk and Kauppinen-Räisänen 2019). Tourist food shopping, especially in relation to the local grocery market, has not been reported in an urban context with the exception of recent work by Ye et al. (2021), in which they identified a geographical mismatch between tourist grocery demand and the location of grocery supply across London.

The rest of this paper attempts to bring urban tourism and retail location modelling together. Adding grocery demand driven by urban tourism in major cities has not been considered or incorporated into retail location modelling to date. Similarly, the quantification of the impacts of tourist food shopping expenditures on the local grocery market has not been undertaken at a small area level. To address both challenges, a tourist-based SIM is built next, based on the urban tourist grocery demand layers created by Ye et al. (2021). The model is developed and calibrated to simulate the flows of tourists from their temporary accommodation or key attractions to nearby grocery stores, with the help of novel large-scale individual origin-destination flow data from Foursquare.

6.3 Tourism spatial expansion in London

London is one of the world's leading tourist cities. In 2018, London hosted almost 12 million domestic visitors and over 19 million international tourists, alongside 319.2 million day visitors (VisitBritain 2019; VisitEngland 2018). Tourists are attracted by a unique concentration of tourism attractions generally located in the Centra Activities Zone (CAZ), along with a diversity of cultural, arts, retail, entertainment and night-time economy functions (Great London Authority (GLA), 2021). The CAZ has a constraint on space and there is little capacity for the development of other strategic functions. Hence, the GLA has for decades endeavoured to encourage tourism outside central London, in more outer districts (Maxim, 2017). Encouragingly, the GLA and Creative Tourist Consultants (2015) shows that 58% of the tourists in London are repeat visitors who are more familiar with the capital and more amenable to explore the outside districts (Inkson, 2019). Therefore, the latest London Plan 2021 (the spatial development strategy for London) re-emphasises this vision: "boroughs in outer and inner London beyond the CAZ are encouraged for new serviced accommodation in town centres to help spread the benefits of tourism to the whole of the capital" (GLA, 2021). Against the background of this policy, the growth in sharing economy accommodation in London has helped to facilitate the rapid expansion of tourism into areas outside the conventional 'tourist bubble'. For example, the leading service platform Airbnb, accounting for around a third of the accommodation sector in London, claims that over 72% of their listed properties are located outside the main hotel areas in London (Airbnb, 2018). A series of studies have provided evidence for this shift of tourism in London, dispersing tourists into residential neighbourhoods, suburban green spaces and peripheral shopping sites (Maitland, 2008, 2010, 2013, 2019; Pappalepore et al., 2014; Maxim, 2020). Ye et al. (2021) showed how different subgroups of tourists have penetrated into residential areas of London and we build on this analysis in the sections below.

Ye et al. (2021) identified four major types of tourist groups in London and produced the corresponding tourist demand layers at the Lower Layer Super Output Area (LSOA) level, based on the estimated small-area distribution of tourist populations and their associated grocery expenditures (Table 6.1). The four tourist population groups comprise: 1) visitors using the sharing-economy accommodation platform Airbnb; 2) tourists staying at hotels and other serviced accommodation (i.e. B&Bs, guest houses, motels and hostels); 3) 'free' guests who spend their nights with relatives or friends; 4) day trip visitors who only undertake leisure activity-based trips during the day/evening without using overnight accommodation. As shown in Table 6.1, It is estimated that a total additional grocery expenditure of over £14m per week is generated from both domestic visitors and international tourists. This tourism demand is significant, contributing almost 6% of total grocery demand in London. Given that tourists are expected to exhibit some tendency to cluster spatially, due to clusters of accommodation and key attractions, we argue that overlooking these additional expenditures underestimates tourism's economic impacts at the local level and also results in an underestimation of demand at some grocery stores.

Table 6.1 Total expenditure estimates for demand groups in London.

Demand group	**Expenditure estimate (£m per week)**	**Proportion of total expenditure**
Residential	199.8	83.4%
Workplace	25.8	10.8%
Tourism (total)	14.06	5.9%
Airbnb	4.72	
Hotel and serviced accommodation	1.35	
Free guest/own home	3.70	

Day trip visitors	4.29	

6.4 Modelling methodology

6.4.1 Model formulation

The small area estimation of tourist grocery demand produced in Ye et al. (2021) is used as the essential demand input of the new tourism SIM. Following the classic production-constrained SIM, our tourism model version can be represented as:

$$S_{ij}^{t} = \sum_{t} A_{i}^{t} O_{i}^{t} W_{j}^{\alpha} exp^{(-\beta^{t} d_{ij})} \quad (6.1)$$

Where:

S_{ij}^{t} is the predicted expenditure flow between LSOA i and store j by tourist type t (1: Airbnb guest; 2: serviced accommodation tourist; 3: tourist staying as a free guest with friends and relatives; 4: day trip visitor);

O_{i}^{t} is the total tourist demand in LSOA i by tourist type t as estimated in Ye et al., (2021);

W_{j} is the overall attractiveness of store j, where α represents the additional or perceived relative attractiveness of store j.

d_{ij} is the distance between LSOA i and store j, expressed as $exp^{-\beta^{t}}$ for tourist type t.

A_{i}^{t} is a balancing factor which takes account of competition and ensures that all demand from LSOA i by tourist type t is allocated to stores within London. The balancing factor thus ensures that:

$\sum_{j} S_{ij}^{t} = O_{i}^{t}$, which is calculated as: $A_{i}^{t} = \frac{1}{\sum_{j} W_{j}^{\alpha} exp^{(-\beta^{t} d_{ij})}}$. (6.2)

To ensure complete coverage of the demand side, residential and workplace grocery demand layers were also generated at the LSOA level and incorporated within the models, although they act only as a baseline and are not the focus of our discussion in this paper. The development and calibration of the non-tourist SIM in London are detailed in Appendix C.

6.4.2 Model calibration

In common with the application of SIMs in the grocery sector, the model has two calibration parameters, α and β, which reflect brand attractiveness and distance decay respectively. In previous model calibrations of the residential and workplace SIMs, the attractiveness parameter α usually adopts different values for each retailer, calibrated by retailer's market shares in the region, to capture the variations in brand loyalty and the relative attractiveness of different brands to consumers (Newing et al., 2015; Waddington et al., 2019). However, for tourist customers, such brand preferences may not be that significant. On the one hand, tourists usually are not always familiar with the grocery supply side in (foreign) destinations and also, it is hard to simulate their brand preferences without a comprehensive tourist survey. Therefore the α value in the model is set at a value of 1 (no additional brand attractiveness) for all brands for our tourist consumers.

Thus, the main calibration that needs to be undertaken is the distance deterrence parameter β, which is set independently for our residential, workplace and tourist layers to account for differences in personal mobility and the relative importance of distance decay to each consumer type. In common with similar model-building exercises (Thompson et al. 2012; Newing et al., 2015), we use an indicator of average trip distance (ATD) by small-area geography to obtain realistic distance-decay parameters.

As detailed in Appendix C, the demand side of the model is disaggregated by household type according to the Output Area Classification (OAC) in London, and for each household type k, the ATD is calculated as the mean distance from each demand zone to the nearest three supermarkets. Thus, β^k in the residential SIM is calibrated utilising these indicative observed ATDs as presented in Table A.4. Similarly, the β^k for the workplace SIM is calibrated according to the average trip length (walking) for eat/drink purpose which is 0.965 km (0.6 miles) in England as reported by the (National Travel Survey (NTS), 2017). In contrast, a major issue in this paper is the calibration of β for the tourist SIM. Since tourists are not well-covered by traditional data sources used for model calibration, such as loyalty cards or large-scale consumer surveys, this study harnesses rich location-based social networks (LBSN) data from the geolocation networking service Foursquare to infer mobility patterns associated with tourists in relation to grocery shopping in London, as outlined below.

Foursquare check-in data have been adopted in previous urban mobility analyses, given its volume of data and ability to capture many different human activities (Noulas et al. 2012). Unlike other geotagged social media data, Foursquare check-in data can only be generated when users are physically present at venues for a certain amount of time, and thus is explicitly associated with actual venues involving consumption activities and other Points of Interest (POI) (Martí et al., 2017; Salas-Olmedo et al., 2018). Therefore, researchers have used Foursquare data to measure venue popularity and interactions between different venues to simulate retail activity at a fine spatial scale (Karamshuk et al., 2013; D'Silva et al., 2018; Doan and Lim, 2019). For example, Piovani et al. (2016) leveraged Foursquare data to validate their proposed retail location choice model and found a high correlation between Foursquare-estimated pedestrian flows and retail turnover of the neighbouring outlets.

The Foursquare data used here consists of two different data sets. First, a venue dataset can be built up capturing urban POIs categorised into a well-structured venue classification schema, which in turn suggests the human activities at the location. Second, movement datasets consisting of origin-destination pairs and the time the movement occurred, offer spatiotemporal information of users' movement between venues. The Foursquare venue and movement datasets used in this research were retrieved via Foursquare Labs Inc. (in May 2019), which contains over 7.6 million origin-destination interactions among 22,689 POIs in London. We identified 726 mentioned grocery outlets and an associated set of 226,800 origin-destination flows from the datasets. Specifically, the flows associated with trips from hotels, residential buildings and tourist attractions were selected to calibrate the travel behaviour of hotel stayers, Airbnb guests and day visitors respectively. For the free guest/own home tourist who stays with residents in London, we deem that their food shopping movements are in line with the hosts and thus follows the ATD used in the residential SIM (see Appendix C). Therefore, the observed ATD for grocery shopping was determined by tourist type, accounting for the distance between accommodation/attraction and grocery store used, as shown in Table 6.2. The parameter β has been set to minimise the difference between observed ATD and the corresponding model derived ATD (by tourist type) within our SIM.

Table 6.2 Beta values for the disaggregate tourist model.

Demand	$\boldsymbol{\beta}$	Predicted ATD (km)	Observed ATD (km)	ATD pred./ATD ob.
Airbnb travellers	7.543	0.5520	0.552	0.9999
Hotel guests	5.796	0.5260	0.526	0.9999
Free guest/Own	0.952	2.0941	2.0934	1.0003

home				
Day visitor	1.586	1.353	1.353	1.0001

For those familiar with calibrating β within SIMs the values for Airbnb and hotel might look exceptionally high – but these values are needed to stop any person travelling outside the very tight catchment area seen in the observed ATD. Whilst the calibration process could be improved further with the additional survey or transaction data, the large-scale foursquare data permits model calibration in relation to inferred tourist shopping trip-making behaviours, enabling us to demonstrate the potential utility of a tourist SIM in supporting business and policy decision making, as explored in the following section.

6.5 Model results

Section 6.4 outlined the model development and calibration process which generated our custom-built SIM - a model containing interactions between 4,835 LSOAs and 1,759 grocery stores in London. The estimated flows are disaggregated by consumer type of residents, workers and four sub-divided tourist groups as outlined in Table 6.2. The calibration routine utilised appropriate indicators of trip-making behaviours by demand type and additionally captured brand attractiveness for residential and workplace populations as presented in Appendix C. In the following subsections we compare the tourist-included SIM to the non-tourist SIM to highlight the impact of tourist demand on store level revenue uplift and local grocery provision rates. We consider the change of grocery service provision ratio and also model the impacts of new store openings. Our specific focus is on the impacts of new store openings on tourist accessibility to grocery shopping opportunities and the contribution of different types of tourist demand to store level revenues.

6.5.1 Model revenue uplifts

Given the spatial variations in tourist demand, most stores receive some revenue uplift from tourists, but the majority (75%) receive less than 10% revenue uplift (which can still be a significant amount). However, 10% of stores are estimated to enjoy a significant uplift (20% or more) whilst a small number (2-3%) have an uplift of over 50% (mostly stores in inner London). Figure 6.1 shows the spatial distribution of revenue uplift.

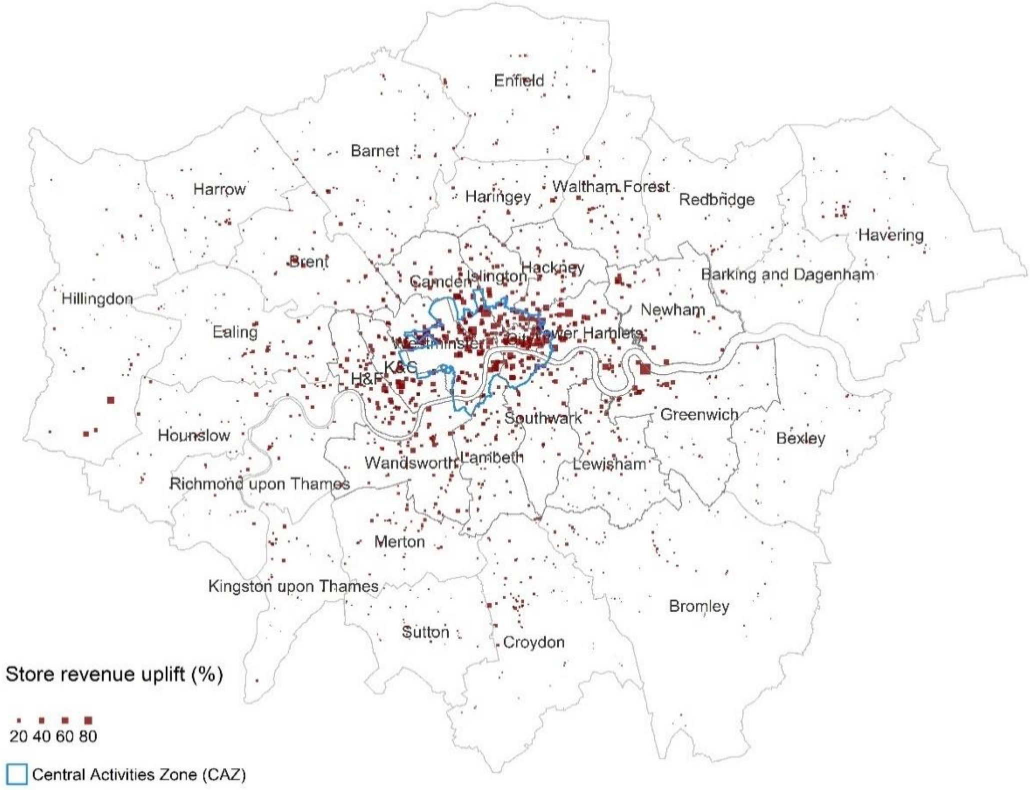


Figure 6.1 The spatial distribution of sales uplift at a store level.

A selected list of store revenues for different retailers is presented in Table 6.3. These stores experience substantial uplifts of revenue due to strong local tourist demand. It is also noticeable that the stores close to transport hubs usually have significant sales uplift as a large amount of tourist demand passes through these hubs. The list also includes several stores that are located outside the CAZ but which benefit from considerable sales uplift driven by tourist demand. Closer inspection reveals that in many cases this is driven by concentrations of Airbnb properties (such as within Earl's Court and West Kensington), or clusters of tourists associated with peripheral attractions (e.g. Richmond or Stratford).

Table 6.3 Selected store performance before vs. after tourist demand.

Brand	Postcode	Location	Floorspace (sqft)	Uplift (%)
Co-Op	WC1H 8BD	Kings Cross	2,130	78.7
Co-Op	SE10 0ER	The O2	2,130	77.8
Others	W2 1RH	Paddington Station	505	66.2

Sainsburys	E1 6HT	Shoreditch High Street	2,130	60.4
Co-Op	WC1E 7DB	The British Museum	1,800	57.4
Co-Op	SW5 9QB	Earl's Court Station	2,790	51.8
M&S	W2 1HA	Paddington station	2,345	47.5
Sainsburys	W2 1HB	Paddington station	2,691	45.2
Tesco	E1 6NF	Commercial St, Spitalfields	2,090	44.3
Waitrose	E20 1EH	Stratford	2,130	23.84
Co-op	W14 9EX	West Kensington	1,683	28.90
Lidl	W12 8PP	Westfield	7,586	16.67
Sainsbury’s	TW10 6NQ	Richmond	2,130	13.48
Tesco	N17 9NE	Tottenham	2,000	9.73

6.5.2 Tourist grocery accessibility and changes of provision

The outputs of SIMs can be specified as performance indicators which can measure accessibility (Clarke and Wilson 1994; Clarke et al., 2002). Unlike typical accessibility measures which examine the size and location of facilities against distance travelled (Hansen, 1959), SIM-derived indicators also consider the likely interactions between facilities and demand. By recalculating the share of floorspace according to the interaction flows, a SIM-derived Hansen style accessibility indicator is used to quantify the provision of grocery retail opportunities available to the customers of a nearby origin. The aggregate level of provision in an LSOA (w_i) can be measured as

$$w_i = \sum_j \frac{S_{ij}^t}{S_{*j}^t} W_j \tag{6.3}$$

The level of provision per tourist capita by LSOA (v_i) and the count of tourists estimated (T_i) by Ye et al. (2021) are utilised to account for accommodation capacity and utilisation to identify the areas which have poor grocery accessibility for tourists, as follows:

$$v_i = \frac{w_i}{T_i} \tag{6.4}$$

Figure 6.2 shows the level of provision per tourist over the LSOAs in London. It is clear that the CAZ and its surrounding boroughs have relatively lower levels of grocery store rates of provision, but these areas offer abundant other types of food and beverage services to mitigate this poor accessibility to grocery stores. Other locations of low grocery supply relative to estimated demand appear in proximity to Heathrow Airport and large sports stadiums, the latter principally associated with day visitors. The areas beyond CAZ with low grocery provision rates compared to tourist locations can be considered as areas of insufficient grocery provision for tourist, and they become areas of opportunity for additional grocery stores. The model is used to explore various new store opportunities in the next section.

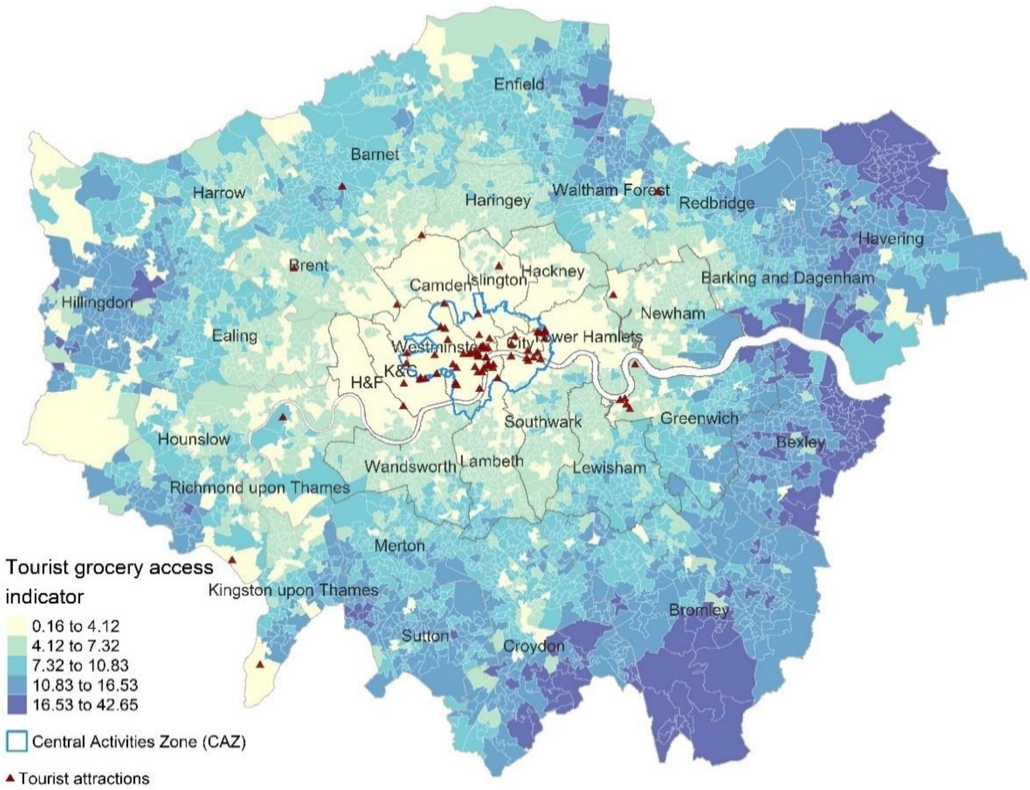


Figure 6.2 Interaction-based indicator of the grocery accessibility to tourist customers.

Figure 6.3 illustrates the differences of grocery provision in the LSOAs over London, when comparing the level of provision per person capita (v_i) derived from the non-tourist SIM and tourist-incorporated SIM. Besides the aforementioned CAZ and airport district, we see a strong decline of grocery provision rates in the inner borough areas outside the CAZ and the fringe areas of the inner boroughs.

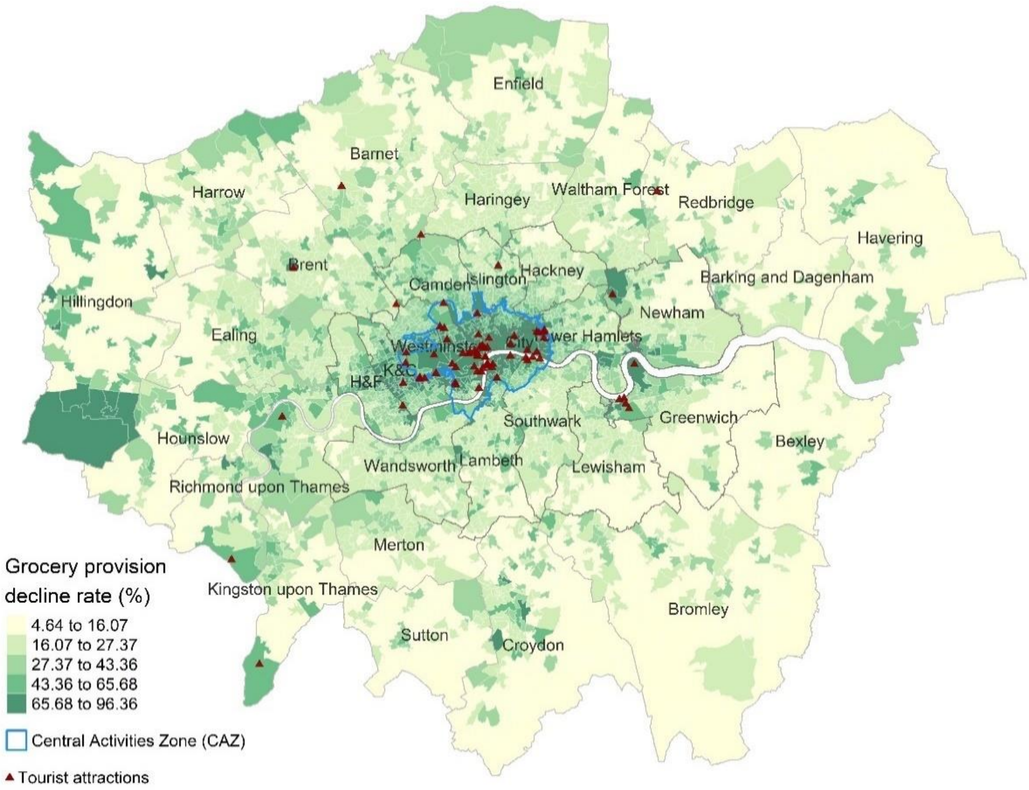


Figure 6.3 Changes of grocery provision ratio after incorporate tourist demand in the model.

6.5.3 'What-if' modelling

In this section, we examine a series of 'what-if?' scenarios related to new store openings. We have chosen three new store openings in areas identified to be insufficient tourist grocery provision (Figure 6.2). Two of the new store developments are based on actual retail provision plans. We use these scenarios to illustrate the importance of incorporating tourism demand within the spatial modelling process and to highlight the potential contribution of tourist expenditures to new store revenues.

The three case scenarios are described below:

a. *A new store in an area of major housing/tourist accommodation growth.* Our first case study is located in the borough of Tower Hamlets, to the east of the CAZ. Although not a traditional hotel district, Tower Hamlets has experienced a rapid growth of tourist accommodation. From 2016 to 2019, the number of properties listed on Airbnb in Tower Hamlets more than doubled (from 3,916 to 8,436 properties). In a report for the London Borough of Tower Hamlets, it is stated that up to 9,085 square metres of new food store floorspace could be supported in the area by 2031 (Carter Jonas, 2016). Also, in

June 2016 one supermarket brand acquired a 15,500 sqft property in Whitechapel Town Centre (a town within tower Hamlets) (Carter Jonas, 2016). Therefore, we use our SIM to predict the performance of the proposed new store in this area and offer insight into the potential customer composition of the new store.

b. *A new store to provide convenience retail opportunities in an area of considerable commercial development.* The second case scenario is the Paddington Opportunity Area (in the western part of the CAZ), as noted in the Westminster City Plan 2019 – 2040 (City of Westminster, 2019). The Plan supports our analysis in suggesting that an increase of retail floorspace is needed to not only meet new resident's needs but also to serve more visitors, tourists and workers. Hence, we add a supermarket of 15,000 sqft to explore how it serves the local customer demand in the proposed areas and how the tourist demand is better met after the new development.

c. *A new store in an area identified to have limited retail provision when considering the tourists in the catchment.* Royal Dock is chosen as our last case scenario, which is one of the most prominent areas in London with low food shopping provision as identified in Figure 6.2a. Currently, Royal dock and its neighbourhoods have no large-format grocery store and only two small size convenience stores (less than 2,013 sqft) within a 1km buffer. We experiment with a new store of 5,000 sqft in the Royal Dock area and examine its performance and impact on the existing stores close by.

The model is rerun with these additional stores added. The level of provision per tourist in the 1km catchment area of pre- and post- new stores opening are compared in Figure 6.4. The Whitechapel and Paddington cases demonstrate a significant improvement in tourist provision rates across the whole catchment area, particularly around the immediate vicinity of the new store sites. The impact of the Royal Dock new store catchment is not as obvious, perhaps unsurprising given the relatively small (5,000 sqft) store proposed here. However, the ATD by tourists in the Royal Dock area is observed to decrease from 2.15km to 1.55 km under this scenario, considerably improving accessibility within this area.

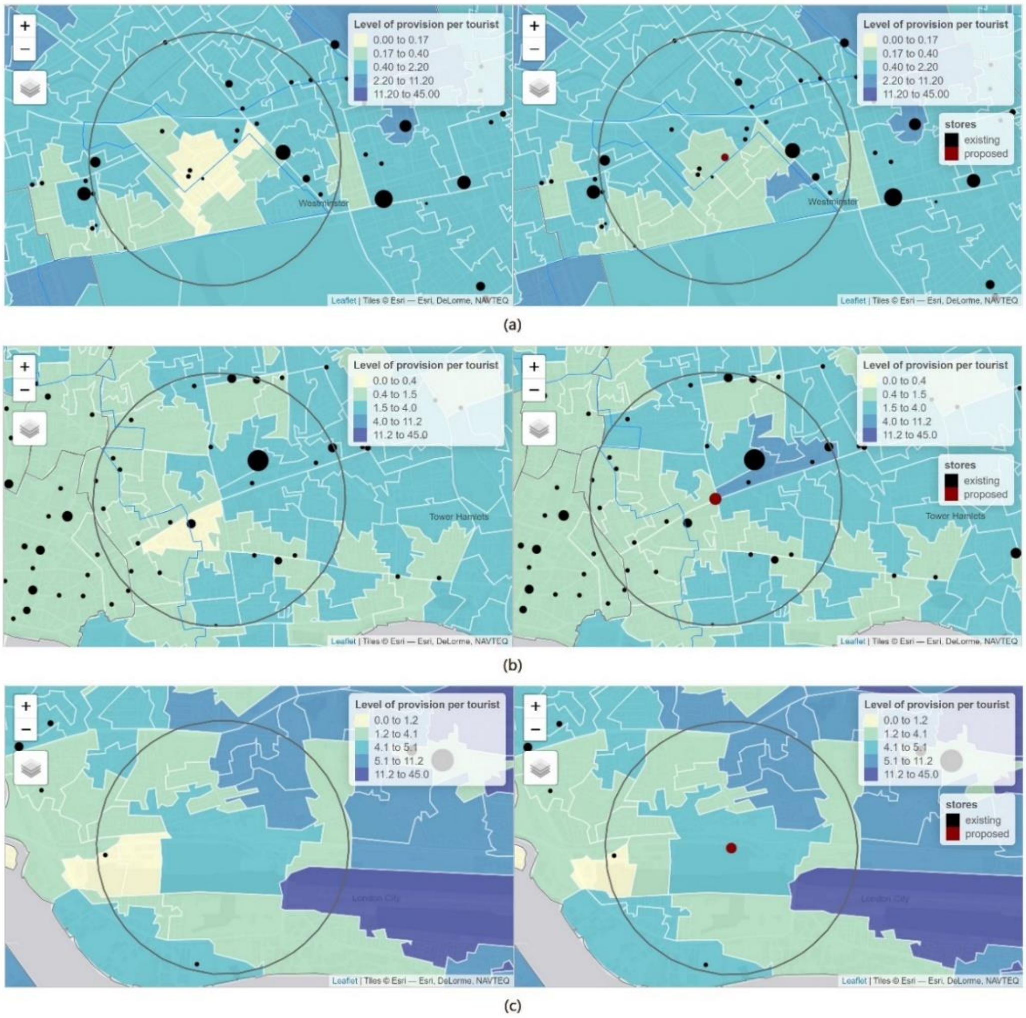


Figure 6.4 Level of provision per tourist change of pre- and post- new store opening in all the three scenarios: (a) Whitechapel Town Centre; (b) Paddington Opportunity Area; (c) Royal Dock.

Table 6.4 presents the store performance of the three proposed stores, based on modelled estimates from our SIM, and the proportion of each demand type accounted for by our modelled flows. The trading intensity – a measure of store efficiency - of the three stores is much higher than the benchmark sales density according to one of the UK's top grocery brands (£17.10 per sqft (Tesco PLC, 2020)), especially for the Paddington store and Royal Dock store which are each modelled to achieve a trading intensity of over £35 per sqft. The high trading intensity indicates a promising store performance for the three proposed stores.

The split of revenue by demand type shown in Table 6.4 can be useful for in-store operations and local marketing campaigns to meet the needs of different types of demand in the catchment areas. The residential demand is the dominant contributor to all three new stores, as we would still expect, but the results also suggest that tourist demand accounts for 11.6% of revenues at the Whitechapel store, 18.9% at the Paddington store and 12.7% at the Royal Dock store. The tourist compositions of each area are very different. Half the tourists in the Whitechapel store catchment are Airbnb guests who are able to self-cater during their stay and thus are likely to have higher demand for fresh food. The store also attracts a quarter of their tourist sales from day visitors who tend to purchase snacks or ‘food-on-the-go’. In contrast, the Royal Dock store is estimated to attract a large volume of day visitors, but a lower proportion of Airbnb guests. The tourist sales from the Paddington store are mainly from overnight tourists staying in both hotel and Airbnb (summed up to over 87%) in the immediate vicinity of that store. Hence, the shelfing and product range of the three stores could be different when targeting the different types of tourist customers.

Table 6.4. Weekly store revenue (£) and trade intensity (£ per sqft per week) estimates and composition.

Store location	Floorspace (sqft)	Total revenue (£)	Trade intensity (£/sqft)	Non-tourist (%)		Tourist only (%)			
				Resident	Worker	Airbnb	Hotel	Free guest	Day visitor
Whitechapel Town Centre	15,500	368,762	23.79	88.4%		11.6%			
				70.5%	17.9%	51.3%	11.7%	11.5%	25.5%
Paddington Opportunity Area	15,000	525,299	35.02	81.1%		18.9%			
				65.3%	15.9%	38.2%	49.0%	6.8%	6.1%
Royal Dock	5,000	190,503	38.1	87.3%		12.71%			
				80.8%	6.5%	24.0%	21.7%	5.5%	48.8%

The opening of new stores inevitably impacts revenues at nearby stores. The deflections which result from these new stores are calculated by comparing the revenues of existing stores within a 1km buffer of the new stores as summarised in Table 6.5. There are 22 existing stores in the 1km buffer of the new Whitechapel store, and on average an 8.6% deflection is caused by the new store, but the revenue from tourists at these stores decreases less (6%), which suggests that the new store absorbs more non-tourist demand than tourist demand from the existing stores close by. The other two proposed stores lead to considerable revenue decreases of surrounding stores and draw in a greater proportion of tourist expenditure since their tourist revenue deflections are higher. Although a rather small-size store, the proposed Royal Dock store results in an average 24.2% decrease for the two existing nearby stores. There is an estimated excellent store performance for the Royal Dock new store – as in Table 6.4 the trading intensity is as high as £38.1 per sqft.

Table 6.5. Number of existing stores close by and the mean percentage of deflections.

Store location	**Number of stores in 1km buffer**	**Total revenue deflection (%)**	**Tourist revenue deflection (%)**
Whitechapel Town Centre	22	8.6	6.0
Paddington Opportunity Area	16	18.3	23.1
Royal Dock	2	24.2	26.9

The 'what-if' analysis shows that in all three scenarios, new stores could enhance tourist provision rates in the areas currently deficient in food shopping supply and provide more accessible grocery shopping opportunities for tourists. After incorporating tourist demand alongside residential and workplace demand, it is possible in this new model to estimate not only the new store revenues and the uplift from tourist demand but also the disaggregation of store revenue by demand type. The in-depth insights of disaggregated demand types could enable the retailers in urban tourism destinations to prepare and make better informed strategic and operational decisions about the new stores and to maximise sales by tailoring the store product range to meet the different requirements of different types of tourist demand.

6.6 Conclusions

Cities are experiencing a spatial expansion and redistribution of urban tourists. The prevalence of self-catering accommodation such as Airbnb and other new tourism trends foster this development – tourists are now venturing into conventional residential neighbourhoods, increasingly sharing local services which have traditionally served only a local residential customer base. Previous research has recognised that in host urban communities of advanced economies like London, tourism now coexists and overlaps with other processes of consumption and production of urban space (Maitland, 2013; Cocola-gant, 2018). It has been suggested that more research is needed for understanding the new demand from urban tourism and its impacts on local residential provision rates (Guttentag, 2015).

This research attempts to bridge urban tourism research with retail location planning to investigate the impacts of tourist grocery shopping demand on the local grocery market. Using London as an example, the paper has highlighted the existing geographical gaps between high tourist demand areas and the local food shopping supply, particularly in peripheral boroughs. A new SIM has been built to model interactions between the tourist demand side origins (alongside underlying residential and workplace demand) and the grocery store network. Given the lack of open survey data on tourist travel behaviour, this research has utilised the large volume of Foursquare data to investigate tourist travel patterns, helping to calibrate the nature and magnitude of distance decay of tourist travel movements by tourist group. We also demonstrated, via 'what-if' analysis, that new-store performance evaluation should account for the impacts of tourist demand, offering better insights as to the likely customer composition for each new store, and assessing the impacts on the performance of stores in immediate proximity.

The research demonstrates the importance of taking tourist demand into account when conducting retail location analysis in urban destinations. The individual store product selection and local marketing campaigns might be more focused around the accommodation stock in the area. The model could allow the store location planning team to estimate the potential sales from each subgroup. This could provide a significant tool to use for in-store product preparation and management. The research shows substantial uplift of store revenue inflows in areas where tourist demand is high, and this extra demand should not be ignored. We believe the addition of urban tourism into retail location models is another important contribution to improving the usefulness of these models in practical applications.

Future research could attempt to improve the current model calibration process and validate the model outcomes, but both would require better empirical data. GPS tracking records and mobile phone roaming data have acted as useful sources to unravel tourist travel behaviour among destination locations. We believe those data harvested in urban destinations might inform more accurate distance decay parameters when modelling in the same areas. Also, any data from industry sources would be worthwhile obtaining, particularly to help validate whether the estimated store sales uplifts from tourists are in line with their estimations (if possible) of consumer type differentiation.

Acknowledgement

This study is supported and funded by the Chinese Scholarship Council (No. 201709370048) and the University of Leeds.

References

Airbnb 2018. Airbnb UK Insights Report. , pp.1–56. [Accessed 13 September 2019]. Available from: https://www.airbnbcitizen.com/wp-content/uploads/2018/10/AirbnbUKInsightsReport_2018.pdf.

Birkin, M., Clarke, G. and Clarke, M. 2010. Refining and operationalizing entropy-maximizing models for business applications. Geographical Analysis. 42(4), pp.422–445.

Birkin, M., Clarke, G.P. and Clarke, M. 2017. Retail Location Planning in an Era of Multi-Channel Growth [Online]. London: Routledge. Available from: https://www.taylorfrancis.com/books/9781317064541.

Björk, P. and Kauppinen-Räisänen, H. 2019. Destination foodscape: A stage for travelers' food experience. Tourism Management. 71, pp.466–475.

Brown, G., Lee, I.S., King, K. and Shipway, R. 2015. Eventscapes and the creation of event legacies. Annals of Leisure Research. 18(4), pp.510–527.

Buning, R.J. and Lulla, V. 2020. Visitor bikeshare usage: tracking visitor spatiotemporal behavior using big data. Journal of Sustainable Tourism. 29(4), pp.711–731.

Carter Jonas 2016. Town Centre Retail Capacity Study 2016: Final Draft Report [Online]. [Accessed 5 September 2020]. Available from: https://www.towerhamlets.gov.uk/Documents/Planning-and-building-control/Strategic-Planning/Local-Plan/Town_Centre_Retail_Capacity_Study_2016.pdf.

City of Westminster 2019. City Plan 2019 - 2040: Convenience Retail Evidence Topic Paper [Online]. [Accessed 5 September 2020]. Available from:

https://www.westminster.gov.uk/sites/default/files/ev_e_002_economy_and_employment_topic_paper_wcc_november_2019_0.pdf.

Clarke, G., Eyre, H. and Guy, C. 2002. Deriving indicators of access to food retail provision in British cities: Studies of Cardiff, Leeds and Bradford. Urban Studies. 39(11), pp.2041–2060.

Clarke, G.P. 2020. Regional science in business In: M. M. Fischer and P. Nijkamp, eds. Handbook of Regional Science. Berlin, Heidelberg: Springer, pp.129–139.

Clarke, G.P. and Wilson, A.G. 1994. A new geography of performance indicators for urban planning In: C. S. Bertuglia, G. P. Clarke and A. G. Wilson, eds. Modelling the City: Performance, Policy and Planning. London: Routledge, pp.55–81.

Cocola-gant, A. 2018. Struggling with the leisure class: Tourism, gentrification and displacement.[Online] Available from: http://e-journal.uajy.ac.id/14649/1/JURNAL.pdf.

D'Silva, K., Kasthuri, J., Noulas, A., Mascolo, C. and Misra, A. 2018. The Role of Urban Mobility in Retail Business Survival. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2(3), p.100.

Davies, A., Dolega, L. and Arribas-Bel, D. 2019. Buy online collect in-store: exploring grocery click&collect using a national case study. International Journal of Retail and Distribution Management. 47(3), pp.278–291.

Dennett, A. 2018. Modelling Population Flows Using Spatial Interaction Models. Australian Population Studies. 2(2), pp.33–58.

Doan, T.-N. and Lim, E.-P. 2019. Modeling location-based social network data with area attraction and neighborhood competition. Data Mining and Knowledge Discovery. 33(1), pp.58–95.

Dolega, L., Pavlis, M. and Singleton, A.D. 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. Journal of Retailing and Consumer Services. 28, pp.78–90.

Freytag, T. and Bauder, M. 2018. Bottom-up touristification and urban transformations in Paris. Tourism Geographies. 20(3), pp.443–460.

Great Britain Day Visitor Survey (GBDVS) 2018. The Great Britain Day Visitor 2018 Annual Report. , p.254. [Accessed 10 February 2020]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-

Library/documents/England-documents/260139488_-_kantar_tns_-_gbdvs_2017_annual_report_v5r.pdf.

Great London Authority (GLA) 2021. The London Plan [Online]. [Accessed 20 March 2021]. Available from: www.london.gov.uk.

Greater London Authority (GLA) and Creative Tourist Consultants 2015. Take A Closer Look: A Cultural Tourism Vision for London, 2015-2017 [Online]. Available from: https://www.london.gov.uk/sites/default/files/cultural_tourism_vision_for_london_low_res_version.pdf.

Gutiérrez, J., García-Palomares, J.C., Romanillos, G. and Salas-Olmedo, M.H. 2017. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. Tourism Management. 62, pp.278–291.

Guttentag, D. 2015. Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. Current Issues in Tourism. 18(12), pp.1192–1217.

Hansen, W.G. 1959. How Accessibility Shapes Land Use. Journal of the American Planning Association. 25(2), pp.73–76.

Inkson, C. 2019. Unplanned Expansions: Renting Private Homes to Tourists In: A. Smith and A. Graham, eds. Destination London: The Expansion of the Visitor Economy [Online]. University of Westminster Press, pp.37–59. Available from: https://www.jstor.org/stable/j.ctvhrd0t9.6.

Ioannides, D., Röslmaier, M. and van der Zee, E. 2019. Airbnb as an instigator of ‘tourism bubble’ expansion in Utrecht’s Lombok neighbourhood. Tourism Geographies. 21(5), pp.822–840.

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]. New York: ACM Press, p.793. Available from: http://arxiv.org/abs/1306.1704%0Ahttp://dx.doi.org/10.1145/2487575.2487616.

Maitland, R. 2013. Backstage Behaviour in the Global City: Tourists and the Search for the ‘Real London’. Procedia - Social and Behavioral Sciences. 105, pp.12–19.

Maitland, R. 2008. Conviviality and everyday life: the appeal of new areas of London for visitors. International Journal of Tourism Research. 10(1), pp.15–25.

Maitland, R. 2010. Everyday life as a creative experience in cities. International Journal of Culture, Tourism and Hospitality Research. 4(3), pp.176–185.

Maitland, R. 2019. Extending the Frontiers of City Tourism: Suburbs and the Real London In: A. Smith and A. Graham, eds. Destination London: The Expansion of the Visitor Economy [Online]. University of Westminster Press, pp.15–35. Available from: https://www.jstor.org/stable/j.ctvhrd0t9.5.

Mak, A.H.N., Lumbers, M., Eves, A. and Chang, R.C.Y. 2012. Factors influencing tourist food consumption. International Journal of Hospitality Management. 31(3), pp.928–936.

Martí, P., Serrano-Estrada, L. and Nolasco-Cirugeda, A. 2017. Using locative social media and urban cartographies to identify and locate successful urban plazas. Cities. 64, pp.66–78.

Maxim, C. 2017. Challenges faced by world tourism cities – London's perspective. Current Issues in Tourism., pp.1–20.

Maxim, C. 2020. Challenges of World Tourism Cities: London, Singapore and Dubai In: A. M. Morrison and J. A. Coca-Stefaniak, eds. Routledge Handbook of Tourism Cities [Online]. London: Routledge, ? Available from: https://www.taylorfrancis.com/books/9780429521331.

Mynttinen, S., Logrén, J., Särkkä-Tirkkonen, M. and Rautiainen, T. 2015. Perceptions of food and its locality among Russian tourists in the South Savo region of Finland. Tourism Management. 48, pp.455–466.

National Travel Survey 2017. Average trip length by main mode, for eat/drink purposes only: England, 2002 to 2017. Department for Transport statistics. [Online]. [Accessed 16 July 2020]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/858316/ntsq03006.ods.

Newing, A. 2013. Incorporating seasonal visitor demand in retail location modelling. University of Leeds.

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

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

Newing, A., Clarke, G.P. and Clarke, M. 2013. Visitor expenditure estimation for grocery store location planning: a case study of Cornwall. The International Review of Retail, Distribution and Consumer Research. 23(3), pp.221–244.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M. and Mascolo, C. 2012. A Tale of Many Cities: Universal Patterns in Human Urban Mobility J. A. Añel, ed. PLoS ONE. 7(5), p.e37027.

Novy, J. 2018. 'Destination' Berlin revisited. From (new) tourism towards a pentagon of mobility and place consumption. Tourism Geographies. 20(3), pp.418–442.

Oppewal, H. and Timmermans, H. 2001. Discrete Choice Modeling: Basic Principles and Application to Parking Policy Assessment In: G. Clarke and M. Madden, eds. Regional Science in Business. Berlin: Springer-Verlag, pp.97–114.

Oshan, T. 2016. A primer for working with the spatial interaction modeling (SpInt) module in the python spatial analysis library (PySAL). Region. 3(2), pp.R11–R23.

Pappalepore, I., Maitland, R. and Smith, A. 2014. Prosuming creative urban areas. Evidence from East London. Annals of Tourism Research. 44(1), pp.227–240.

Piovani, D., Zachariadis, V. and Batty, M. 2016. Quantifying Retail Agglomeration using Diverse Spatial Data. Scientific Reports. 7(1), pp.1–8.

Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C. and Gutiérrez, J. 2018. Tourists' digital footprint in cities: Comparing Big Data sources. Tourism Management. 66, pp.13–25.

Shabrina, Z. 2020. The impact of the platform economy in cities: the case of Airbnb.

Stors, N. 2020. Constructing new urban tourism space through Airbnb. Tourism Geographies., pp.1–24.

Tesco PLC 2020. Tesco PLC Annual Report and Financial Statements. [Accessed 11 June 2020]. Available from:

https://www.tescoplc.com/investors/reports-results-and-presentations/annual-report-2020/.

Thompson, C., Clarke, G.P., Clarke, M. and Stillwell, J. 2012. Modelling the future opportunities for deep discount food retailing in the UK. International Review of Retail, Distribution and Consumer Research. 22(2), pp.143–170.

VisitBritain 2019. Quarterly Inbound Update Full Year 2018. [Accessed 20 January 2020]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/2018_uk_and_regional_ips_summary.pdf.

VisitEngland 2018. England Domestic Overnight Trips Summary-All Trip Purposes-2018. [Accessed 15 December 2020]. Available from: https://www.visitbritain.org/gb-tourism-survey-2018-overview.

Waddington, T., Clarke, G.P., 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.

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. and Reynolds, J. 2012. Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. Geoforum. 43(6), pp.1076–1087.

Wood, S. and Reynolds, J. 2011. The intrafirm context of retail expansion planning. Environment and Planning A. 43(10), pp.2468–2491.

Ye, Z., Clarke, G. and Newing, A. 2021. Estimating small-area demand of urban tourist for groceries: The case of Greater London. Journal of Retailing and Consumer Services. 58, p.102263.

Chapter 7 Discussion and conclusions

7.1 General overview

This thesis has sought to answer the research question: "***how can location-based social media data be used to explore the spatial behaviour and spending patterns of urban tourists and contribute to more accurate retail location modelling in the grocery sector of urban destinations?***" First, the potential of LBSN data for helping to understand tourist spatial patterns in urban areas hosting significant tourist visits was assessed. By taking the world-leading urban tourism destination London as an example, the thesis has explored (using LBSN data) how the estimated tourist population, activity and mobility patterns could be used in the spatial modelling process to quantify the impacts of tourism on local residential areas with a geographical lens at a local level. In particular, the understudied links between tourism and grocery spending was chosen as a specific research agenda, especially given the growth in short-term self-catering rental accommodation outside core tourist areas, which are thought to generate additional grocery spending in the locality. A wealth of data was collected and analysed in the area of Greater London to produce an estimated tourist grocery demand layer which could be added into retail location modelling. The new SIM built with additional tourist demand is capable of supporting store revenue prediction, trade area analysis and new store location planning in urban destinations which experience an influx of diverse tourist groups.

The thesis has contributed to two main strands of research. On the one hand, this work contributes to LBSN data analytics in the field of urban tourism research. It demonstrates that the concentration and movements of the digital traces left by tourist users from LBSN services have the potential to reveal interesting new spatial patterns on both the activity and mobility dimension of tourists in urban destinations. In this regard, the thesis particularly chose Sina Weibo – a rich but under-utilised LBSN dataset to investigate the spatial travel behaviours of inbound Chinese tourists in London. Despite the limitation of this data only relating to one tourist group, the analysis of the Weibo dataset highlights that tourist trajectories from LBSN sources are capable of extracting tourists' spatiotemporal behaviour. When added to the collation of other English language based multisource

LBSN data sets, we can obtain a rich pattern of tourist movements across an urban area. The data integration of LBSN data and conventional sources are used to model the small-area tourist population distributions, which offers a more nuanced insight into the volume, composition and distribution of various forms of urban tourism, which underpins the quantification and assessment of tourist impact on host communities at a range of geographical scales.

On the other hand, this work also contributes to retail location modelling in urban areas. With a unique focus on tourist grocery demand, this research is the first work to address the linkage of urban tourism and the local grocery market in host urban destinations. To date, non-residential demand in retail models has largely been made up of workplace demand and tourism in seasonal mass-tourist destinations, such as coastal regions (Newing et al., 2015; Waddington et al., 2019). In urban destinations, the findings of this thesis suggest that the grocery demand from tourists and visitors are both important in retail location analysis and in some areas tourism can make a substantial contribution to store revenues and the composition of catchment areas. This research is the first known work that incorporates urban tourist grocery demand from all forms of tourist – overnight and day visitor, international and domestic into the SIM for retail modelling process at the scale of a major city.

Many of the findings in this thesis also identify the challenges when exploiting different types of LBSN data to gain information on urban tourist patterns or during the incorporation of urban tourist demand in the retail location modelling process. The next section discusses these findings in response to the original research aims and objectives and reflects on the challenges of the methodology adopted, along with broader limitations of the work. It also discusses the practical implications of the research in a broader context and identifies future research opportunities.

7.2 Summary and critique of research findings

The research began with three overarching aims and eight specific objectives (Chapter 1). Organised by each of the aims and corresponding objectives, this section discusses the potential contributions and implications alongside the challenges identified during the research.

7.2.1 Reviewing and exemplifying the utility of LBSN data analytics for exploring urban tourist patterns

The prevalence of LBSN platforms provides enormous and ever-growing geospatial big data with high spatial and temporal granularity. With the advance of urban data science, these LBSN data have provided new opportunities to generate knowledge for urban tourism research and shed new light on the practice of tourism planning and destination management. Previous research on tourist behaviour analysis has relied heavily on surveys and questionnaires, in which destination management organisations usually predetermine what issues should be addressed, thus perhaps missing out on other important tourist interests (Miah et al., 2017). The works presented in Chapters 2 and 4 indicate that LBSN services are of tremendous potential by offering large volumes of spatiotemporal data regarding tourist travel behaviour. With the help of spatial analysis techniques (see Chapter 4 and Appendix A), tourist patterns extracted from LBSN data allow important trends to be revealed at a fine spatial and temporal scale alongside outliers being identified for further investigation. Furthermore, it is possible to link different LBSN datasets in the same region to provide a more comprehensive source of data (beyond conventional survey data) to gain insights into urban tourist patterns at high levels of spatial resolution.

Chapter 2 addressed the potential of LBSN data as a means of exploring tourist experiences (J. Li et al., 2018) and reviewed and critically assessed the utility of different sorts of LBSN data in extracting and modelling spatial patterns of urban tourist population density, activity and mobility behaviour. It highlighted that three main types of LBSN data all have their unique opportunities as well as limitations when employed to understand tourist travel patterns. In brief:

- location-based check-in data has the advantage of identifying tourist activities revealed by multipurpose tourist mobility trajectories (but with a focus on only the most popular venues);
- geotagged social media data is suitable to estimate tourist spatiotemporal densities and mobilities;
- tourism service website data offers the most complete and up-to-date data on destination tourism offers and demand-side utilisation but may lack individual user’s trajectories and thus only can be analysed at the collective user level.

Therefore, it is important to adopt multiple LBSN datasets to offer a multidimensional view of tourist spatial patterns. These novel and emerging geospatial big datasets can work as great supplements of reliable surveys and statistics to offer invaluable sources to investigate tourist presence, activities, movements and interaction within a host city at a high spatiotemporal level of resolution.

Through the case study in Chapter 4 and Appendix A, the research shows that as an emerging geospatial data source, LBSN data offer new opportunities to answer the question ‘how do tourists experience the city?’ In prior work, the applications of LBSN data relating to more specific tourism research questions have been reviewed, such as geotagged Twitter data for travel demand modelling (Abbasi et al., 2015) or Flickr geo-photos used to illustrate tourist experiences (Miah et al., 2017). Chapter 4 used the under-researched Weibo dataset in international tourism studies to introduce a more general LBSN data analytic framework which could be applied to extract and analyse urban tourist spatial travel behaviour from the LBSN data sources containing users’ travel records. This LBSN data analytics framework descends from Big Data analytics, in which a series of spatial analyses are deployed to collect, collate, wrangle, geovisualise and analyse the geospatial information associated with tourist travel behaviours from LBSN datasets. The framework is summarised as follow, with the technical details of the case study reported in Chapter 2 and Appendix A :

(1) Use APIs or web scraping tools to collect LBSN data;

(2) Identify tourist users;

(3) Form AOIs or tourist hotspots in destinations by spatial clustering algorithms such as DBSCAN;

(4) Infer tourist footprints;

(5) Geovisualise tourist density, by type of activity when possible;

(6) Construct travel trajectories of individual tourist users according to individual time-series LBSN records;

(7) Aggregate individual trajectories to collective travel movements;

(8) Uncover tourist mobility patterns based on those travel movements;

(9) Understand destination attraction features by individuals.

In the case study (see Chapter 4 and Appendix A), Weibo check-in datasets within Greater London were collected as a unique but typical location-based check-in data set. The results showed that LBSN data is a valuable data source generated by tourists (inbound Chinese tourists in London in this case) to extract their spatial travel patterns at both an individual and a collective level, and more importantly, to reveal previously uncovered tourist interests and preferences in term of their activity choices, travel routes and core-peripheral attraction visits. It provides a novel way to create a tourist segmentation based on the multipurpose travel behaviours of tourists, which revealed that almost half of the sampled Chinese tourists are traditional visitors of landmark attractions, undertaking common mobility patterns and travel routes. It also revealed that a second large group, comprising around a fifth of Weibo users, show strong shopping preferences at a high diversity of popular shopping venues. Given that Chinese tourists have been identified by UNWTO (2014) as the largest international tourist expenditure group globally, Chapter 4 examined the shopping-related consumption activities for each segment to demonstrate the heterogeneity of the shopping venue choices and roles in their multipurpose travel itineraries. The identified overlap between tourists' and residents' utilisation of local retail services informed subsequent work in this thesis to investigate how different forms of tourists may share local grocery retail services which were most likely targeted at the resident population only.

7.2.2 Developing a methodology for urban tourist population modelling and demand estimation based on data collated from conventional and LBSN sources

Tourists and visitors are recognised as temporary populations, important though for understanding daily urban movements and consumption patterns within destination cities (Charles-Edwards and Bell, 2015). Comprehensive, spatiotemporally detailed and accurate estimates of tourist populations are thus a prerequisite for assessing their social and economic impacts within local neighbourhoods of host urban destinations. However, there is no single dataset that captures the spatial distribution of tourist populations at a fine geographical scale, albeit disparate LBSN data emerge as new means of exhibiting varied urban tourist patterns (as discussed in Chapters 2 and 4). This thesis has developed a methodology to model estimates of tourist population and demand in urban areas. This has been demonstrated in Chapter 5 and Appendix B, where a diversity of LBSN datasets (alongside conventional statistics and surveys) were amalgamated to generate a more holistic representation of the urban tourist market in London.

By reference to the tourism administrative survey outcomes (Table 5.1), this research disaggregated the tourist population in London into four main tourist groups - day trip visitors and three types of overnight tourists: tourists using traditional commercial accommodation, travellers spending their nights at Airbnb properties and guests staying with relatives or friends. Tourist arrivals and nights spent are reported at an aggregate scale within official surveys or statistics (typically at a city or borough level in the case of London). These official sources are reliable and consistent but limited in spatiotemporal detail and thus there is potential to combine these with the novel emerging LBSN datasets to produce a series of small-scale tourist population distribution maps by tourist type.

Based on the discussion in Chapters 2 and 3, it was shown that geotagged social media data from Twitter, Flickr or other services are possible sources to produce estimates of the daytime tourist population, whereas tourism accommodation service websites are regarded as useful indicators of overnight tourist distributions (Spalding et al., 2017; Patel et al., 2017; Batista e Silva et al., 2018). In this thesis, the density of tourist geotweets was utilised as a proxy to disseminate borough level daytime tourist population into small area day visitor population mapping; Airbnb listing and reservation records were used to create the short-term rental guest distribution surface and the serviced accommodation locations were generalised from Ordnance Survey and OpenStreetMap. The tourist population estimates produced in Chapter 5 and Appendix B were used within the subsequent analysis to understand the small area distribution of different tourist groups and as inputs to tourist demand estimation and modelling.

Chapter 3 demonstrated how the spatial extension of tourism has expanded from the urban central area to more peripheral suburban neighbourhoods in many urban destinations. This now means that both tourists and local residents co-habitate in many areas and contribute more equally to local expenditure patterns. Tourists now utilise the services and facilities initially designed for the use of local residents and workers. In this regard, increasing research seeks to examine and assess the impact of emerging tourism on residential areas. The thesis draws on the prior evidence of the spatial extension of tourism into the ‘off the beaten track’ areas of London (Maitland and Newman, 2009; Maitland, 2013) and endeavours to bridge urban tourism with grocery retailing. The tourist grocery spending patterns created in Chapter 5 suggest that tourists, interweaved with the local population (cf Chapter 3), contribute a considerable amount to grocery sales. According to the estimations in Chapter 5, tourists contribute more than 10% of the grocery demand in nearly 12% of the LSOAs in London. These demand estimations are based on the generated tourist population distribution and potential grocery expenditure rate per person or household (as detailed in Chapter 5). Since grocery retailing is not normally considered as a key tourism product, and its expenditure has not been reported in tourism satellite accounts, the best efforts were made in Chapter 5 to find reliable sources of expenditure rates for different tourist types.

The produced high spatial and temporal resolution tourist population estimates are valuable sources to understand the role of urban tourism in the host destination at a fine spatial scale. The small-scale estimates of tourist population distribution could provide empirical evidence to support in-depth studies into the tension of urban tourists as a floating population and the overlapping local dwellers (such as overcrowding, deterioration of public spaces, environmental pressures, etc.) or to assess the service accessibility for these transient users (i.e. public transport, food and drink, etc.) (Novy and Colomb, 2019; Shabrina et al., 2019). Associated with the phenomenon of tourist gentrification in urban destinations as touched upon in Chapter 3, such small-area population mapping of tourists in both day and night time enables scholars to investigate the role of urban tourism in the housing issue, demographic change and urban transformation of the host communities (Cocola-gant, 2018).

7.2.3 Incorporate urban tourist demand into retail location modelling

An examination of how urban tourist demand interacts with the local grocery supply was conducted in Chapter 6. A disaggregated tourist SIM was built and compared to a traditional model based on residential and workplace demand (non-tourist model: see Appendix C) to show the impact of tourists on store-level catchment areas and store performance/revenues.

The estimated tourist demand layers of the four different types of tourists and visitors were fed into the SIM as separate layers and calibrated independently using bespoke model parameters to capture the mobility patterns associated with each group. Specifically, the observed tourist grocery shopping trips of Airbnb and serviced accommodation users were informed by the Foursquare movements between accommodation sites and grocery stores; the distance decay parameter of day visitors was calibrated by the Foursquare movements from attraction locations to grocery outlets;

for the free guests staying with their friends and relatives or at second home the average travel distance was defined as being the same as the residential model in London.

A major challenge during model building is the validation of the model outputs. Previous research incorporating non-residential demand in SIMs have been supported by actual sales data via industry collaboration (Newing et al., 2015; Waddington et al., 2019). As this was not available for this project, this thesis turned to the reported industry data (i.e. market share and trading intensity of retailers) concerning retailer performance to validate the effectiveness of the model. As outlined in Appendix C.4, these are the best empirical sources regarding grocery store performance. Further validation could be implemented via collaboration with retailers who could provide more comprehensive store performance datasets which would assist in model calibration and validation, but is unlikely to provide ‘whole of the market’ data.

The incorporation of tourist grocery demand into the grocery shopping SIM offers a more comprehensive insight into the local composition of grocery retail demand. The findings in Chapter 6 show that for some stores tourists make up a quarter to a third of the expenditure in the modelled store catchment area. In this respect, the neglect of tourist demand will impair the predictive power of conventional non-tourist retail models and underestimate store revenue prediction. Although we lack actual individual store revenues, the comparison of the tourist-included SIM against the tourist-excluded SIM suggests the revenue uplifts of each store (and their spatial variation across London) have a close relationship with the distribution of Airbnb units and their utilisation. The ‘what-if’ scenario analysis presented in Chapter 6 also demonstrated that the tourist SIM built upon urban tourist grocery demand can be used to assist future store expansion plans on the supply side. The model could estimate the revenue change of any proposed new stores both in terms of residence-based and tourist-based expenditures.

In the domain of grocery retailing, Chapter 6 also suggests that with new tourist demand trends, local grocery markets may face challenges and opportunities from the spatial re-distribution of tourism and its associated consumption activities. The estimated results from the model could inform local marketing campaigns, to make them more focused around the complexity of the population in the trade area, as driven by the volume and range of accommodation stock in the area. The model allows the disaggregation of store customer types to estimate the potential sales from each subgroup. This could be significant in store preparation and management. Where model outputs estimate substantial uplift of store revenues based on tourism, tourist demand should not be ignored. Although using grocery shopping as an example, the study offers a viable approach to evaluate the impact of urban tourism on other aspects of the local economy (non-food goods for example). Similarly, although London has been used as a case study, the ideas and methodology should prove applicable elsewhere.

Besides the above contribution in retail location planning, the proposed tourist SIM has wider implications for the urban studies community. As presented in Chapter 6, the model is capable of quantifying the grocery provision ratios in local areas. After incorporating the tourist demand, the outputs of the model estimate a decline in the level of grocery provision per capita in many areas, some of them are located in the peripheral suburban areas which may not be well prepared for the new influx of tourists. As many destinations have seen a growth in short-term rental accommodation such as Airbnb, grocery shopping has been included as a desired facility in the property description (Stors, 2020) and may increasingly become a part of tourist food consumption. Therefore, the areas that are predicted to experience a decrease of grocery provision ratios due to tourist inflows should consider optimising store operations by serving longer opening hours or a greater range of products tailored for tourist needs.

7.3 Limitations of the work

7.3.1 Limitation of Weibo datasets

The thesis used Sina Weibo check-in and POI datasets in London as a case study to demonstrate the opportunities and potential of LBSN data for exploring urban tourist patterns and their spatiotemporal distribution, activities and movements. These datasets offer a large volume of Chinese tourist observations at a high spatial resolution in London. However, the quantity of big data cannot guarantee data quality (Brunsdon and Comber, 2020). As outlined in Chapter 2, with Weibo data, like other LBSN datasets, the issues of bias, skewness and representation are often raised when applying the data in human behaviour analytics (Roick and Heuser, 2013; Malleson and Birkin, 2014; Lansley et al., 2018). There are also limits to this research when leveraging Weibo datasets to gain information about Chinese tourist patterns in London.

The active users of these LBSN services are usually only a small fraction of the overall tourist population. Internet connection accessibility and free Wi-Fi provision in the destination cities may also influence the motivation and ability of tourists to engage with social media (Ferreira et al., 2020). These shortcomings of social media datasets raise the classic representation issue (Yao et al., 2019) - how well the extracted urban tourist patterns represent the activities of the entire population. From 2017 to 2019, the average annual number of Chinese inbound visits to the UK was over 850,000. Normally, London hosts more than half of the inbound visits to the UK. In the case of China, due to nearly all the direct flights from China to the UK landing in London, London is included as a major stop of almost all the Chinese tourist visits to the UK (VisitBritain, 2019a), which means the annually Chinese inbound visits in London should be more than 425,000. On the other hand, according to the collected Weibo check-in dataset (see Appendix A.2), between 2016-01-01 and 2018-08-28, there were 22,118 check-ins from 6,534 Weibo users assigned as Chinese tourists. Therefore, approximately 2,300 Weibo tourist users per year have been captured in the dataset, making up only 0.56% of the actual Chinese visits in London.

Meanwhile, although traditional sample-based data collection techniques (including GPS data) in tourism research also involve bias (and representation concerns of sample selections), the data are usually collected in a controlled environment with characterised personal details and sample quotas to fulfil (Shoval and Ahas, 2016). However, such detailed information on users cannot usually be captured using LBSN. Although in this study, the Weibo users are entirely Chinese and it is unnecessary to infer the origin of country of these tourists, other demographic characteristics such as gender, age and social class are very difficult to predict as reviewed in Chapter 2, although some progress has been made based on user names, textual contents or historic location records (Sloan et al., 2015; Huang and Wong, 2016; Longley and Adnan, 2016; Stock, 2018). This is mainly because the Weibo is a Chinese language social networking service based in mainland China and user information is still limited and there are also associated privacy and ethical concerns. Meanwhile, Weibo Data Centre (2021) reports that Weibo users are dominated by the younger generation - almost 80% of their users are under-30s, with more female than male users. Therefore, when using Weibo check-in data to analyse Chinese tourists, the sample dataset over-represents the younger generation and female tourists. Thus, the LBSN data-driven approach in this thesis may provide insights into the spatial patterns of tourist presences and flows to some extent, but it does not offer a means of fully quantifying tourist distributions and dynamics. It does help, however, when data can be cross validated with the findings from different LBSN datasets or combined with published survey data as also shown in this thesis. The skewness issue discussed in Chapter 2 also occurs in the Weibo check-in datasets. Weibo is dominantly used in mainland China and the users may use the service to share their experiences during travel. However, compared to over 1 million tourists during the day in London (GLA, 2014), the Weibo check-in data from Chinese tourists is only a fraction of this number and the spatial distribution is rather sparse. Therefore, the spatial skewness originating from the dataset inevitably affects the results of Chapter 4.

Furthermore, apart from the check-in dataset, the Weibo POI dataset also has its limitation. First, the POI system in Weibo is chaotic with many mistakes. The attractions may have several POIs. Second, the Weibo POIs are categorised in Chinese without a defined classification system. However, as introduced in Chapter 2, tourist activities can be inferred based on the venue category. Therefore, to retrieve a correct and concise category system is crucial. Appendix A.2 and A.3 fully report the approaches that have been used to generate a clean Weibo venue dataset with the help of Foursquare venue category structure. It is also worth mentioning that uncertainty exists when inferring tourist activity type simply from the category of the venue they visit. For example, people may visit an iconic shop to take photos rather than actually consume there. There is the potential of inaccuracy in such an interpretation.

Based on the above discussion, therefore, it is important to be critical when using Weibo and other similar LBSN datasets in urban tourism research. However, apart from complementing existing sources in urban analytics, LBSN data usually offers the most important (or sometimes the only) freely accessed spatiotemporal logs concerning the travel behaviour of individual tourists at many tourism destinations. These new forms of data are valuable in exploiting tourist travel patterns at fine spatiotemporal scales, although with the abovementioned limitations. Thus, it is important to integrate multiple data sources, especially with reliable administrative datasets (Roick and Heuser, 2013). The vast quantities of small area data on tourist travel behaviour from LBSN should be combined with routine official statistics to reweight and enhance the representativeness of that data (Birkin, 2019).

7.3.2 Limitation of tourist population modelling

In this thesis, a wide spectrum of publicly available data from both conventional sources and LBSN services has been used to present the details of the defined four major tourist groups in London. Thus, the production of small-area tourist population estimates is also to some extent limited by the available datasets. These will be summarised below:

(1) Non-serviced accommodation: only the sharing economy accommodation properties registered on Airbnb were included. Although Airbnb is the most successful platform, and one that has built a strong partnership with London's community tourism programme, there are still many other short-term rental and house sharing suppliers offering similar services (GLA, 2017a). For example, there are over 150,000 entire properties providing accommodations for backpackers in London on Couchsurfing; service platforms such as Holiday Lettings, Homeaway and Booking.com each has more than 5,000 registered short-term rentals in London (Inkson, 2019). However, arguably it is almost impossible to obtain a complete estimation of these sharing economies and short-term rental properties. Also, Inkson (2019) identified the duplication of accommodation listings across a number of suppliers, but tourist utilisation of these accommodation services is usually also not as openly accessible as Airbnb. However, it is worth noting the incompleteness of the dataset.

(2) Hotel and serviced accommodation: there is no one source that lists the accommodation stock in this sector along with attributes including location and bedspaces. This research collected and linked two open-source POI databases (Ordnance Survey POI and OpenStreetMap) to gain one consistent service accommodation POI dataset. Error and missing entries may exist in the original datasets. Moreover, the dataset was disaggregated to the LSOA level to evenly distribute the serviced accommodation bedspace stock by one bedspace occupancy rate across the LSOAs in London (as no disaggregated occupancy rate is available). However, in practical terms, the bedspace count of each accommodation establishment can be vastly different, as the type, price, ranking and location all could influence the choices of tourists and lead to an uneven utilisation of the service across London. Therefore, the small-area modelling of these serviced accommodation tourists is a coarse estimate and might not fully represent the area variation of tourist preferences. Detailed and up-to-date serviced accommodation datasets at fine granularity regarding its supply and

utilisation from the hotel industry or tourism survey would benefit future research in this area.

(3) Free guest/own home: the estimates of free guests and own home stayers are derived from IPS, GBTS and 2011 Census by ethnic group, which are reliable in terms of representation. The work here mapped these inbound tourist stays by linking with the usual residence of each corresponding ethnic group. That seemed more appropriate than allocating randomly. However, do say Japanese tourists always stay with Japanese residents? It could be an arbitrary approach to attribute free guest/own home tourists to the ethnic community just based on where are they from. In addition, some of the tourist groups are difficult to map onto a typical ethnic group, such as Singapore, South Africa and Russia (see Appendix B.1). The work assigned tourists from these markets to the closest ethnic groups. There are also 21.5% free guest/own home tourist nights in London are from other minor markets, which were evenly dispersed to the distribution of households in London. Although a coarse method, the numbers in these cases are small. This allocation process might be improved through additional study work, but given the relatively small numbers involved, the gains may be marginal.

(4) Day trip visitor: based on the definition of day trip visitors in London, this research only considers domestic visitors in London as the day trip visitor market. Whether a Twitter user is defined as domestic depends on the most intensive activity centre of a user according to his/her historic geotweets. The spatial variation of day trip visitors over the LSOAs in London was estimated by linking day trip visitor geotweets in London with the borough-level daytime population survey results concerning day trip visitors. The technical details of generating small-area day trip visitor population can be found in Appendix B.2. Since the population estimation is based on geotagged Twitter dataset, the aforementioned limitations of LBSN data analytics may also exist at this step and lead to some uncertainty issues.

7.3.3 Limitation of data sources for the spatial modelling of tourist SIM

The SIM is essentially a spatial modelling approach to simulate the flows between origin and destination. In the case of this study, the tourist SIM is constructed to estimate the grocery shopping flows from the location of tourists to the nearby grocery stores in London. As noted in Chapter 6, the calibration of distance deterrence parameter β is a crucial part of this spatial modelling process. In previous studies, such calibration was aided by empirical data provided by one or more collaborating retailers (Newing, 2013; Waddington et al., 2019). Due to the lack of survey sources concerning tourist travel patterns, in this study the distance decay parameter was calibrated based on the tourist grocery trips derived from Foursquare movement datasets. The skewness and representation issue discussed in Chapter 2 is also likely to influence the effectiveness of the calibration procedure here. Even though there are plenty of tourist observations that can be collected from the Foursquare datasets in London, the venues and movements related to grocery shopping trips is still limited. Among the Foursquare venues in London, only 3.2% of the POIs are grocery stores or supermarkets and the movements that can be used to simulate tourist grocery trips only account for 12.9% of all the harvested Foursquare movement dataset. The proportion of accommodation in Foursquare venues in London is 4.35% and the attraction is 11.4%, which make up 9.2% and 30.2% of the Foursquare movement dataset respectively.

Also, the validation of the model results was not possible to conduct due to the lack of empirical store sales and performance data. In an ideal world, therefore, the tourist SIM would benefit from more data to validate its robustness and accuracy. This means that errors might propagate through the process of model construction and comparison. As presented in Appendix A.4, some reported data regarding retailer performance has been used to validate the accuracy of the model outputs, but more solid and comprehensive sources that would improve the validity of the research were unavailable.

Finally, the building of a tourist SIM in other major urban destinations may confront similar challenges of data availability. The research draws on a broad range of datasets of conventional sources derived from the surveys and census of the UK and the daytime population surveys provided by GLA for London. The data sources used in this research may not be available for other cities. The varied circumstances of data availability in different destinations may lead to distinctive limitations that impede estimating the tourist population in other cities. Furthermore, although the emerging open accessed datasets such as OpenStreetMap, InsideAirbnb and other social media data can be retrieved ubiquitously, it is expected that the LBSN data in some cities may be much sparser than in London and lead to more biased and skewed datasets. This may affect the tourist population modelling and the calibration of SIMs when leveraging LBSN data to simulate tourist presences and movements.

7.4 Future research agenda

The topics in this thesis can be expanded in a broad range of future directions.

As noted above, many aspects of the presented work are limited by data availability. Therefore, the rapid advances in embedded sensor technologies with smartphones and other mobile devices, in conjunction with the increasing open scheme of data sharing, provide immense opportunities to use greater high resolution space-time data on tourist mobilities, which will improve the breadth, depth and scale of tourist travel data sets. These highly accurate and reliable spatiotemporal data would benefit this research and also other applications of LBSN big data. These enriched and refined tourist travel datasets, which also cover long periods of time, will increase the understanding of tourist travel behaviour and the combination of diverse tourist datasets could provide invaluable additional insights to inform the small area estimates of tourist populations in the future. It also possible to implement a real-time data infrastructure of tourist LBSN data collection, integration and analytic for the vision of ‘smart tourism’. But the empirical datasets required to validate the accuracy and validity of the proposed SIM can only be accessed via industry collaboration, which is an obvious area of future research.

Another avenue for future work is to move the research focus from London to other urban destinations. Recently, work on other European cities such as Barcelona, Madrid, Lisbon, Paris and Berlin and some urban regions in Australia have also observed the increase in tourism in residential areas due to the proliferation of Airbnb properties, which has attracted scholars to investigate the undergoing tourist gentrification and displacement process in these cities too. It would be useful to test the methodological framework proposed in this thesis in other cities. It would also be worthwhile to examine what data sets might exist for smaller cities, to again test if the methodology and results could be replicated. This would be important too in the context of grocery retailing given that the grocery sector itself, and the importance of grocery stores as a source of tourist food and drink, will be different in these cities.

In a similar way to studying different cities, it would be of interest to extend the analysis from the grocery sector to the effects on other services and the urban environment, such as housing, food and beverages, nightlife and green space. It should be relatively straightforward to adapt the methodology used here to other key sectors of the service economy. Furthermore, constant annual tracking of the tourist patterns in urban destinations will facilitate the longitudinal analysis of tourist demand trends, which would be advantageous to understand the transformative changes driven by tourists in cities. Especially in the post-Covid era, it is a need for many destinations to investigate the repercussion and recovery of their tourism economy. These LBSN datasets and the extracted tourist patterns will offer great research opportunities in this area.

7.5 Concluding remarks

In light of all the above, the thesis concludes as follows:

Firstly, this thesis has evaluated LBSN data as an emerging means to refine, extend and enlighten the understanding of the spatial travel behaviour of urban tourists, although the skewness and biases of the LBSN data should always be considered. The research recommends that LBSN data should be combined with census and other routine surveys to better inform small-scale tourist population modelling. It provides novel opportunities to understand the micro-geographies of the population in urban areas hosting a significant number of tourists and visitors. The proposed methodology of tourist population modelling can be applied to other tourism cities to create a comprehensive evidence base to quantify and assess tourist impacts on host communities at fine geographical scales. After identifying the spatial expansion of urban tourism in destinations, the thesis has argued that in major urban destinations tourist demand on grocery should not be neglected. In the case study area of London, tourists make up an estimated 6% of the total grocery expenditure, 71.8% of which are now generated outside the CAZ and 32.7% are in the outer boroughs. In regard to tourist type, it has been estimated that Airbnb guests and day trip visitors are the major contributors to tourist grocery demand, followed by free guests and serviced accommodation travellers. The spatially varied tourist grocery demand estimates suggest there are unmet gaps even in large metropolitan cities like London which are often assumed to have sufficient grocery supply. It is likely to be the same in other major urban areas - hence, there may be scope for other major urban destinations to produce similar small-area tourist demand estimates. Building on that, the thesis has later shown the benefits of adding tourist demand into the retail location modelling of grocery retailing in urban destinations. The tourist-incorporated SIM has contributed an effective approach to quantify and assess the tourism impact on local neighbourhoods. It has also been argued that the new model is capable of facilitating more accurate retail location analysis within areas that have many urban tourists. The disaggregated insights into catchment demand composition could assist store operations by tailoring the products and running marketing campaigns (especially to different types of tourists). The addition of tourist demand in SIM enhances the predictive capacity of the model in terms of estimating the store revenue and brand performance; thus the new model is more suitable to be applied in store performance benchmarking, new store site selection, store network optimisation and other retail planning practices at major urban destinations. The next step of the research seeks to use this model in collaboration with a retailer in relation to examining the impacts of real, planned development schemes. To sum up, as cities experience the spatial diffusion of tourists into more suburban areas, grocery stores should be considered as an important source of tourist food and drink and this should be considered more by both retailers and retail planners. Not only will this improve the store location research undertaken in tourist areas by retailers themselves, it will also provide opportunities for gaps in the provision of supply to be proactively identified and addressed by urban planners.

Appendix A Supplementary notes for Chapter 4 (Paper I)

This appendix provides complementary notes on the techniques discussed in Chapter 4.

A.1 Data collection

This case study collected all the Points of interest (POIs) within Greater London by Weibo Open API https://api.weibo.com/2/place/nearby/pois.json with a defined boundary, whose coordinates are referenced by minimum longitude (51.278239), minimum latitude (-0.514428), maximum longitude (51.697120) and maximum latitude (0.343027) respectively. Within this study area there are 2,665 POIs which have been used as 'check-ins' by Weibo users. The detailed information of each POI can be retrieved by https://api.weibo.com/2/place/pois/show.json. The key fields of the Weibo POI dataset are listed in Table A.1. Weibo created a large number of POIs within Greater London for users to check-in at, but Weibo users can also create a new POI when they want to check-in at places or venues that have not been included in the POI dataset. As a result, the category system of POIs in Weibo is quite messy and incomplete. Such a category system cannot be used to explore tourist activity directly, so the category association is indispensable in this case study. This study categorised all the POIs into 8 groups: Art & Entertainment, Outdoors & Recreation, Professional & Other Places, Food, Education, Hotel and Travel & Transport. The detailed approach conducted in the category association is discussed in section A.2.3.

Table A.1 Key fields and description of Weibo POI.

POI ID	Object ID of the POI
POI Title	Textual description of the POI
POI address	Address of the POI
Category	Category of the POI
Check-in Number	Sum number of check-ins at the POI until 2018-08-28
Check-in User Number	Sum number of users who have been checked-in at the POI until 2018-08-28

Lat	Latitude of the POI
Lon	Longitude of the POI

For each POI, only the most recent 1,500 check-ins can be downloaded by API https://api.weibo.com/2/place/pois/users.json, but only 31 out of all the POIs have check-ins of more than 1,500. It means that for most of the POIs, all the check-ins can be retrieved. Every downloaded check-in has a series of information, with the fields most relevant to this study listed in Table A.2. In addition to spatial data, temporal information of check-in time can also be obtained. For an individual user, the first and the last check-in for the collected dataset are designated as t_{min} and t_{max} respectively. As a result, a total of 148,039 check-ins within London were collected and used in this study, among which the earliest check-in was on 2013-04-16 and the latest one was on 2018-08-28. The Weibo check-ins before 2016 are too sparse and the new Sino-UK tourism VISA started from 2016. Therefore, this research only used the check-ins after 2016-01-01, meaning 114,471 check-ins are used in the subsequent analysis.

Table A.2 Key fields and description of downloaded Weibo Check-in.

Check-in ID	Object ID of one check-in
Check-in date	Date (yyyy-mm-dd) of one check-in
Text	Text message contained in one check-in
User-ID/name	Object ID and name of the user who check-in
POI-ID/title	Object ID and title of the place which being checked-in

A.2 Data cleaning and optimizing

The noisy nature of data from social media is a commonly criticized drawback of LBSN data when compared with often well-structured and ‘neat’ survey data. Therefore, data cleaning is a crucial process before any analysis can be conducted. The research utilised two datasets from Weibo: the Weibo check-in dataset and the POI dataset. For the check-in dataset, the research needed to identify the check-ins from tourist Weibo users and also filter out the data from fake tourist accounts. For the POI dataset, the data cleaning and optimising steps include location detection and category association.

A.2.1 Identifying tourists

Weibo is an almost exclusive Chinese user social media service. By using a ‘reasonable time-span’ criterion, it is possible to distinguish each Weibo user as either a Chinese inbound tourist or a resident (Chinese migrants in London). For every Weibo user, the length of stay in London is calculated by the time span:

$$t_{span} = t_{max} - t_{min} \tag{A.1}$$

t_{span} is the time span of each user and used as the estimated length of stay in London; t_{max} is the timestamp of the last Weibo check-in of the user’s record in London; t_{min} is the timestamp of the first Weibo check-in of the user’s record in London.

The time-span criterion was chosen with the help of the International Passenger Survey (IPS). Every year, IPS conducts more than 700,000 interviews, of which over 250,000 are used to produce estimates of overseas travel and tourism for usage by various government sectors including VisitBritain and the national and regional Tourist Boards. IPS provides a quarterly Travelpac dataset containing information on inbound tourists’ country of residence, duration of visit and expenditure. Using the Travelpac data for 2016, the box plot in Figure A.1 shows that the mean length of stay of Chinese tourists (including leisure and business tourists) was 20 nights. Also, according to VisitBritain, Chinese tourists on average stay 19.84 nights when they visit the UK (VisitBritain, 2017). London is usually both the arrival and departure city for Chinese tourists in the UK. Thus, this research identifies 20 continuous days per year as a suitable threshold to separate tourists and short-stay visitors from longer-term students and residents.

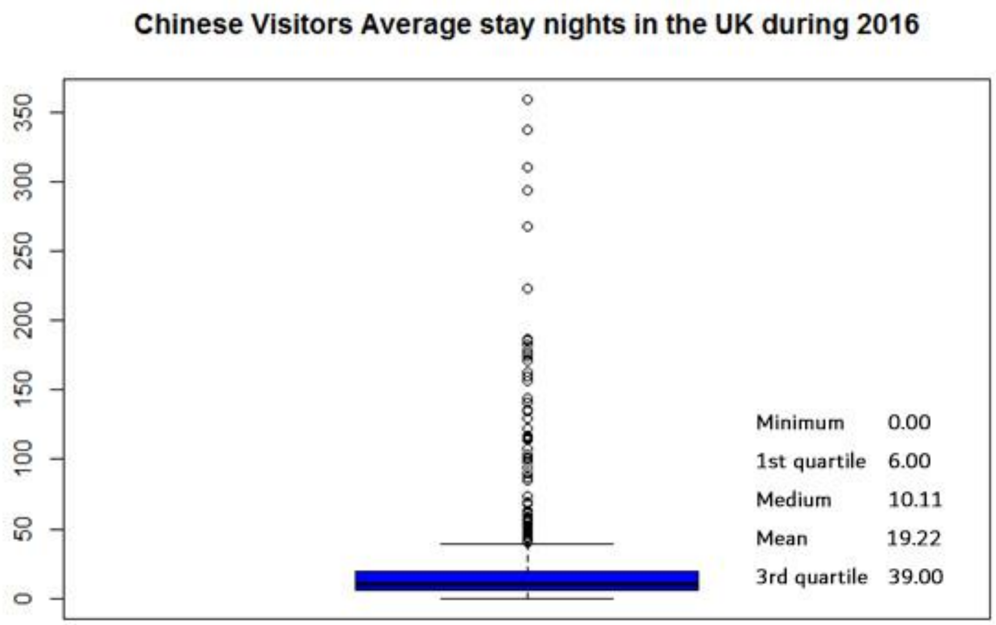


Figure A.1 Chinese visitor average stay nights from IPS TravelPac 2016.

This work adopted 15 different POIs in a day as a preliminary threshold of real users. It is easy to filter spammers and other non-person users out by using this criterion to examine every user's check-in log in Greater London. According to the above criterion, 6,534 users with a total of 22,118 check-ins were assigned as tourists, while 11,947 Chinese users generating 53,811 check-ins were categorised as residents, students or long-stay visitors. There were also 2,364 users who only had check-ins on one day and 26,959 users who only had one single check-in. It was impossible to decide whether they were tourist or resident users. These users and check-ins were not included in the subsequent analysis. Finally, 72 users having intensive check-ins (at more than 15 different locations) on one day were identified as fake accounts. These users and the related 1,412 check-ins from them were excluded.

A.2.2 Location detection: from POI to AOI

The data cleaning of POI datasets is more complicated. The spatial uncertainty of LBSN data is driven by the accuracy of the mobile devices, GPS and the precision in translating place names to coordinates and geographic references, which can result in incorrect and inconsistent POIs. Incorrect POI names are obviously errors, but the inconsistent POIs usually appear at some large venues, such as parks and universities. There could be several different POIs which are closely related to the same venues. The fundamental reason is the issue of scale; in other words, if the research is conducted at a small scale, different POIs will be considered as different places. But at a large scale, the POIs are related to the same AOIs, which means it is better to regard them as a whole and use the collective check-ins as the check-ins of this AOI. Hence, location detection is introduced as a key step for obtaining reasonable and consistent POI datasets.

This thesis used DBSCAN for location detection to identify nearby POIs and aggregate them when they are in the same category. Given that only the Attractions and Education category are more likely to have this multi-POI issue, this step is not applied to the other 6 categories. This category criteria also helps to avoid missing but necessary POIs: some small accommodation places and shopping venues for example, which are important in this research to study tourist shopping activities in detail. In each cluster, the origin POI with the highest number of check-ins is assigned as the new POI; meanwhile, the number of check-ins at all the origin POIs are summed to be the check-in number of the new POI. The origin POIs (now aggregated into new POIs) are then removed from the POI dataset. The evolution of POI optimizing is illustrated in Figure A.1 (POIs having less than 5 check-ins have been erased to avoid random visits). Figure A.2(a) presents the origin POIs around the British Museum, collected from Weibo. There are five different POIs which all located within the geographical extent of the British Museum. After spatial clustering by DBSCAN, the 5 POIs were generalised into only one (the one has the largest number of check-ins among them), and all the check-ins were added to the remaining POI (Figure A.2(b)).

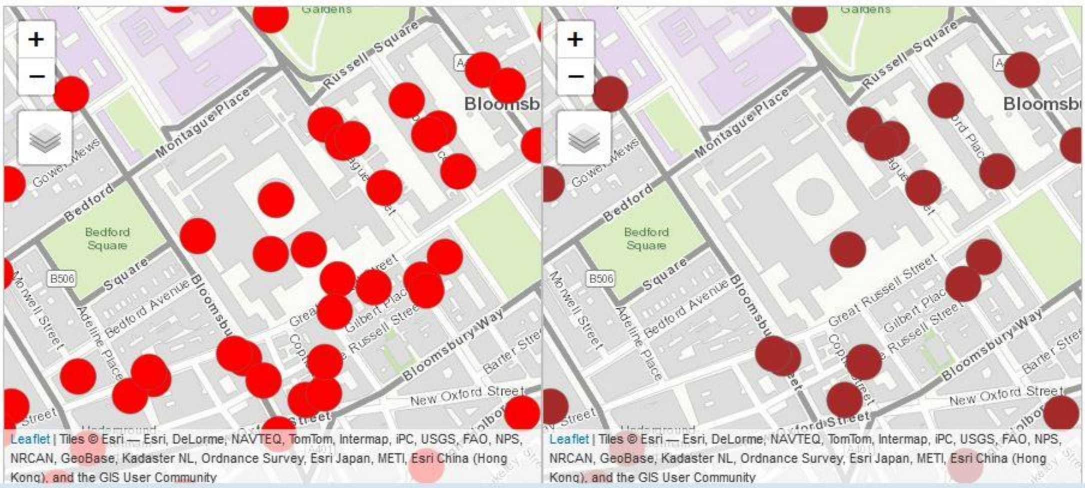


Figure A.2 Evolution of POI optimizing (a) before location detection; (b) after location detection.

A.2.3 Category association of AOI

The next step is to associate categories to the AOIs, with the help of Foursquare venue categories and API. The Foursquare venue database structured all the venues into 10 main categories: Travel & Transport, Arts &

Entertainment, Shop & Service, College & University, Event, Food, Nightlife Spot, Outdoors & Recreation, Residence, and Professional & Other Places. The tourism accommodation venues belong to the ‘Hotel’ subcategory under the main category of “Travel & Transport”. Therefore, the research first identified all the accommodation POIs (they have very few errors because they were created officially by Sina Weibo due to business cooperations), and then adopted the commonly accepted Foursquare venue categories to reclassify all the remaining POIs into 8 groups: Shopping, Attractions, Entertainment, Accommodation, Education, Restaurant, Transport and Others (Figure A.3). The classification of the POI category was based on mapping to Foursquare venues. By using the https://api.foursquare.com/v2/venues/search? and limiting the mapping result as the most likely venue, it is possible to know the venues and their correspondent categories.

For the aim of this case study, all the collected POI locations were associated at an AOI level. After noise filtering, location inference and category association, the study obtained a clean POI dataset of 1,329 AOIs in London and a related 20,223 check-ins from 6,265 tourist users. As will be detailed in the next section, the British Museum, London Heathrow Airport and Trafalgar Square/ the National Gallery are the top three most popular check-in locations and the categories that have most check-ins belong to Attractions and Education.

Activity type	Venue main category	Typical venue detailed category evidenced in check-in data
Visiting arts and entertainment venues	Arts & Entertainment	Museum, Historic site, Concert hall, Art gallery, Theater
Outdoor sightseeing	Outdoors & Recreation	Park, Palace, Scenic lookout, Bridge, Castle
Visiting monuments/building	Professional & Other Places	Monument/landmark, Church, Library
Shopping	Shop & Service	Department store, Market, Souvenir shop, shopping mall
Dining out	Food	Tea room, Steak house, English restaurant, Burger Joint
University-related activity	College & University	College academic building, College library, College Residence Hall
Accommodation	Hotel	Hotel, Hostel, Bed & Breakfast, Boarding House
Travelling	Travel & Transport	Hotel, Train station, Metro station, Airport

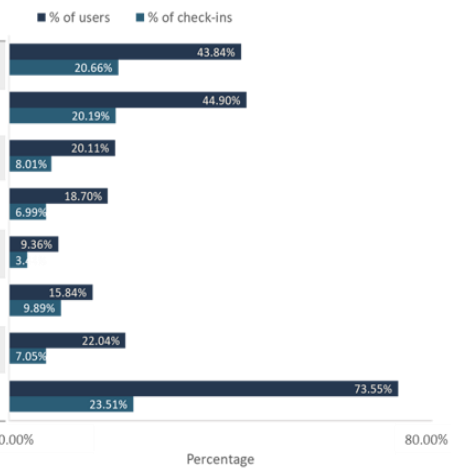


Figure A.3 Activity and venue choices of inferred Chinese tourists as derived from Weibo check in data.

A.3 Kernel Density Estimation

Weibo check-ins from tourist users were geovisualised by different activity types. The first three activity types “Visiting arts and entertainment venues”, “Outdoor sightseeing” and “Visiting monumnets and building”, as listed in Figure A.3, were aggregated as Visiting and Sightseeing activity. Other types of activities included in the geovisualisation were Shopping, Dining out, Travelling and Accommodation. The Weibo tourist check-ins of each activity type were extracted based on the venue main category as shown in Figure A.3.

Kernal Density Estimation (KDE) was used for the geovisualisation of both Weibo tourist check-ins and the Foursquare venue dataset for comparison, presented in Figure 4.1 and Figure 4.2 respectively. The KDE bandwidth are the scalar values of normal reference distribution and applied to both the x and y directions, based on the better-supported rule-of-thumb principle (Venables and Ripley, 2002):

$$bandwith = 1.06 * \min\left(\sigma, {}^{R}\!/_{1.34}\right) * n^{-1/5} \qquad \text{(A.2)}$$

R is the interquartile range of the standard deviation of the points, σ is the variance of the points, n is the number of points. The KDE surfaces are generated via R package MASS by the function *kde2d*, and the bandwidths are calculated by the default function *bandwidth.nrd*.

A.4 K-means clustering

A.4.1 Variable selection

Tourist travel behaviour at the collective level helps us to obtain an overall view of how Chinese tourists travel in Greater London. However, every tourist has different travel movements and activities reflecting their different experiences in London. The main aim of the research of Chapter 4 was to investigate the latent tourist multipurpose travel patterns and attempt to create an LBSN data-driven tourist segmentation. Therefore, this study introduced a series of variables to understand individual tourist travel behaviours.

Previous studies indicated that although of great significance, the similarity of user-activity is not good enough solely to create clusters. Among all the user travel dimensions, the transition between locations is the most effective, while the temporal dimension is the worst in clustering LBSN service users (Lian and Xie, 2011). Therefore, this study described tourist multipurpose travel from three different dimensions. First, the travel characteristic indicators described the basic travel feature of each tourist; second, activity preference indicators characterised each tourist's activity choices in London from the frequency and importance of each type of activity, the popularity and diversity of the venue choices in each type of activity and the multipurpose degree of the daily trips; third, the spatial mobility indicators depicted individual tourist travel behaviour based on the spatial analysis of their footprints and trajectories patterns.

A.4.1.1 Travel characteristics

Travel characteristic measurements were used to calculate the basic check-in behaviour of each user during their stay in London. Four measurements were used to quantify tourist travel mode: 1) length of stay, calculated by the time span as in Equation A.1; 2) number of trips, referring to the number of trip days that each user had at least checked-in in the study area; 3) number of stops, the total number of check-ins from one user, and 4) number of different AOIs, the total number of AOIs one user has checked in at during one trip.

A.4.1.2 Activity preferences

In this study, the category of one venue was used to suggest the activity of the tourists who checked in at this venue. This study considered the 7 main tourist activities in Figure A.3: Visiting arts and entertainment venues, Visiting monuments/building, Shopping, Dining out, University-related activity and Travelling and accommodation.

Tourists' activity preferences according to their check-ins are described from 5 different dimensions: 1) activity frequency is calculated by how often one activity accounts for the whole journey; 2) main activity frequency refers to how often the tourist takes one activity as his/her main activity in daily trips; 3) activity diversity means how many different venues the tourist visited under the same activity type; 4) activity popularity measures the total popularity of the venues visited by the tourist, aggregated by activity type; and 5) multipurpose degree quantifies how many different activities the tourist visits along the daily trip. The former four indices were calculated for each of the 7 activities, so there are a total of 28 variables. The last variable, named "multipurpose degree" is designed to measure the average number of multipurpose activities of the tourist. Therefore, the activity preference indices have 29 variables in sum.

A.4.1.3 Mobility patterns

Tourist mobility patterns were quantified by a series of variables from 3 main aspects: travel behaviour, spatial measurement and networks analysis.

(a) Travel behaviour

Tourist travel behaviour can be very different. Two measurements were used to further describe the travel behaviour of each tourist: 1) Mean transit AOI is the number of AOI tourist visits per day, which shows whether tourists take a busy or relaxed trip; 2) return probability is measured by the probability of returning behaviour, that is how often a tourist returned to the previously visited places.

(b) Spatial pattern of footprints

Spatial pattern measurements can quantify tourist movements by using geographical calculators. Since the aim was to explore how Chinese tourists experience Greater London during their stay, the check-ins at London Heathrow Airport were excluded from the spatial pattern analysis of tourist travel footprints.

The travel distance of each tourist was calculated by two quantitative indices: 1) mean travel distance refers to the average length of tourist's daily route; 2) mean travel placement is the average length between the two stops in a daily route. Point patterns of tourist' footprints can be quantified by Standard Ellipses Deviation (SED). SED is a centrographic measure to summarise the size and shape of a point set: a larger area means more dispersion points and a low eccentricity shows a more even distribution of the points (Huang and Wong, 2016). Therefore, by adopting SDE, the study calculated another two indices – 3) dispersion and 4) even distribution of each tourist's footprints from Weibo check-in data.

(c) Trajectory pattern from AOI networks

Two quantitative indices were designed to describe tourist's trajectory patterns: 1) total weight of routes indicates to what extent a tourist's travel route is similar to the majority of the tourists captured in these data, and 2) total degree of AOIs suggests to what degree a tourist visits core AOIs and/or peripheral AOIs.

After quantifying individual tourist travel behaviour from the 41 variables outlined in Table 4.2 for all the 5,634 tourists, the research only used the 1,171 active tourist users for understanding tourist multipurpose travel patterns. These 1,171 tourist users were chosen because they had daily routes consisting of more than 3 stops and could be used to create a segmentation of tourist multipurpose travel patterns. K-means was first used to classify all the 1,171 Weibo tourists; then, the research used LDA to understand the multipurpose trip characteristics of tourists within each classification group.

A.4.2 Standardisation

Before clustering, the variables need to be first standardised to make variables comparable. The standardisation includes transforming the variables to make them have mean zero and standard deviation as in Equation A.3.

$$R_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad \text{(A.3)}$$

Where R_i is the standardised value of variable X for tourist object i, X_i is the initial value, X_{min} is the minimum value of the variable X for all the tourist object, and X_{max} is the maximum value of the variable X for all the tourist object. R package ‘scale’ is used at here to scale the input data. After standardisation of all the 41 variables, the value of each variable ranges from –1 to 1, eliminating the potential errors due to different numeric scales.

A.4.3 Variable correlation

The R package ‘Hmisc’ function *rcorr* was used to create the correlation matrix of all the 41 variables in Table 4.2. The visualisation of the correlation coefficient is shown in Figure A.4, generated by R package ‘corrplot’ function *corrplot*. The variables in Figure A.4 are ordered by hierarchical clustering order. The hierarchical clustering analysis used the ward method, which is a popular clustering method of hierarchical agglomeration (Ward, 1963). Using Ward's Method, we start out with all sample units in n clusters of size 1 each. The algorithm initially forms n-1 clusters by considering the smallest error sum of squares and largest values. Then, in successive iterations, every step reduces one cluster and forms the clusters of observation until the results of error from the squares is minimised or alternatively the value maximised. In this case, n starts from 41 and stops at 8. The result of Ward clustering is visualised in Figure A.4 and the series number of variables within each rectangle represents the high correlation of these variables.

The eight clusters formed by the Ward hierarchical clustering are numbered from 1 to 8 as shown in Figure A.4. Except cluster 2, the other seven clusters correspond to the seven tourist activities and each cluster includes the four indicators of activity preference in Table 4.2. This finding is in line with expectations and demonstrates that there are no two activities correlated with each other. However, there are other variables which show significant correlation with some of the activity clusters. Cluster 5 consists of the variables related to tourist dining out activity but also correlates with the 41st variable ‘total centrality’, which suggests that the Chinese tourists who show a strong preference for dining out have an obvious tendency to also visit the most core attractions in London. Also, the travelling and accommodation cluster (No. 3) presents a high correlation with the length of stay and number of daily trips and return probability. This is also reasonable because the travelling and accommodation activities mainly include the check-ins at the transport hubs and hotels, and these variables all relate to travel in the city. The last cluster which correlates with other non-activity variables is Cluster 8, which is based on the ‘Visiting landmark and building’ activity. This activity correlates with number of stops, number of different attractions, multipurpose degree and mean transit attraction. This is probably because tourists are more likely to combine landmark and building visits with other activities (suggesting multi-place visit behaviour). Finally, the results for Cluster 2 indicate that the four mobility pattern variables are correlated, especially mean distance and mean displacement. But the two variables measure different aspects of travel and thus both are kept in the subsequent analysis.

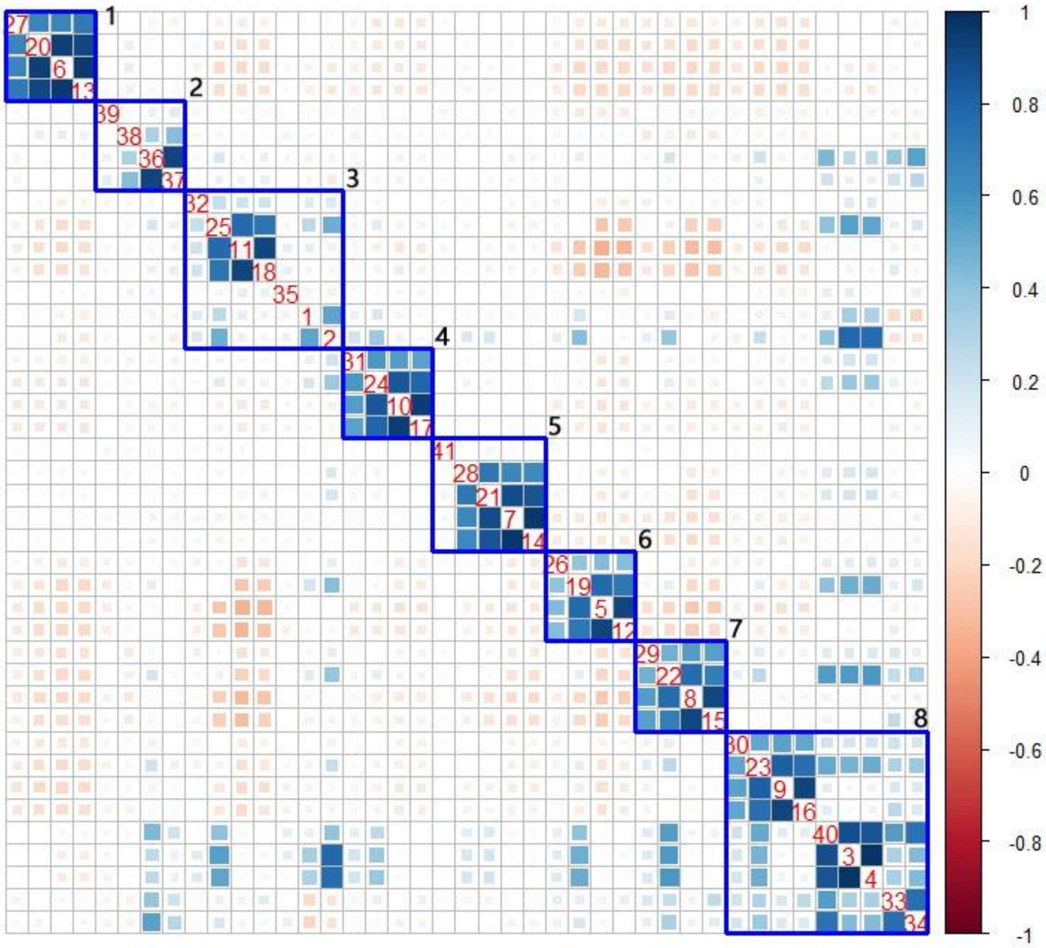


Figure A.4 Variable correlation matrix with the hierarchical clustering results.

A.4.4 Clustering

K-means clustering is a technique that creates groups by partitioning the objects in the shortest distance of similarity into subsets and reallocating the assignments until an acceptable set of groups is obtained. The K-means algorithm is efficient for clustering high-dimensional objects, therefore has been widely used in building neighbourhood classifications and geodemographic discriminators to support service delivery. This research aimed to identify the clusters of 41-dimensional objects (in Table 4.2), thus K-means was chosen for the clustering algorithm to identify various types of multipurpose trip patterns.

The only prior knowledge needed for k-means is to specify the number of clusters (k). K-means iteratively allocates objects into clusters until the variability within clusters is minimised and the variability between clusters is maximised. In this research, k was determined with the help of the R package ‘NbClust’. The ‘NbClust’ package provides the optimal number of clusters under 30 index solutions. Among the outputs of the NbClust, 12 out of 30 indices suggests five as the optimal k. The goodness-of-fit of the clustering is usually evaluated by two criterion: the total within-cluster Sum of Square which measures the Euclidean sum of squared deviations of each object to the cluster mean, and the total between-cluster sum of square which calculates the distance between each cluster. This research used the Calinski-Harabasz index in the R package ‘vegan’ to evaluate the result. The Calinski-Harabasz index is also known as the Variance Ratio Criterion and is calculated as the ratio of the sum of between-cluster dispersion and the sum of inter-cluster dispersion for all clusters. Figure A.4 shows the Calinski criterion of the k-means clustering using different k values (3 to 12), and the right plot indicates that when k=5, the Calinski criterion has the highest score. Therefore, the cluster number of 5 was chosen in this research to classify tourists into separated subsets.

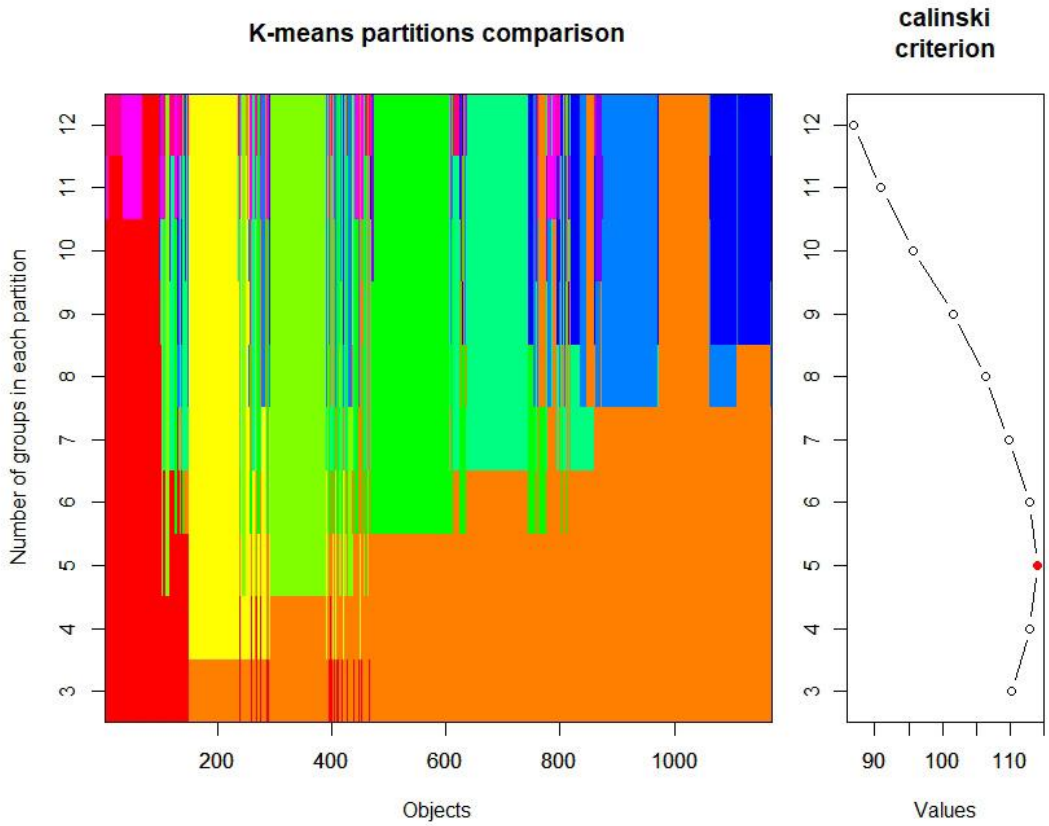


Figure A.5 Determination of the value of k for K-mean clustering.

A.5 Describing the clusters with LDA

Different from partitioning clustering algorithms like K-means, the LDA algorithm is a generative statistical model in natural language processing (NLP) which is widely adopted in document topic modelling. It assumes that each document contains a mixture of topics and each topic has a certain probability of mentioning a word. LDA has been widely used to identify topics from documents and calculates the proportion of different topics in each document by examining word distributions in the documents (Blei et al., 2003). Previous research has successfully extended the topic modelling approach to infer user life-style patterns from their Foursquare check-ins (Hasan and Ukkusuri, 2015; Qu and Zhang, 2013). In our work, we treat the daily trajectory of each Weibo tourist user as a “document” and the subcategory of each checked-in place as a “word”. The “topics” are hidden destination choice patterns in the tourist multipurpose trips.

When applying K-means to the 41 variables in Table 4.2, the clustering is based on not only the diversity and popularity of tourist activity choices, but also a spatial dimension reflecting how tourists visit and transit between attractions. For each cluster, LDA can be applied at the subcategory level to identify the detailed activity choice patterns. Therefore, to combine the advantages of both K-means and LDA, this study first used K-means to classify all the Weibo tourists into different clusters, then employed the LDA algorithm to further identify hidden multipurpose destination choice patterns by the subcategory sequence of tourist’s daily trips.

This research work first filtered out the outliers in each of the five tourist segments generated by K-means cluster and then applied the LDA algorithm to the subcategory sequences of remaining cluster members’ daily trajectories. The classification and further topic modelling results can be described as follows:

Cluster 1 – Traditional tourists (48.4%) is a group of tourists that only have a high proportion of check-ins at Arts & Entertainment venues such as museums and performing art venues. They have a longer length of stay but few of their daily trips are shared on Weibo. They travel to the most core attraction areas but their travel routes can be very different from the majority. The tourists in this cluster have the lowest level of multipurpose trips. The topic modelling results further reveal that although their daily trips include fewer activity types, they have more diverse subcategory types under the same activity: they mainly visit museums, performing art venues and stadiums, and incorporate other sightseeing activity at a variety of attractions. Their shopping activities may occur at department stores (LDA group 4) or souvenir shops (LDA group 1 and 2).

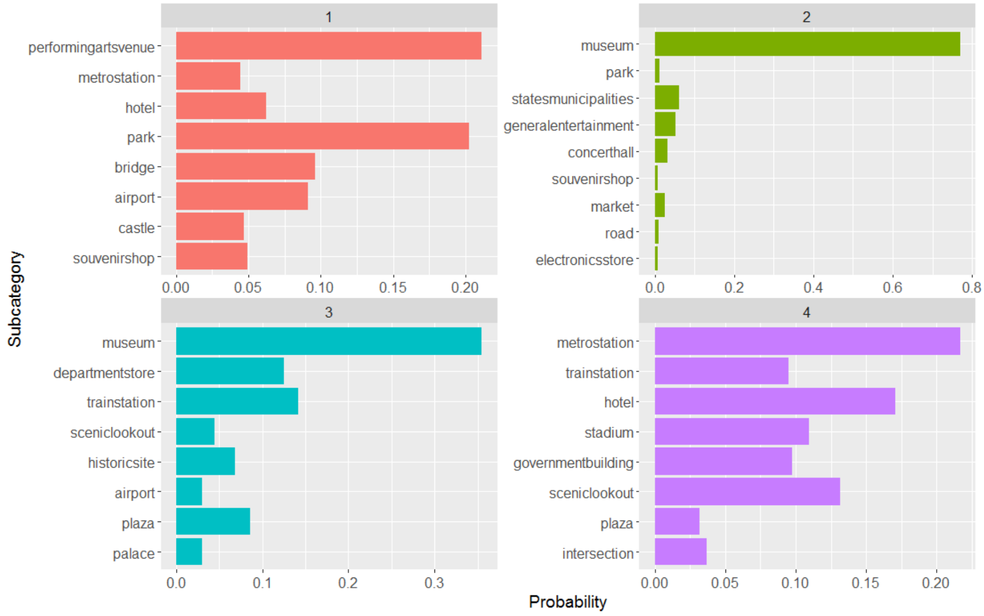


Figure A.6 Topic modelling result of Cluster 1 (Traditional tourists) based on subcategory.

Cluster 2 – Shopping enthusiasts (20.5%) is the most multipurpose group of tourists with clear shopping activities. Their greater number of stops at different attractions and the largest travel displacement both suggest their rather busy schedules. They travel by the most popular routes but only remain in a small region of London. They also enjoy visiting and sightseeing at different types of attractions.

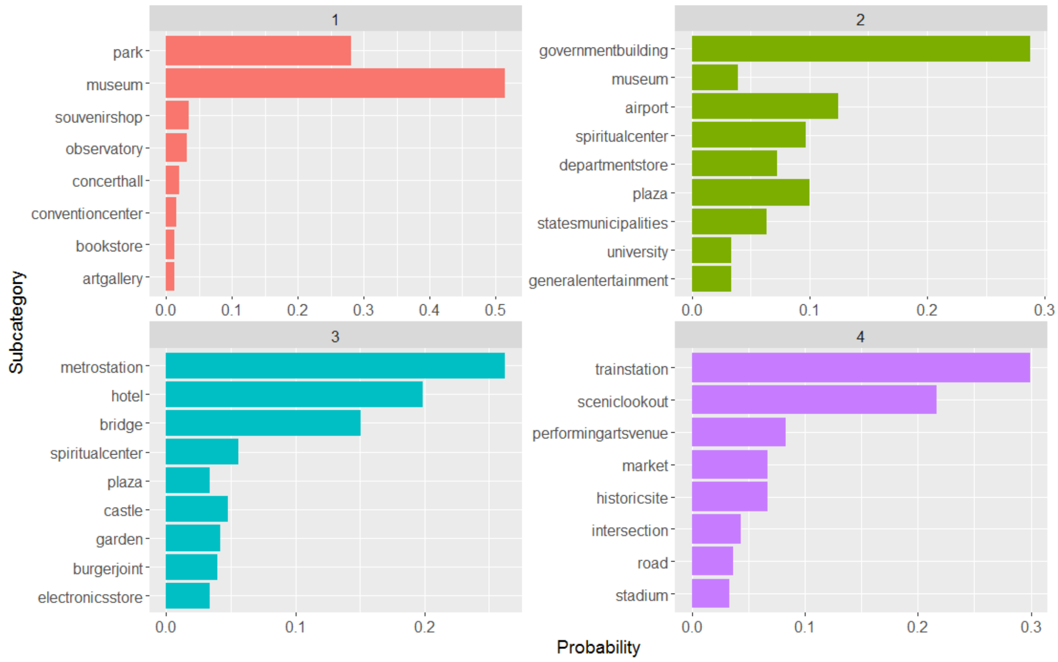


Figure A.7 Topic modelling result of Cluster 2 (Shopping enthusiasts) based on subcategory.

Cluster 3 - Gourmets (11.6%) is a group of tourists who have a very strong preference for dining out. They do not stay long in London, but they tend to share most of their trips on Weibo. They enjoy shopping a lot but it is rare for shopping to be their main trip purpose and they only visit the most popular shopping venues including department stores (LDA group 2 and 3), souvenir shops (LDA group 3) and markets (LDA group 4). They travel among the core part of the attraction networks but take rather relaxed trips (fewer visited attractions per trip). The topic modelling of this cluster shows consistent results of less sightseeing attractions but more restaurants compared with other clusters. Their choices of dining out venues can be diverse: Asian restaurants, English restaurants, French restaurants, Portuguese restaurants, Modern European restaurants, burger joints, and steakhouses.

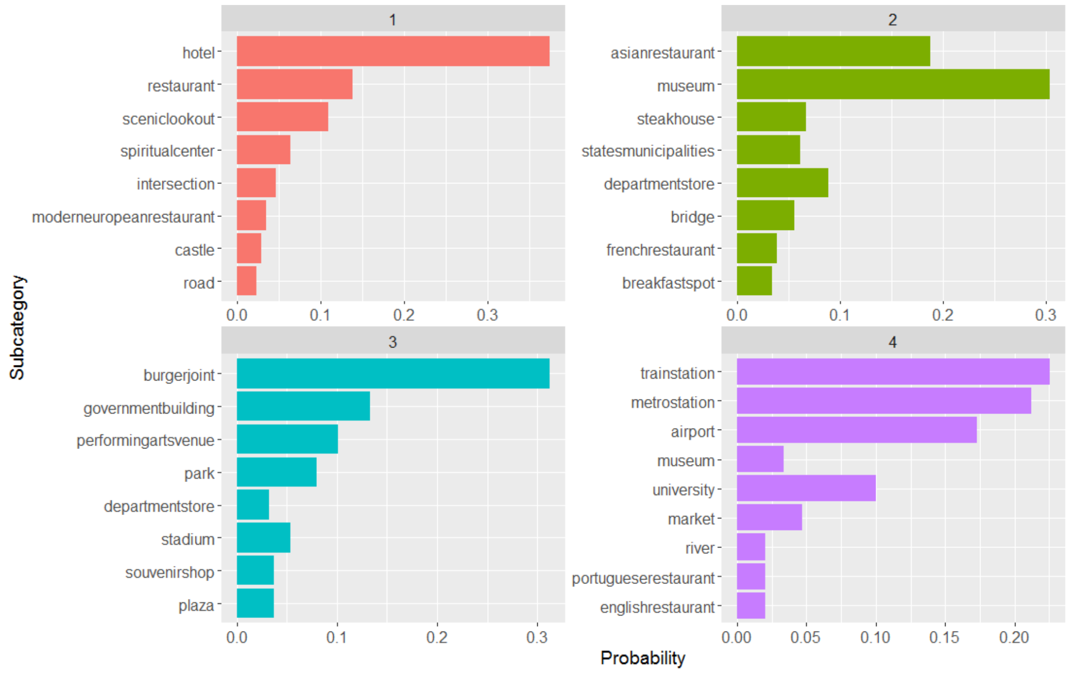


Figure A.8 Topic modelling result of Cluster 3 (Gourmets) based on subcategory.

Cluster 4 – Education (9.8%) is a highly multipurpose travel group with a dominant activity related to university or combined university and travelling, from both the results of K-means and LDA. They also enjoy outdoor sightseeing and shop at a variety of places including department stores, markets, electronic stores, souvenir shops, and bookshops, but each individual tourist does not visit many diverse shopping venues. This is the only cluster which shows an apparently periodic behaviour. Also, the spatial mobility pattern suggests that their footprints are the least evenly distributed and they may travel in uncommon routes to some peripheral areas in the attraction networks including some unpopular shopping venues.

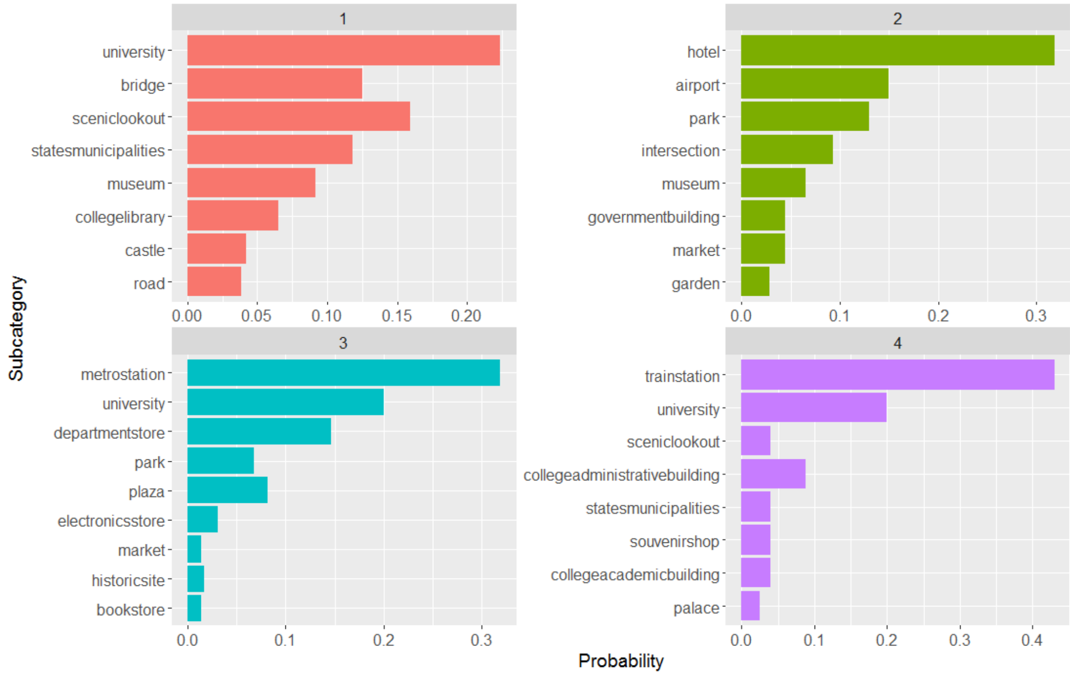


Figure A.9 Topic modelling result of Cluster 4 (Education) based on subcategory.

Cluster 5 – Outdoor sightseeing (9.6%) consists of tourists who mainly visit very popular sightseeing attractions. Their footprints are widely dispersed during their short stays, but they do not have very busy daily schedules, which is in line with their large mean displacement between attractions. The result of LDA also suggests that the tourists in this cluster visit and sightsee at various places but have a clear preference for religious buildings (group 3) and famous monuments/ buildings (group 2 and 4). The topic modelling results reveals that they only occasionally shop at department stores and souvenir shops.

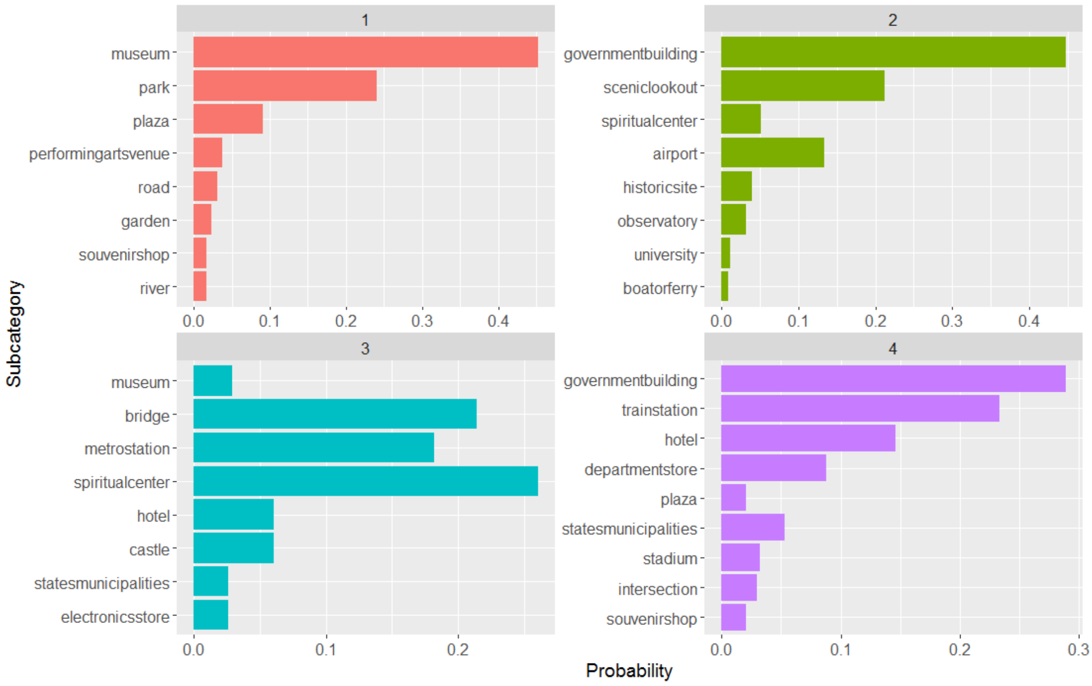


Figure A.10 Topic modelling result of Cluster 5 (Outdoor sightseeing) based on subcategory.

According to this Chinese tourist segmentation, among all the three dimensions, activity preferences significantly distinguish our sample tourists: tourists with different activity preferences are separated from each other, and those with similar activity engagements are clustered together, which shows the capability of the proposed framework for unveiling diverse tourists' activity preferences. The K-means results show that each segment has a varying degree of activity-based multipurpose patterns: Cluster 2 (Shopping enthusiasts) is the most multipurpose travel group whereas Cluster 1 (Traditional tourists) and 4 (Education) are the lowest. The topic modelling results also indicate that even within the same segments, the multipurpose travel patterns can be different: the Shopping enthusiasts (Cluster 5) and the Gourmets (Cluster 3) show more heterogeneous multipurpose travel patterns than the other clusters. This is probably because the visiting and sightseeing venues have less different subcategories than the shopping venues or restaurant types, hence, these tourists can combine visiting and sightseeing activities with their own specific preferences in much more variable ways. Therefore, the segmentation results indicate that based on tourist daily trajectory, it is possible to cluster tourists by their dominant activity engagements. However, tourists with the same activity preferences may still have distinct multipurpose travel patterns, especially for tourists who are fond of consumption-related activities such as shopping and dining out.

This segmentation also incorporates a spatial dimension, which reflects the mobility patterns of tourists with different multipurpose travel patterns. Shopping enthusiasts (Cluster 5) is a highly multipurpose tourist group but is also distinguished by its long mean travel distance and displacement but small dispersed footprints and great travel route weights. It helps to depict a persona for this segment: a group of tourists who travel within a small area with condensed and diverse services, although they enjoy plenty of different activities in their daily trips. They are more like tourists following the guide books rather than genuine city explorers. Another example is the art and entertainment segment Cluster 1 (Traditional tourists). This group shows very low travel route weights but high attraction degrees with a small footprint area, which suggests that the tourists in this group may not usually follow the common travel routes but they still remain in the core areas of the constructed attraction networks, rather than explore the peripheral areas. Their rather short travel distance and displacement also infer that this group enjoy travelling in a more relaxed way. In contrast, for the Education segment (Cluster 4), although also having low travel route weights but high attraction degrees, their travel area can be much larger and they show a unique periodic behaviour which is rarely seen in other segments. Contrary to other segments, in which the low mean distance always appears with the low displacement, the Education segment (Cluster 4), shows low mean distance but high displacement. These features help us to depict the persona of the university-related tourists: they do not travel actively every day but when they travel they tend to visit far distance places in the peripheral area of the attraction networks by transit from the core areas, and these visits are often repeated.

With a particular focus on tourist shopping activity, this Chinese tourist segmentation shows how shopping activities in the multipurpose trips of each tourist segment can be very different. Visiting and sightseeing clusters (Cluster 1) take account of the majority of Chinese tourists, but for these groups of tourists, shopping activity usually appears as a low proportion of their daily travel, especially when the dominant activity is visiting the museum – these tourists' shopping activities are even lower except at souvenir shops. But the department store can be a more common choice when their main activity is outdoor sightseeing or visiting performing art venues. In contrast, Shopping enthusiasts (Cluster 2) and Education (Cluster 4) show higher enthusiasm for shopping activity, but the Shopping enthusiasts segment is characterised by more diverse shopping venue choices but popular stores while the Education segment tends to have few shopping stops per trip, but with a preference for visiting less popular stores.

Therefore, these results suggest that by understanding tourist multipurpose travel patterns, and how they can include shopping activity in their daily trips, can be vital for practical tourism planning and destination management, especially for a world tourism city like London with abundant tourism resources on offer.

Appendix B Supplementary notes for Chapter 5 (Paper II)

This appendix provides complementary technical notes for the work presented in Chapter 5. In this Chapter, the tourist population was modelled using four disaggregated tourist segments at the LSOA level and the grocery demand for each was estimated separately. The four tourist groups in London considered in this research were: Airbnb guest, travellers staying overnight at a hotel and other serviced accommodation, free guests living with relatives or friends, and day trip visitors. The three former tourist types are overnight tourists who may generate different levels of grocery expenditure at the grocery stores near their temporary accommodation sites, thereby uplifting the store revenue in the locality. On the other hand, the day trip visitors may shop at the grocery outlets near the attraction and the venues, or the transport hubs they visit. This appendix adds more details of the methods used to create the population distributions of the four types of tourists.

B.1 Airbnb guest

The AirDNA datasets were retrieved in October 2019 from the CDRC (Consumer Data Research Centre) of the University of Leeds. The datasets used in this research include the total property data for London and the reservation records of each property updated to June 2018. There are 207,117 Airbnb properties according to the AirDNA dataset. The reservation records were used to calculate the actual utilisation of each property in the previous year (from June 29, 2017 to June 28, 2018) and to link to the property dataset for further analysis (see Table B.1). There were 106,974 properties (51.6% of all the listed properties) that had been booked and used during the time span. This study used the location and utilisation of these Airbnb properties in London during the one-year span to estimate the spatial distribution pattern of Airbnb guests in London.

Table B.1 The data structure of the property dataset.

Data field	Example
Property ID	3623312
Utilisation	590
Host.ID	17279844
Listing.Title	1Bed flat Piccadilly/Trafalgar, London sleeps3

Property.Type	Serviced apartment
Listing.Type	Entire home/apt
Created.Date	16/07/2014
Last.Scraped.Date	28/06/2018
Neighborhood	Westminster
Average.Daily.Rate..Native.	120.24
Annual.Revenue.LTM..Native.	14188
Occupancy.Rate.LTM	0.8
Number.of.Bookings.LTM	29
Number.of.Reviews	55
Bedrooms	1
Bathrooms	1
Max.Guests	3
Calendar.Last.Updated	28/06/2018
Superhost	FALSE
Minimum.Stay	2
Count.Reservation.Days.LTM	118
Count.Available.Days.LTM	29
Count.Blocked.Days.LTM	129
Latitude	51.509384
Longitude	-0.132466

The utilisation of the property was calculated using the reservation days during a year and the maximum guest number of the property. For example, the property (ID: 3623312) had been reserved for 118 days in one year. The capacity of the property is 3 guests. Therefore, in a timespan of one year, this property had hosted a maximum of 590 guests. When aggregating the utilisation of all the 106,974 Airbnb properties by LSOA level, the spatial distribution of Airbnb guest in London within one year in 2018 is presented as Figure 5.1.

B.2 Tourists stay overnight at serviced accommodation

The spatial pattern of tourists also depends on the distribution of the serviced accommodation they stay in. The serviced accommodation mainly includes traditional hotels, B&Bs, guest houses and hostels according to the glossary of 'Hotels and similar accommodation' in Eurostat[1]. Since there is no freely accessible dataset ready to use for this research, the location and

[1]https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Tourist_accommodation_establishment

utilisation of the hotel and other serviced accommodation was collected from Ordnance Survey and OpenStreetMap.

Ordnance Survey has a clear classification that can be used to extract the point dataset of hotels and other serviced accommodation establishments. The classification code ‘01010003’, ‘01010005’ and ‘01010006’ were used to identify the ‘bed & breakfast and backpacker accommodation’, ‘hostels’, and ‘hotels, motels, country houses and inns’ accommodation establishments in the Ordnance Survey POI dataset.

For the OpenStreetMap, the hotel POIs were downloaded by the R package ‘osmdata’ with the function add_osm_feature , selecting the region name as “Greater London”. The feature key was defined as “tourism” and the values of designated POI types were determined as ‘hotel’, ‘guest_house’, ‘hostel’ and ‘motel’, linking to the introduction of tourism POI types available from OpenStreetMap Wiki[2].

The two accommodation establishment datasets were then combined together using the R package ‘sf’ function st_is_within_distance. The POIs from the two datasets within a distance of 20 metres were identified as duplicates and combined as one POI. Finally, the research obtained a point dataset of 2,042 geolocated serviced accommodation in London. Figure B.1 shows the spatial distribution of these serviced and non-serviced accommodation establishments.

The bedspace of serviced accommodation in London was audited by VisitEngland in 2016 and released as the Accommodation Stock Audit (VisitEngland, 2016); The data is available at the borough level (see Table B.2). The next step is to distribute the utilised bedspace of each borough into the location of each serviced accommodation location. According to the England Occupancy Survey, the occupancy rate in London was quite stable during the year before the pandemic at an average of 60.2% (VisitEngland, 2019). Therefore, this research used this occupancy rate to evenly distribute the bedspaces in each borough to the location of serviced accommodation. The resultant spatial distribution of the serviced accommodation bedspaces in London is shown in Figure 5.3.

---

[2] https://wiki.openstreetmap.org/wiki/Key:tourism

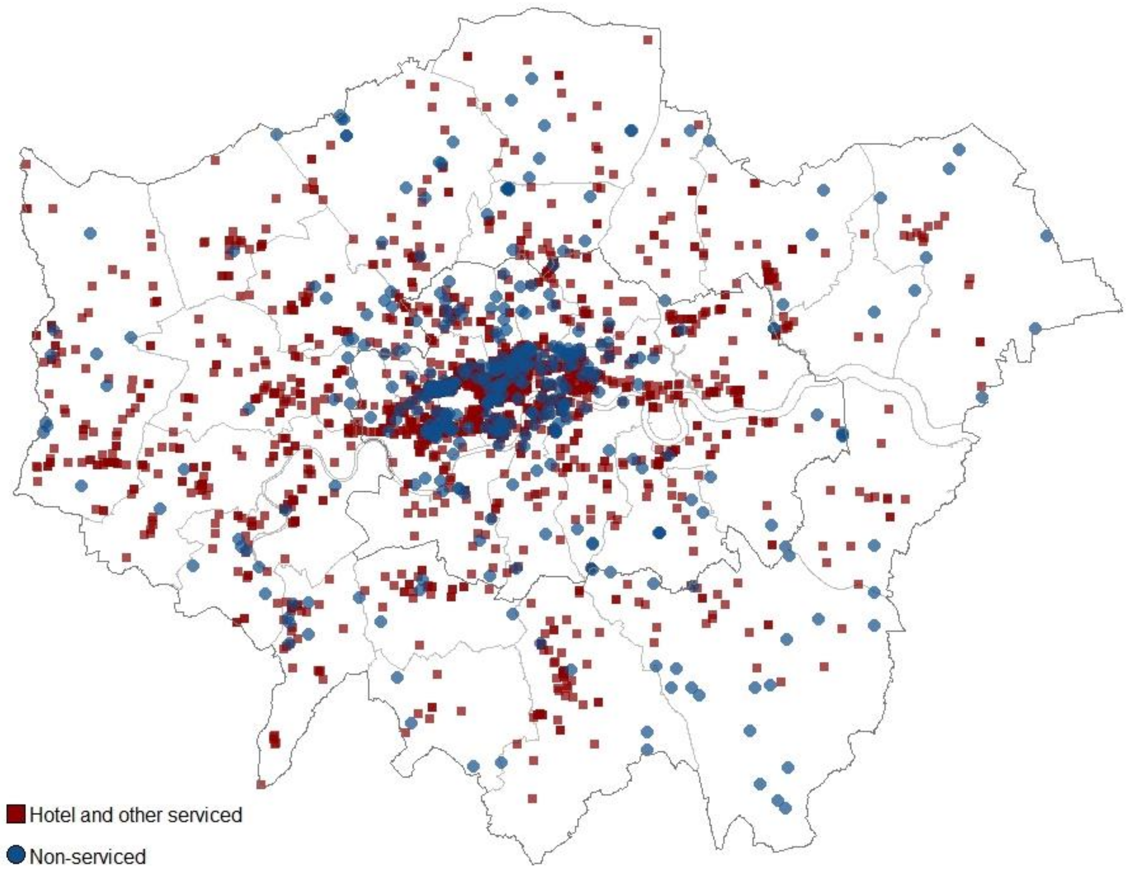


Figure B.1 The spatial distribution of serviced and non-serviced accommodation in London.

Table B.2 The traditional serviced accommodation stock by borough in London.

Borough	Serviced.estab	Serviced.room	Serviced.bedspace
City of London	59	6264	14543
Westminster	548	49746	107543
Kensington & Chelsea	475	31111	69696
Hammersmith & Fulham	102	7636	17249
Wandsworth	34	1534	3521
Lambeth	69	8539	19632
Southwark	34	3148	7765
Tower Hamlets	94	11180	25938
Hackney	21	1705	3940
Islington	62	5140	12593
Camden	244	25517	58371
Brent	44	2926	6681
Ealing	66	2501	6119
Hounslow	63	5116	12782
Richmond upon Thames	112	3311	7277

Kingston upon Thames	25	1088	3002
Merton	13	449	1305
Sutton	16	312	709
Croydon	55	3177	7489
Bromley	32	801	1785
Lewisham	21	276	621
Greenwich	48	2252	5089
Bexley	15	544	1198
Havering	19	716	1857
Barking & Dagenham	11	630	1576
Redbridge	39	1246	3046
Newham	47	4376	10369
Waltham Forest	21	1181	2704
Haringey	16	225	499
Enfield	13	655	1607
Barnet	42	2174	5084
Harrow	39	1129	2578
Hillingdon	83	11019	23996

B.3 Free guest with relatives or friends

Free guests staying with relatives or friends is a significant accommodation type for the tourist in London. Its economic impact, however, has been largely overlooked. The estimation of the spatial distribution of free guests in London is based on the free guest tourist nights reported by IPS (International Passenger Survey). The IPS regularly collects information about the numbers and types of visitors entering and leaving the UK since 1961. Normally, the IPS conducts 700,000 to 800,000 interviews a year, among which over 250,000 interviews are used for the estimation of inbound and outbound travel and tourism, on a monthly, quarterly and annual basis (ONS, 2016). The interviews are conducted with a random sample of passengers as they enter or leave the UK. The IPS estimates that approximately 90% of the passengers have a chance of being sampled on the survey. The robustness of the IPS estimate is reported as ranging from 95% confidence interval of +/- 1.9% of the estimate for total visits to the UK by overseas residents, to confidence intervals of +/- over 50 per cent for some estimates relating to visits to the UK from some countries (ONS, 2020). The study results are used by various government departments, including the ONS, the Department for Transport, the Home Office, HM Revenue and Customs, VisitBritain and the national and regional Tourist Boards. Particularly, VisitBritain, as a member of the IPS Steering Group, uses the data to not only produce overseas tourism reports and market-specific analysis, but also provides the UK figures to international bodies such as UNWTO, Eurostat and the European Travel Commission.

According to IPS, free guest with relatives or friends ranks in second place of the accommodation types of all the staying visits in the UK. In 2018, 26.43% of the visits in London were free guests staying with their relatives or friends, making up for 37.43% of all nights (VisitBritain, 2019b). These free guests living with local residents, bring an expected uplift to local grocery demand. To better quantify this grocery demand uplift at the small-area level, the research aimed to investigate the spatial distribution of the free guest tourists in London. A basic rationale of our methodology was that during travelling in a foreign country, free guest tourists are more likely to live with their relatives and friends sharing the same ethnic group. Therefore, it is sensible to distribute these tourist nights into the households of the same ethnicity in London. The methodology of estimating the overseas free guest tourists in London follows the steps listed below and the results are reported in Chapter 5:

(1) Identify the ethnic groups of the main source countries of the free guest tourists;

(2) Investigate the spatial distribution of each ethnic group at the LSOA level in London;

(3) Equally distribute the total nights of each source country into the LSOA in London according to the usual residents of each corresponding ethnic group population;

(4) Calculate the free guest nights in each LSOA and visualise the result.

IPS offers the inbound accommodation choices of the main source markets in the UK (Table B.1, created based on https://www.visitbritain.org/inbound-accommodation-research), but the origin countries of free guest tourists in London is unknown. According to Table B.3, the total free guest nights in the UK is 114,773,759. IPS also reports that 41,523,000 nights are spent by free guests in London, hence it is known that London accounts for approximately 36.18% of all the free guest nights in the UK. Therefore, we use 36.18% to calculate the total nights staying in London for each country.

Table B.3 Free guest tourist nights in London and their corresponding ethnic group.

Country	**Total nights in the UK**	**Total nights in London**	**Ethnic group (detailed) in 2011 Census**
Total	114,773,759	41,523,000	
USA	9,688,924	3,505,271	North American
Poland	8,576,304	3,102,747	Polish
France	7,741,148	2,800,603	European Mixed
Spain	7,655,698	2,76d9,689	European Mixed
Australia	6,861,947	2,482,524	Australian / New Zealander
India	6,825,483	2,469,332	Anglo Indian Indian or British Indian
Germany	4,472,134	1,617,935	European Mixed
Ireland	3,903,299	1,412,140	Irish
Canada	3,892,466	1,408,221	North American
Italy	3,519,890	1,273,430	Italian
Netherlands	2,298,544	831,570	European Mixed
New Zealand	1,898,212	686,738	Australian / New Zealander
Switzerland	1,736,376	628,188	European Mixed
South Africa	1,718,748	621,811	African White African White and Black African
UAE	1,573,877	569,399	Arab African/Arab White and Arab
Hungary	1,437,847	520,186	European Mixed
China	1,328,379	480,583	Chinese
Portugal	1,322,695	478,526	European

			Mixed
Belgium	1,249,267	451,961	European Mixed
Sweden	1,233,655	446,313	European Mixed
Argentina	1,223,444	442,619	Argentinian
Denmark	1,084,957	392,517	European Mixed
Greece	1,039,323	376,008	Greek Greek Cypriot
Norway	962,994	348,393	European Mixed
Czech Republic	870,474	314,921	European Mixed
Singapore	858,900	310,734	Chinese Malaysian
Hong Kong	745,299	269,635	Chinese
Brazil	646,399	233,855	Brazilian
Malaysia	625,631	226,342	Malaysian
Saudi Arabia	462,099	167,179	Arab African/Arab White and Arab
Russia	440,871	159,499	Commonwealth of (Russian) Independent States
Austria	403,703	146,052	European Mixed
Thailand	391,181	141,522	Thai
Japan	341,679	123,613	Japanese
Finland	287,452	103,995	European Mixed
Mexico	287,037	103,845	Mexican
South Korea	254,112	91,933	Korean
Luxembourg	162,182	58,674	European Mixed
Other	24,751,132	8,954,497	

In the 2011 Census, the standard 18 ethnic group response categories are expanded into 251 detailed ethnic groups at the OA level by country. The data at LSOA level can be downloaded via Nomis under the table of

QS211EW.csv. The core approach to creating the spatial distribution of free guest tourists is to allocate the tourist nights of each origin of country to the corresponding household with the same ethnic group. The allocation results are shown in Table B.3 and are used to generate the spatial distribution of the free guest tourists of each ethnic group. Of course this brings some likely errors. For some countries, it is straightforward to assign the source market country to a specific ethnic group, such as Brazil to “Brazilian”, Poland to “Polish”, and Japan to “Japanese”. Also, it is sensible to combine the tourist nights from the 14 European countries to the “European Mixed” ethnic group. But for some other countries, the allocation is not that straightforward and the origin countries are assigned based on the most related main ethnic groups in the census. This may result in the estimated spatial distributions not being as accurate as others. These origin countries are Singapore to “Chinese” and “Malaysian”, Saudi Arabia to “Arab”, “African/Arab”, and “White and Arab”, South Africa to “African”, “White African”, and “White and Black African” as shown in Table B.3.

B.4 Day visitors in London

Although very little information exists on actual small-area visitor numbers and their associated expenditure, in this section the research tries to demonstrate that the use of headline figures from national surveys such as the London Daytime Population (Table B.4) can be linked with novel social media data to generate small-area estimates of the day visitor distribution and their expenditure.

Table B.4 Day trip visitors across borough in Greater London.

Code	**Boroughs**	**Day Trip Visitors**	**% of total daytime population**	**% of total daytime tourists**	**% of total daytime population (excl. 0-4 & school children)**
E12000007	London	736,400	7.33%	66.99%	8.94%
E09000001	City of London	97,572	17.64%	80.16%	17.73%
E09000002	Barking and Dagenham	11,235	6.30%	81.76%	9.03%
E09000003	Barnet	18,066	5.07%	72.53%	6.55%
E09000004	Bexley	12,867	6.08%	76.85%	8.22%
E09000005	Brent	13,252	4.51%	69.88%	5.86%

E09000006	Bromley	24,767	8.16%	78.76%	10.61%
E09000007	Camden	36,759	7.42%	60.21%	8.08%
E09000008	Croydon	27,175	7.78%	78.98%	10.27%
E09000009	Ealing	14,354	4.38%	64.34%	5.67%
E09000010	Enfield	24,163	7.86%	80.97%	10.42%
E09000011	Greenwich	29,782	11.68%	86.05%	15.18%
E09000012	Hackney	18,639	7.08%	83.31%	8.99%
E09000013	Hammersmith and Fulham	17,807	7.73%	63.00%	9.12%
E09000014	Haringey	11,951	5.30%	76.66%	6.94%
E09000015	Harrow	20,882	9.11%	83.90%	11.72%
E09000016	Havering	15,743	7.02%	76.42%	9.03%
E09000017	Hillingdon	15,992	4.48%	52.84%	5.53%
E09000018	Hounslow	12,861	4.42%	68.28%	5.53%
E09000019	Islington	24,759	7.55%	80.94%	8.42%
E09000020	Kensington and Chelsea	21,204	8.39%	37.91%	9.57%
E09000021	Kingston upon Thames	10,982	6.61%	67.43%	8.43%
E09000022	Lambeth	18,145	6.10%	65.93%	7.42%
E09000023	Lewisham	17,484	7.64%	82.44%	10.31%
E09000024	Merton	10,551	5.44%	69.29%	7.01%
E09000025	Newham	23,823	7.78%	76.44%	10.38%
E09000026	Redbridge	11,613	4.58%	73.27%	6.46%
E09000027	Richmond upon Thames	12,905	6.80%	67.38%	8.91%
E09000028	Southwark	26,383	6.33%	76.59%	7.41%
E09000029	Sutton	12,597	7.09%	82.12%	9.62%
E09000030	Tower Hamlets	17,161	4.32%	58.65%	5.09%
E09000031	Waltham Forest	9,514	4.27%	76.73%	5.83%
E09000032	Wandsworth	20,130	7.55%	70.90%	9.65%
E09000033	Westminster	75,282	8.39%	42.79%	8.79%

The research downloaded a sample of geotweets by streaming from REST API from Sept. 21, 2018, to Oct. 01, 2019 in the Greater London area. All together, 1,316,218 geotweets from 325,077 users were collected. In order to explore further the spatiotemporal distribution of tourist geotweets, the structure and format of an individual tourist's geotweets are as shown in Table B.5.

Table B.5 The data structure of the geotweets.

Data field	Example
Geotweet ID	1099613290685110016
User ID	438097407
Latitude/longitude	51.5096/-0.2043
Venue	Kensington, London
Venue type	City
Text	Notting Hill is a 1999 romantic comedy-drama film directed by Roger Michell. The screenplay was written by Richard Curtis. https://t.co/dwBNtBSQaE
Text language	en
Create time	Sun Feb 24 10:13:32 +0000 2019
Timestamp	1.551e+12
Home	Bangkok, Thailand
User Language	en
Screen name	GotzyRedLips

The procedure to extract the geotweets from daytime visitor has three main steps:

(1) Identify users' origin of country to exclude the international tourist Twitter user in London

(2) Collect all the daytime geotweets from these users when they stay in London. The daytime geotweets are defined as the geotagged tweets that are posted during 9 am to 5 pm.

(3) For each of these users, trace back the historic geotweets, to identify the usual residence (or the workplace). The usual residence is defined as the location that the user posted there on different dates with a total of more than 10 geotweets. These locations are recognised as the usual residence of the user, and the geotweets there are excluded from the day trip visitor geotweets datasets.

(4) Create the spatial distribution of day trip visitors in London based on the density of the user in the LSOAs. It is the user rather than the geotweets that is used to depict the spatial pattern. This is to avoid any potential skewness which may occur when some users generated

much more geotweets than others leading to the over representativeness of some venues.

The technical details of each step is reported as follows:

Step 1:

According to the UNWTO definition of a tourist, the following criteria have been used to distinguish the international tourists in London:

i. The user's usual environment is outside the UK;
ii. The user's stay in the UK is less than one year;
iii. The user has at least one overnight stay.

Each user's usual environment has been gleaned from his/her historic geotweets by downloading from Twitter usertimeline API. After identifying the international tourists in the UK and excluding their geotweets, a set of 836,008 geotweets from 106,272 Twitter users belong to the local users and are used in the research for further analysis (representing 32.7% of all the users and 63.5% of the geotweets in the collected sample dataset).

Step 2:

Among these geotweets, there are 450,652 geotweets generated between 9 am to 5 pm, which makes up 53.9% of the local users' geotweets. research for modelling the distribution of day trip visitors.

Step 3:

The last step is to exclude the geotweets that are posted at the users' usual residence or workplace location. For each user, the locations that may refers to usual residence or workplace are identified as the place that the user has visited more than once and generated a sum of more than 10 geotweets there. After applying this criteria to each of the users in the datasets after Step 2, the 313,233 geotweets from 65,437 users are used to model the distribution of day trip visitors.

The tourist geotweets density by LSOA (Figure B.2) clearly shows a high spatial concentration of visitors in Central London, particularly Westminster, City of London, Camden, Kensington and Chelsea, Islington and Hammersmith and Fulham. Some LSOAs in Brent, Tower Hamlets and Newham borough also show dense tourist geotweets.

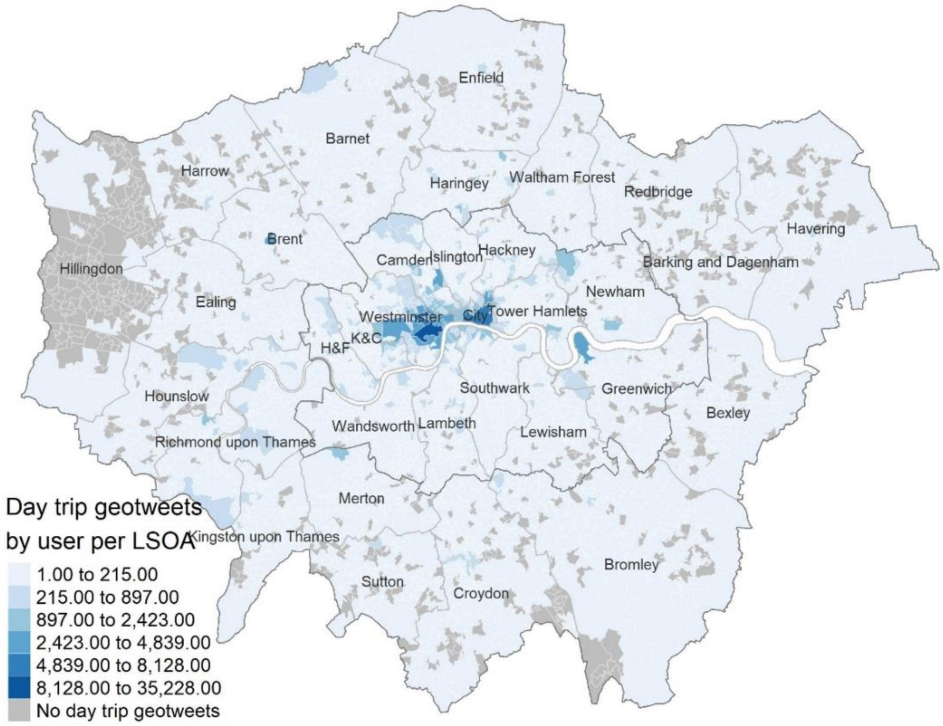


Figure B.2 The distribution of sampled Twitter users that are identified as day trip visitors in London.

Step 4:

Based on this dataset, Figure B.3 presents a proxy of day trip visitors over the LSOAs in London. The proportion of day trip tourist users in each LSOA has been calculated, nested within the borough (see Figure 5.8). By Linking to the population survey of daytime visitors as shown in Table B.4, the spatial distribution of day trip visitor in London is created as in Figure B.3.

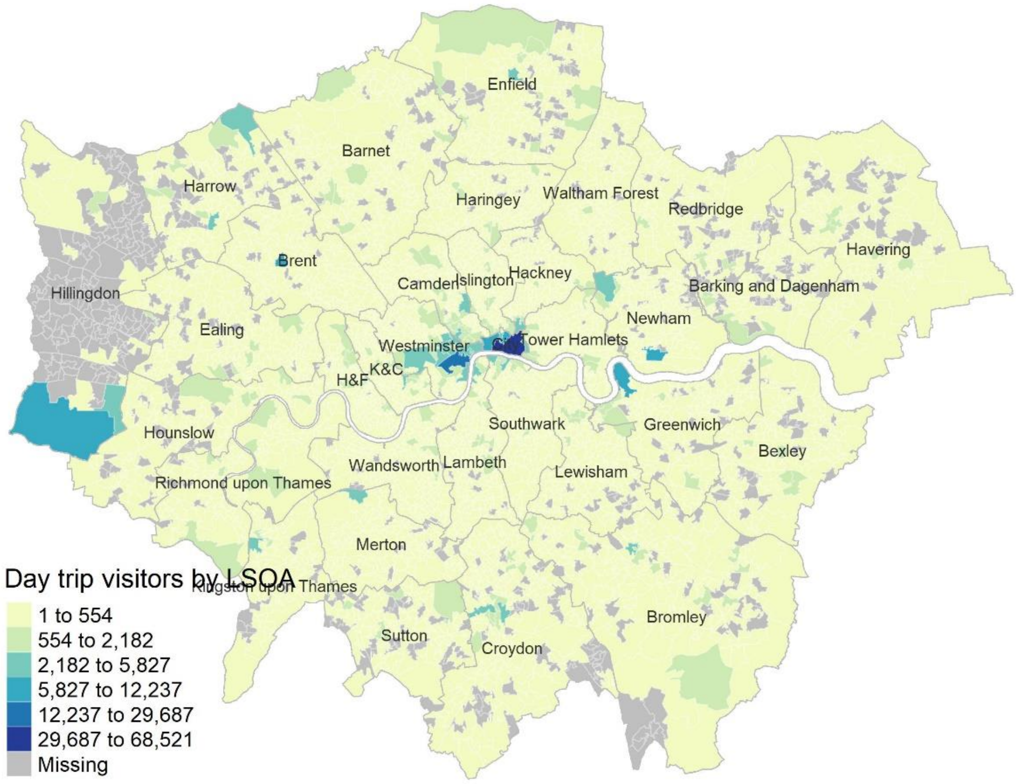


Figure B.3 Day trip visitor distribution in London.

Appendix C Supplementary notes for Chapter 6 (Paper III)

In this appendix, the model developments and calibration of the non-tourist SIM are detailed. The non-tourist SIM has been built on residential and workplace demand in London and is used to compare with the tourist SIM (seen in Chapter 6) to investigate the impacts of urban tourist demand on local grocery demand and potential supply.

C.1 Demand estimation

The conventional demand side of grocery shopping consists of expenditure from residents and workers in London.

C.1.1 Residential demand

The residential demand is segmented by household type according to the Output Area Classification (OAC). OAC is a UK geodemographic classification created from the 2011 census data by ONS at the Output Area (OA) level. It consists of 8 Supergroups, 26 Groups and 76 Subgroups (ONS, 2011). For each of the 26 OAC groups, ONS further reports their detailed household expenditure in their annual Living Costs and Food Survey (LCF) (see Table C.1) (ONS, 2017). This research uses the ‘Food and non-alcoholic’ expenditure released in the LCF 2017 as the expenditure rates of residential household spending per week.

There are 25,053 OAs in London in the 2011 census. Residential grocery demand in London is estimated at the OA level, using disaggregated expenditure rates from LCF by OAC group, alongside household counts capturing the number of residential households in each OA from the 2011 Census, and the OA level small-area geodemographic classification from the OAC. The produced demand estimate is associated with the residential population and is calculated as:

$$O_i^k = e^k n_i^k \tag{C.1}$$

Where:

O_i^k is a measure of the total residential grocery demand available in OA i by household type k;

e^k is a measure of the average weekly grocery expenditure for household type k, taking from the LCF survey;

n_i^k is the number of households of type k in OA i.

Table C.1 London OAC and household expenditure per week according to the LCF survey (ONS, 2019).

OAC group	Food & non-alcoholic drinks (£/week)	Description	OAC group	Food & non-alcoholic drinks (£/week)	Description
1A	0	Farming Communities	4C	69	Asian Traits
1B	0	Rural Tenants	5A	78.4	Urban Professionals and Families
1C	0	Ageing Rural Dwellers	5B	74.9	Ageing Urban Living
2A	0	Students Around Campus	6A	73.3	Suburban Achievers
2B	55.9	Inner-City Students	6B	76.2	Semi-Detached Suburbia
2C	18.5	Comfortable Cosmopolitans	7A	46.6	Challenged Diversity
2D	62	Aspiring and Affluent	7B	2.4	Constrained Flat Dwellers
3A	49.5	Ethnic Family Life	7C	0	White Communities
3B	54.4	Endeavouring Ethnic Mix	7D	22.1	Ageing City Dwellers
3C	36.8	Ethnic Dynamics	8A	52.2	Industrious Communities
3D	55.8	Aspirational Techies	8B	70.4	Challenged Terraced Workers
4A	63.8	Rented Family Living	8C	26.3	Hard-Pressed Ageing Workers
4B	64.4	Challenged Asian Terraces	8D	52.6	Migration and Churn

C.1.2 Workplace demand

The workplace demand is produced using the census-based Workplace Zones and Workplace Zone Statistics (WZS) in London published by ONS (2014). Workplace Zones are a small-area geography created from the 2011 Census, optimised to capture workplace population based upon the places of work (ONS, 2014). Therefore, the WZS gives greater geographical precision to the location of workplace populations than residence-based geographies. In addition, since no specific survey reports the average grocery expenditure of workers in London, the research follows the advice of Waddington et al. (2019) and uses £5 per person as the expenditure rate of workers in London. Therefore, workplace demand can be calculated as:

$$O_w = en_w \tag{C.2}$$

Where:

O_w is a measure of the total worker grocery demand available in Workplace Zone w;

e is assumed as £5 for the expenditure rate of workers in London;

n_w is the number of workers in the Workplace Zone w.

As such, combined with the tourist demand as estimated in Chapter 5, the three parts of the disaggregated demand side of grocery shopping in London is summed up as in Table C.2. The residential, workplace and tourism demand is fed into the SIM separately and the model parameters are calibrated separately too. Thus their unique impacts on the flows and revenues can be clearly seen.

Table C.2 The three parts of disaggregated grocery demand in London.

	Residential	**Workplace**	**Tourism**
Spatial unit	OA	Workplace Zone	LSOA
Counts of unit	25,053	8,154	4,835
Expenditure rates	Disaggregated by 26 OAC group (Table C.1)	£5	Disaggregated by four tourist types (Table 5.6)
Expenditure estimates (£ millions)	199.8	25.8	14.06

C.2 Supply side

The model allocates the estimated demand to 1,759 separate stores that fall within Greater London. These stores include all supermarkets and convenience stores from the newly released Geolytix Retail Point dataset in 2020. However, the Geolytix dataset only offers the floorspace of the stores at the band level (<3,013 sqft; 3,013 sqft ~ 15,069 sqft; 15,069 sqft ~ 30,138 sqft; >30,138 sqft) instead of the actual size. Therefore, the research utilised the CACI grocery outlets data (up to 2014) to infer the actual size of these stores based on the geographical coordinates of the store points. For the store built after 2014, their size is estimated by the average size of stores in the same band of floorspace but whose actual size is available from CACI 2014 dataset. Besides the major ten retailer brands, other brands such as Booths, Costco, Whole Foods Market, Budgens, etc., are combined as 'Others'. The spatial distribution of the grocery stores in London and their aggregated floorspace by LSOA is presented as in Figure 5.10. The brands and detailed information are listed in Table C.3.

Table C.3 10 main grocery brands in Greater London and detailed information.

Brand	Count of stores	Sum of floorspace (sqft)	Floorspace share in London (%)
Aldi	39	463,202	2.39
Asda	43	1,230,611	9.10
Co-Op	265	863,604	6.32
Iceland	124	809,354	4.83
Lidl	90	1,020,584	5.64
M&S	174	1,736,572	5.25
Morrisons	30	900,875	6.57
Sainsbury's	367	3,631,624	4.81
Tesco	466	3,194,173	24.65
Waitrose	92	1,298,820	21.16
Others	69	752,647	9.28
Sum	**1759**	**15,902,065**	**100.00**

C.3 Model development

Based on the demand estimated above, the non-tourist SIM built upon the residential and workplace demand in London is represented as:

$$S_{non-tourist} = S_{resi} + S_{work} = S_{ij}^{kn} + S_{wj}^{n} = A_i^k O_i^k W_j^{\alpha^n} exp^{(-\beta^k c_{ij})} + A_w O_w W_j^{\alpha^n} exp^{(-\beta c_{wj})} \quad \text{(C.3)}$$

Where:

S_{ij}^{kn} represents the predicted expenditure flow between residential demand zone OA i and store j of brand n by residential household type k; O_i^k is the total residential demand in OA i by household k as measured in Section C.1.1.

S_{wj}^{n} represents the predicted expenditure flow between Workplace Zone w and store j of brand n; O_w is the total workplace demand in Workplace Zone w as measured as in Section C.1.2;

W_j is the overall attractiveness of store j, and α^n represents the additional or perceived relative attractiveness of store j by store brand n; C_{ij} and C_{wj} are the distance from origin zone i or w to store j, and incorporate the distance decay parameter $exp^{-\beta^k}$ for household type k and $exp^{-\beta}$ for workplace demand;

A_i^k and A_w are the balancing factors that take account of competition and ensure that all demand from OA i by household type k or Workplace Zone w is allocated to stores within London. The balancing factors thus ensure that: $\sum_j S_{ij}^{kn} = O_i^k$ and $\sum_j S_{wj}^{n} = O_w$. They are respectively calculated as: $A_i^k = \frac{1}{\sum_j W_j^{\alpha^n} exp^{\left(-\beta^k C_{ij}\right)}}$ and $A_i^b = \frac{1}{\sum_j W_j^{\alpha^n} exp^{\left(-\beta C_{wj}\right)}}$. (C.4; C.5)

C.4 Model calibration

C.4.1 Distance decay parameter β

The goal of calibrating the distance deterrence parameter β is to replicate the average trip distance (ATD) of the consumers. For the residents of each OAC supergroup k, β^k is calibrated with the corresponding observed ATD shown in Table C.4. These observed ATD are calculated by the mean distance from each OA demand zone to the nearest three supermarkets. The β value of OAC supergroup 1 is empty because the LCF reports no expenditure on ‘Food & non-alcoholic drinks’ for this group in London and the only 10 OAs in London belong to this supergroup are not included. Meanwhile, the β value of the workplace SIM is calibrated according to the average trip length by walking mode for eat/drink purpose which is 0.9656 km (0.6 miles) in England as reported by the National Travel Survey of London (National Travel Survey (NTS), 2017).

Table C.4 Beta values for the disaggregate model.

	Demand	β	Predicted ATD	Observed ATD	ATD pred./ATD ob.
resident (OAC SPRGRP)	2	2.983	0.7380	0.738	1.0000
	3	2.855	0.8000	0.800	1.0000
	4	2.292	1.0802	1.080	0.9999
	5	1.973	1.1999	1.200	0.9999
	6	1.916	1.3994	1.400	0.9999
	7	1.899	1.2198	1.220	1.0002
	8	1.966	1.2893	1.290	1.0001
worker		2.208	0.9656	0.9656	1.0000
tourist	Airbnb	7.543	0.5520	0.552	0.9999
	Hotel	5.796	0.5260	0.526	0.9999
	Free guest/Own home	0.952	2.0941	2.0934	1.0003
	Day visitor	1.586	1.353	1.353	1.0001

C.4.2 Relative attractiveness parameter α

Next, we calibrate α according to the market share in Table C.5. Although Kantar monitors market shares by major grocery retailers in Greater Britain (Kantar, 2021), the observed market share of grocery retailers in London has not been reported publicly. Obviously, the Kantar national market shares are not suitable to be taken as observed market shares to be used in the calibration of the London model. This research uses the same approach of Waddington (2017) to estimate the regional market share of grocery retailers by calculating the market share per sqft. of floorspace (V_b) using the national figures. This method is also in line with the corroborated relationship between brand's presence level and market share in prior studies (Hughes et al., 2009; Thompson, 2013). As such, the regional market share ($Nmksh_b$) is calculated as:

$$V_b = \frac{Nmksh_b}{Nflsp_b} \tag{C.6}$$

$$Nprop_b = V_b * \ Rflsp_b \tag{C.7}$$

$$Nmksh_b = \left(\frac{Nprop_b}{\sum_n Nprop_b}\right) * 100 \tag{C.8}$$

Where:

V_b represents the market share value of one sqft of floorspace for brand b, which is calculated by the Kantar national market share $Nmksh_b$ and the reported national total floorspace $Nflsp_b$ of each brand b. The $Nflsp_b$is estimated via the same abovementioned method based on CACI 2014 and Geolytix 2020 datasets;

$Nprop_b$ is the proportion of regional market share making up national market share for brand b, which is measured by the brand market share value V_b along with the regional floorspace $Rflsp_b$;

$Nmksh_b$ is the estimated brand regional market share for each brand b of all the n brand, which is upscaled from $Nprop_b$. $Nmksh_b$ is then used to calibrate the residential SIM.

The estimated market share of the ten main grocery retailers in London is shown in Table C.5. The estimated regional market shares of several major retailers in London have substantial differences compared to the national market shares. For example, Aldi, Asda, Co-Op and Morrisons have a much lower value, whereas Sainsbury’s and Waitrose have a higher market share in London. These estimated market shares have been used to calibrate the relative attractiveness parameter α for the residential SIM. The predicted regional market share of residential SIM and the goodness-of-fit following calibration on α in the SIM are shown in Table C.6.

Table C.5 Estimated regional market shares in London.

Brand	**National market share $Nmksh_b$ (%)**	**National floorspace $Nflsp_b$ (sqft)**	**London proportion of national market share $Nprop_b$ (%)**	**Estimated London regional market share $Nmksh_b$ (%)**
Aldi	8.20	8,230,193	0.46	4.59
Asda	15.00	19,483,528	0.95	9.42
Co-Op	6.20	14,301,185	0.37	3.72
Iceland	2.20	6,865,431	0.26	2.58
Lidl	6.10	8,809,931	0.71	7.03
M&S	3.20	11,441,413	0.49	4.83
Morrisons	10.00	14,879,456	0.61	6.02

Sainsbury's	15.30	22,658,026	2.45	24.39
Tesco	26.80	33,037,051	2.59	25.77
Waitrose	5.00	6,307,335	1.03	10.24
Others	2.00	10,550,273	0.14	1.42
Sum	**100.00**	**156,563,822**	**10.06**	**100.00**

After calibrating the α value of the residential SIM, the workplace SIM follows the same attractiveness parameter α, which means the brand attractiveness to the workers remains the same as to the residents. The workplace SIM outputs in Table C.6 shows that when α remains the same, Aldi, Asda, Lidl, Iceland and Morrisons are observed to have lower predicted workplace market shares than the predicted residential market shares in London, whereas the market shares of M&S and Waitrose show a significant rise. This may be due to the format of the stores since the workers are more likely to visit convenience stores during their work break time. As noted in Chapter 6, the α value for the tourist SIM is 1.0 without further disaggregation. Consequently, the market shares of the residential, workplace and tourist SIM are reported as in Table C.6.

Table C.6 The predicted market share of each disaggregated SIM.

Brand	predicted residential market share (%)	predicted workplace market share (%)	Predicted tourist market share (%)	Predicted overall market share (%)	Estimated Market share in London (%)
Aldi	4.56	2.34	1.32	4.13	4.59
Asda	9.50	7.59	5.03	9.04	9.42
Co-Op	3.80	2.95	8.83	4.01	3.72
Iceland	2.85	1.89	4.65	2.85	2.58
Lidl	7.17	4.21	3.70	6.65	7.03
M&S	4.85	11.21	9.61	5.82	4.83
Morrisons	6.34	4.00	3.92	5.95	6.02
Sainsburys	24.32	22.75	24.21	24.14	24.39
Tesco	25.09	24.12	22.27	24.82	25.77
Waitrose	10.10	17.57	12.42	11.04	10.24
Others	1.42	1.36	4.03	1.57	1.42
Goodness-of-fit for the calibration of market share in residential SIM Correlation: 0.99975; R^2=0.99950					

C.5 Model validation

Since there is no empirical data of individual store sales, it is not possible to validate the robustness of the custom-built SIM. This research, however, endeavours to make use of industry reported brand performance data. For example, in Colliers International (2018), the UK supermarket investment review mentioned that the current trading intensity of the ASDA Mitcham supermarket is around £15 sqft. In the SIMs built in this thesis, the sale density of this store is £15.31 in the non-tourist SIM and £15.65 when adding the tourist SIM. This result shows that in the ASDA Mitcham store case at least, the model outputs are very close to the industry report.

For each of the stores in the model, if any empirical performance data can be accessed, it will enable the examination of the validity of the SIM. Any evidence suggesting the tourist contribution to store revenue will also be valuable to validate whether the tourist SIM accurately reflects the impact of tourists at the individual store level.

List of References

Abbasi, A., Rashidi, T.H., Maghrebi, M. and Waller, S.T. 2015. Utilising Location Based Social Media in Travel Survey Methods *In*: *Proceedings of the 8th ACM SIGSPATIAL International Workshop on Location-Based Social Networks* [Online]. New York: ACM Press, pp.1–9. Available from: http://dl.acm.org/citation.cfm?doid=2830657.2830660.

Adelfio, M., Serrano-Estrada, L., Martí, P., Kain, J.H. and Stenberg, J. 2020. Social Activity in Gothenburg's Intermediate City: Mapping Third Places through Social Media Data. *Applied Spatial Analysis and Policy*. **13**(4), pp.985–1017.

Agryzcov, T., Nolasco-Cirugeda, A., Oliver, J.L., Serrano-Estrada, L., Tortosa, L. and Vicent, J.F. 2015. Using data from Foursquare Web Service to represent the commercial activity of a city. *International Journal of Computer, Control, Quantum and Information Engineering. World Academy of Science, Engineering and Technology*. **9**(1), pp.69–76.

Agryzcov, T., Tortosa, L., Vicent, J.F. and Wilson, R. 2019. A centrality measure for urban networks based on the eigenvector centrality concept. *Environment and Planning B: Urban Analytics and City Science*. **46**(4), pp.668–689.

Airbnb 2018. Airbnb UK Insights Report. , pp.1–56. [Accessed 13 September 2019]. Available from: https://www.airbnbcitizen.com/wp-content/uploads/2018/10/AirbnbUKInsightsReport_2018.pdf.

Ashworth, G. and Page, S.J. 2011. Urban tourism research: Recent progress and current paradoxes. *Tourism Management*. **32**(1), pp.1–15.

Ashworth, G.J. 2009. Questioning the urban in urban tourism *In*: G. Maciocco and S. Serreli, eds. *Enhancing the City: New Perspectives for Tourism and Leisure*. New York: Springer, pp.207–220.

Back, A. 2020. Temporary resident evil? Managing diverse impacts of second-home tourism. *Current Issues in Tourism*. **23**(11), pp.1328–1342.

Baeza-Yates, R. 2020. Biases on Social Media Data: (Keynote Extended Abstract) *In*: *The Web Conference 2020 - Companion of the World Wide Web Conference, WWW 2020* [Online]. ACM, pp.782–783. [Accessed 16 January 2022]. Available from: https://doi.org/10.1145/3366424.3383564.

Bahrehdar, A.R., Adams, B. and Purves, R.S. 2020. Streets of London: Using Flickr and OpenStreetMap to build an interactive image of the city. *Computers, Environment and Urban Systems*. **84**.

Barchiesi, D., Moat, H.S., Alis, C., Bishop, S. and Preis, T. 2015. Quantifying international travel flows using Flickr. *PLoS ONE*. **10**(7), pp.1–8.

Batista e Silva, F., Freire, S., Schiavina, M., Rosina, K., Marín-Herrera, M.A., Ziemba, L., Craglia, M., Koomen, E. and Lavalle, C. 2020. Uncovering temporal changes in Europe’s population density patterns using a data fusion approach. *Nature Communications*. **11**(1).

Batista e Silva, F., Marín Herrera, M.A., Rosina, K., Ribeiro Barranco, R., Freire, S. and Schiavina, M. 2018. Analysing spatiotemporal patterns of tourism in Europe at high-resolution with conventional and big data sources. *Tourism Management*. **68**, pp.101–115.

Batty, M. 2008. Spatial interaction *In*: K. Kemp, ed. *Encyclopedia of Geographic Information Science*. Thousand Oaks: SAGE, pp.416–418.

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A. and Urquhart, R. 2021. Incorporating E-commerce into Retail Location Models. *Geographical Analysis*., gean.12285.

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

De Beule, M., Van den Poel, D. and Van de Weghe, N. 2014. An extended Huff-model for robustly benchmarking and predicting retail network performance. *Applied Geography*. **46**, pp.80–89.

Bhaduri, B. 2008. Population Disitribution During the Day *In*: S. Shekhar and H. Xiong, eds. *Encyclopedia of GIS*. New York: Springer-Verlag, pp.880–885.

Birkin, M. 2019. Spatial data analytics of mobility with consumer data. *Journal of Transport Geography*. **76**(7), pp.245–253.

Birkin, M., Clarke, G. and Clarke, M. 2010. Refining and operationalizing entropy-maximizing models for business applications. *Geographical Analysis*. **42**(4), pp.422–445.

Birkin, M., Clarke, G.P. and Clarke, M. 2017. *Retail Location Planning in an Era of Multi-Channel Growth* [Online]. London: Routledge. Available from: https://www.taylorfrancis.com/books/9781317064541.

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., Harland, K. and Malleson, N. 2013. The classification of space-time behaviour patterns in a British city from crowd-sourced data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. **7974 LNCS**(PART 4), pp.179–192.

Birkin, M. and Heppenstall, A. 2011. Extending Spatial Interaction Models with Agents for Understanding Relationships in a Dynamic Retail Market. *Urban Studies Research*. **2011**(1986), pp.1–12.

Brown, G., Lee, I.S., King, K. and Shipway, R. 2015. Eventscapes and the creation of event legacies. *Annals of Leisure Research*. **18**(4), pp.510–527.

Brunsdon, C. and Comber, A. 2020. Big issues for big data: challenges for critical spatial data analytics. *Journal of Spatial Information Science*. (21).

Buning, R.J. and Lulla, V. 2020. Visitor bikeshare usage: tracking visitor spatiotemporal behavior using big data. *Journal of Sustainable Tourism*. **29**(4), pp.711–731.

Bustamante, A., Sebastia, L. and Onaindia, E. 2019. Can Tourist Attractions Boost Other Activities Around? A Data Analysis through Social Networks. *Sensors*. **19**(11), p.2612.

Charles-Edwards, E. and Bell, M. 2015. Seasonal flux in Australia's population geography: Linking space and time. *Population, Space and Place*. **21**(2), pp.103–123.

Chen, J., Becken, S. and Stantic, B. 2020. Using Weibo to track global mobility of Chinese visitors. *Annals of Tourism Research*., p.103078.

Chen, M., Arribas-Bel, D. and Singleton, A. 2019. Understanding the dynamics of urban areas of interest through volunteered geographic information. *Journal of Geographical Systems*. **21**(1), pp.89–109.

Chen, Y., Mahmassani, H.S. and Frei, A. 2018. Incorporating social media in travel and activity choice models: conceptual framework and exploratory analysis. *International Journal of Urban Sciences*. **22**(2), pp.180–200.

Cheng, Z., Caverlee, J., Lee, K. and Sui, D.Z. 2011. Exploring Millions of Footprints in Location Sharing Services. *Icwsm*. **2010**(Cholera), pp.81–88.

Choi, M.J., Heo, C.Y. and Law, R. 2016. Progress in Shopping Tourism. *Journal of Travel and Tourism Marketing*. **33**(1), pp.S1–S24.

Chua, A., Servillo, L., Marcheggiani, E. and Moere, A. Vande 2016. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tourism Management*. **57**, pp.295–310.

Clarke, G., Eyre, H. and Guy, C. 2002. Deriving indicators of access to food retail provision in British cities: Studies of Cardiff, Leeds and Bradford. *Urban Studies*. **39**(11), pp.2041–2060.

Clarke, G.P. 2020. Regional science in business *In*: M. M. Fischer and P. Nijkamp, eds. *Handbook of Regional Science*. Berlin, Heidelberg: Springer, pp.129–139.

Cocola-gant, A. 2018. *Struggling with the leisure class: Tourism, gentrification and displacement*.[Online] Available from: http://e-journal.uajy.ac.id/14649/1/JURNAL.pdf.

Colliers International 2018. *UK SUPERMARKET INVESTMENT REVIEW*.

Comito, C., Falcone, D. and Talia, D. 2016. Mining human mobility patterns from social geo-tagged data. *Pervasive and Mobile Computing*. **33**, pp.91–107.

Cromarty, H. and Barton, C. 2018. *The growth in short-term lettings (England)* [Online]. Available from: www.parliament.uk/commons-library%7Cintranet.parliament.uk/commons-library%7Cpapers@parliament.uk%7C@commonslibrary.

D'Silva, K., Kasthuri, J., Noulas, A., Mascolo, C. and Misra, A. 2018. The Role of Urban Mobility in Retail Business Survival. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*. **2**(3), p.100.

Daggitt, M.L., Noulas, A., Shaw, B. and Mascolo, C. 2016. Tracking urban activity growth globally with big location data. *Royal Society Open Science*. **3**(4), p.150688.

Davies, A., Dolega, L. and Arribas-Bel, D. 2019. Buy online collect in-store: exploring grocery click&collect using a national case study. *International Journal of Retail and Distribution Management*. **47**(3), pp.278–291.

Deane, S. 2021. 2021 Airbnb Statistics: Usage, Demographics, and

Revenue Growth. *Stratosjets*. [Online]. [Accessed 17 January 2022]. Available from: https://www.stratosjets.com/blog/airbnb-statistics/.

Department for Digital Culture Media & Sport (DCMS) 2021. *The Tourism Recovery Plan* [Online]. [Accessed 1 September 2021]. Available from: http://www.nationalarchives.gov.uk/doc/.

Dickinson, J.E., Lumsdon, L.M. and Robbins, D. 2011. Slow travel: issues for tourism and climate change. *Journal of Sustainable Tourism*. **19**(3), pp.281–300.

Doan, T.-N. and Lim, E.-P. 2016. Attractiveness versus Competition: Towards an Unified Model for User Visitation *In*: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management* [Online]. New York: ACM Press, pp.2149–2154. Available from: http://dl.acm.org/citation.cfm?doid=2983323.2983657.

Doan, T.-N. and Lim, E.-P. 2019. Modeling location-based social network data with area attraction and neighborhood competition. *Data Mining and Knowledge Discovery*. **33**(1), pp.58–95.

Dogru, T., Mody, M. and Suess, C. 2019. Adding evidence to the debate: Quantifying Airbnb’s disruptive impact on ten key hotel markets. *Tourism Management*. **72**, pp.27–38.

Dolega, L., Pavlis, M. and Singleton, A.D. 2016. Estimating attractiveness, hierarchy and catchment area extents for a national set of retail centre agglomerations. *Journal of Retailing and Consumer Services*. **28**, pp.78–90.

Edwards, D., Griffin, T. and Hayllar, B. 2008. Urban Tourism Research. Developing an Agenda. *Annals of Tourism Research*. **35**(4), pp.1032–1052.

Fekete, E. 2015. Race and (Online) Sites of Consumption. *Geographical Review*. **105**(4), pp.472–491.

Ferreira, A.P.G., Silva, T.H. and Loureiro, A.A.F. 2015. Beyond Sights: Large Scale Study of Tourists' Behavior Using Foursquare Data *In*: *2015 IEEE International Conference on Data Mining Workshop* [Online]. IEEE, pp.1117–1124. Available from: http://ieeexplore.ieee.org/document/7395793/.

Ferreira, A.P.G., Silva, T.H. and Loureiro, A.A.F. 2020. Uncovering spatiotemporal and semantic aspects of tourists mobility using social sensing. *Computer Communications*. **160**, pp.240–252.

Ferreri, M. and Sanyal, R. 2018. Platform economies and urban planning: Airbnb and regulated deregulation in London. *Urban Studies*. **55**(15), pp.3353–3368.

Filieri, R. 2015. Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. *Tourism Management*. **51**, pp.174–185.

Fotheringham, A.S. 2017. Spatial interaction *In*: D. Richardson, N. Castree and M. Goodchild, eds. *International Encyclopedia of Geography*. Atlanta, GA: Wiley.

Foursquare 2019. Foursquare. *Foursquare Labs Inc*.

Franklin, A. and Crang, M. 2001. The trouble with tourism and travel theory? *Tourist Studies*. **1**(1), pp.5–22.

Freytag, T. and Bauder, M. 2018. Bottom-up touristification and urban transformations in Paris. *Tourism Geographies*. **20**(3), pp.443–460.

García-Palomares, J.C., Salas-Olmedo, M.H., Moya-Gómez, B., Condeço-Melhorado, A. and Gutiérrez, J. 2018. City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities*. **72**(2018), pp.310–319.

Georgiev, P., Noulas, A. and Mascolo, C. 2014. Where Businesses Thrive:

Predicting the Impact of the Olympic Games on Local Retailers through Location-based Services Data *In*: *Eighth International AAAI Conference on Weblogs and Social Media* [Online]., pp.151–160. Available from: http://arxiv.org/abs/1403.7654.

Giglio, S., Bertacchini, F., Bilotta, E. and Pantano, P. 2019. Machine learning and points of interest: typical tourist Italian cities. *Current Issues in Tourism*., pp.1–13.

Girardin, F., Calabrese, F., Fiore, F.D., Ratti, C. and Blat, J. 2008. Digital Footprinting: Uncovering Tourists with User-Generated Content. *IEEE Pervasive Computing*. **7**(4), pp.36–43.

Girardin, F., Fiore, F., Ratti, C. and Blat, J. 2008. Leveraging explicitly disclosed location information to understand tourist dynamics: A case study. *Journal of Location Based Services*. **2**(1), pp.41–56.

Girardin, F., Vaccari, A., Gerber, A., Biderman, A. and Ratti, C. 2009. Quantifying urban attractiveness from the distribution and density of digital footprints. *International Journal of Spatial Data Infrastructures Research*. **4**, pp.175–200.

Great Britain Day Visitor Survey (GBDVS) 2018. The Great Britain Day Visitor 2018 Annual Report. , p.254. [Accessed 10 February 2020]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/England-documents/260139488_-_kantar_tns_-_gbdvs_2017_annual_report_v5r.pdf.

Great Britain Day Visits Survey (GBDVS) 2020. *The Great Britain Day Visitor 2019 Annual Report* [Online]. [Accessed 20 March 2021]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/gbdvs_2019_annual_report_-_a.pdf.

Great Britain Tourism Survey (GBTS) 2020. *The GB Tourist 2019 Annual Report* [Online]. [Accessed 19 March 2021]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/gb_tourist_annual_report_2019_final.pdf.

Great Britain Tourism Survey （GBTS） 2019. Online Data Browser: Great Britain domestic overnight trips. [Accessed 16 January 2020]. Available from: https://gbtsenglandlightviewer.kantar.com/ViewTable.aspx.

Great London Authority (GLA) 2021. *The London Plan* [Online]. [Accessed 20 March 2021]. Available from: www.london.gov.uk.

Greater London Authority (GLA) 2017. *AirBnB submission: Draft London Plan*.

Greater London Authority (GLA) 2014. Daytime Population of London 2014. *Greater London Authority (GLA)*. [Online]. [Accessed 19 July 2019]. Available from: https://data.london.gov.uk/dataset/daytime-population-borough.

Greater London Authority (GLA) and Creative Tourist Consultants 2015. *Take A Closer Look: A Cultural Tourism Vision for London, 2015-2017* [Online]. Available from: https://www.london.gov.uk/sites/default/files/cultural_tourism_vision_for_london_low_res_version.pdf.

Grinberger, A.Y., Shoval, N. and McKercher, B. 2014. Typologies of tourists' time-space consumption: a new approach using GPS data and GIS tools. *Tourism Geographies*. **16**(1), pp.105–123.

Gunter, U. and Önder, I. 2021. An Exploratory Analysis of Geotagged Photos From Instagram for Residents of and Visitors to Vienna. *Journal of Hospitality and Tourism Research*. **45**(2), pp.373–398.

Gutiérrez, J., García-Palomares, J.C., Romanillos, G. and Salas-Olmedo, M.H. 2017. The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism Management*. **62**, pp.278–291.

Guttentag, D. 2015. Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*. **18**(12), pp.1192–1217.

Hall, C.M. 2014. Second Home Tourism: An International Review. *Tourism Review International*. **18**(3), pp.115–135.

Hallyar, B., Griffin, T. and Edwards, D. 2008. *Urban tourism precincts*. Oxford: Elsevier.

Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T. and Kremer, P. 2018. Geolocated social media as a rapid indicator of park visitation and equitable park access. *Computers, Environment and Urban Systems*. **72**, pp.38–50.

Hardy, A. 2003. An investigation into the key factors necessary for the development of iconic touring routes. *Journal of Vacation Marketing*. **9**(4), pp.314–330.

Hargittai, E. 2020. Potential Biases in Big Data: Omitted Voices on Social Media. *Social Science Computer Review*. **38**(1), pp.10–24.

Hasan, S. and Ukkusuri, S. V. 2015. Location contexts of user check-ins to model urban geo life-style patterns. *PLoS ONE*. **10**(5).

Hasnat, M.M. and Hasan, S. 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies*. **96**, pp.38–54.

Hauff, C. 2013. A study on the accuracy of Flickr's geotag data *In*: *SIGIR 2013 - Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* [Online]., pp.1037–1040. [Accessed 1 January 2022]. Available from: http://www.flickr.com/.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V. and Di Minin, E. 2018. Social Media Data Can Be Used to Understand Tourists' Preferences for Nature-Based Experiences in Protected Areas. *Conservation Letters*. **11**(1).

Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P. and

Ratti, C. 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*. **41**(3), pp.260–271.

Hecht, B. and Stephens, M. 2014. A tale of cities: Urban biases in volunteered geographic information *In*: *Proceedings of the 8th International Conference on Weblogs and Social Media*. AAAI Press, pp.197–205.

Hollenstein, L. and Purves, R. 2010. Exploring Place through User-Generated Content: Using Flickr to Describe City Cores. *Journal of Spatial Information Science*. **1**(1), pp.21–48.

Hsieh, H.P., Lin, F., Li, C. Te, Yen, I.E.H. and Chen, H.Y. 2019. Temporal popularity prediction of locations for geographical placement of retail stores. *Knowledge and Information Systems*. **60**(1), pp.247–273.

Hu, F., Li, Z., Yang, C. and Jiang, Y. 2018. A graph-based approach to detecting tourist movement patterns using social media data. *https://doi.org/10.1080/15230406.2018.1496036*. **46**(4), pp.368–382.

Hu, W. and Jin, P.J. 2017. An adaptive hawkes process formulation for estimating time-of-day zonal trip arrivals with location-based social networking check-in data. *Transportation Research Part C: Emerging Technologies*. **79**, pp.136–155.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W. and Prasad, S. 2015. Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*. **54**, pp.240–254.

Huang, Q. and Wong, D.W.S. 2016. Activity patterns, socioeconomic status and urban spatial structure: what can social media data tell us? *International Journal of Geographical Information Science*. **30**(9), pp.1873–1898.

Huang, X.-T. and Wu, B.-H. 2012. Intra-attraction Tourist Spatial-Temporal Behaviour Patterns. *Tourism Geographies*. **14**(4), pp.625–645.

Huff, D.L. 1963. *A Probabilistic Analysis of Shopping Center Trade Areas* [Online]. Available from: https://www.jstor.org/stable/3144521.

Hughes, R., Hallsworth, A.G. and Clarke, G. 2009. Testing the effectiveness of the proposed UK 'competition test'. *Service Industries Journal*. **29**(5), pp.569–590.

Inkson, C. 2019. Unplanned Expansions: Renting Private Homes to Tourists *In*: A. Smith and A. Graham, eds. *Destination London: The Expansion of the Visitor Economy* [Online]. University of Westminster Press, pp.37–59. Available from: https://www.jstor.org/stable/j.ctvhrd0t9.6.

Ioannides, D., Röslmaier, M. and van der Zee, E. 2019. Airbnb as an instigator of 'tourism bubble' expansion in Utrecht's Lombok neighbourhood. *Tourism Geographies*. **21**(5), pp.822–840.

iPropertymanagement 2020. Airbnb Statistics 2020 : User & Market Growth Data. *iPropertyManagement.com*.

IPS 2019. Inbound accommodation research. [Accessed 4 February 2020]. Available from: https://www.visitbritain.org/inbound-accommodation-research.

Jansson, A. 2019. The mutual shaping of geomedia and gentrification: The case of alternative tourism apps. *Communication and the Public*. **4**(2), pp.166–181.

Jia, S. (Sixue) 2020. Motivation and satisfaction of Chinese and U.S. tourists in restaurants: A cross-cultural text mining of online reviews. *Tourism Management*. **78**, p.104071.

Jiang, Y., Li, Z. and Ye, X. 2019. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. *Cartography and Geographic Information Science*. **46**(3), pp.228–242.

Jin, C. and Xu, J. 2018. Using user-generated content data to analyze tourist mobility between hotels and attractions in cities. *Environment and*

Planning B: Urban Analytics and City Science.

Joseph, K., Tan, C.H. and Carley, K.M. 2012. Beyond ‘local’, ‘categories’ and ‘friends’ *In*: *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* [Online]. New York: ACM Press, pp.919–926. Available from: http://dl.acm.org/citation.cfm?id=2370216.2370422.

Judd, D.R. 2003. Urban tourism and the geography of the city. *EURE*. **29**(87), pp.51–62.

Kádár, B. 2014. Measuring tourist activities in cities using geotagged photography. *Tourism Geographies*. **16**(1), pp.88–104.

Kang, S., Lee, G., Kim, J. and Park, D. 2018. Identifying the spatial structure of the tourist attraction system in South Korea using GIS and network analysis: An application of anchor-point theory. *Journal of Destination Marketing and Management*. **9**, pp.358–370.

Kantar 2021. Great Britain - Grocery Market Share. *Kantar Worldpanel*. [Online]. Available from: https://www.kantarworldpanel.com/global/grocery-market-share/great-britain.

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]. New York: ACM Press, p.793. Available from: http://arxiv.org/abs/1306.1704%0Ahttp://dx.doi.org/10.1145/2487575.2487616.

Kesorn, K., Juraphanthong, W. and Salaiwarakul, A. 2017. Personalized Attraction Recommendation System for Tourists Through Check-In Data. *IEEE Access*. **5**, pp.26703–26721.

Khan, N.U., Wan, W. and Yu, S. 2020. Spatiotemporal analysis of tourists and residents in Shanghai based on location-based social network’s data from WeiBo. *ISPRS International Journal of Geo-Information*. **9**(2), p.70.

Kim, Y., Kim, C. ki, Lee, D.K., Lee, H. woo and Andrada, R.I.T. 2019. Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*. **72**, pp.249–256.

Kitchin, R. 2017. Big data—Hype or revolution *In*: L. Sloan and A. Quan-Haase, eds. *The SAGE Handbook of Social Media Research Methods*. SAGE, p.728.

Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography*. **3**(3), pp.262–267.

Kotiloglu, S., Lappas, T., Pelechrinis, K. and Repoussis, P.P. 2017. Personalized multi-period tour recommendations. *Tourism Management*. **62**, pp.76–88.

Kowalski, Ł. 2019. Comparing spatial-interaction and hybrid agent-based modelling approaches: An application to location analysis of services. *JASSS*. **22**(1).

Lansley, G., Goodchild, M. and Longley, P. 2018. Big Data and Geospatial Analysis. *In: De Smith, M and Goodchild, M and Longley, P, (eds.) Geospatial Analysis: A comprehensive guide to principles, techniques and software tools. (pp. 547-570). The Winchelsea Press: Edinburgh. (2018)*., pp.547–570.

Lansley, G. and Longley, P. a. 2016a. Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services*. **30**, pp.271–278.

Lansley, G. and Longley, P. a. 2016b. The geography of Twitter topics in London. *Computers, Environment and Urban Systems*. **58**, pp.85–96.

Law, C. 2002. *Urban Tourism: The visitor economy and the growth of large cities*. London: Continuum.

Lee, D. 2016. How Airbnb short-term rentals exacerbate Los Angeles's affordable housing crisis: Analysis and policy recommendations. *Harvard Law & Policy Review*. **10**, pp.229–253.

Lew, A. and McKercher, B. 2006. Modeling tourist movements: A local destination analysis. *Annals of Tourism Research*. **33**(2), pp.403–423.

Li, D., Zhou, X. and Wang, M. 2018. Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities*. **74**, pp.249–258.

Li, J., Xu, L., Tang, L., Wang, S. and Li, L. 2018. Big data in tourism research: A literature review. *Tourism Management*. **68**, pp.301–323.

Li, L., Goodchild, M.F. and Xu, B. 2013. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*. **40**(2), pp.61–77.

Li, Y., Steiner, M., Wang, L., Zhang, Z.-L. and Bao, J. 2013. Exploring venue popularity in Foursquare *In*: *2013 Proceedings IEEE INFOCOM* [Online]. IEEE, pp.3357–3362. Available from: http://ieeexplore.ieee.org/document/6567164/.

Lian, D. and Xie, X. 2011. Collaborative activity recognition via check-in history *In*: *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks - LBSN '11*.

Lim, K.H., Chan, J., Karunasekera, S. and Leckie, C. 2018. Tour recommendation and trip planning using location-based social media: a survey. *Knowledge and Information Systems*.

Liu, B., Huang, S. (Sam) and Fu, H. 2017. An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management*. **58**, pp.132–141.

Liu, X. and Wang, J. 2015. The geography of Weibo. *Environment and Planning A*. **47**(6), pp.1231–1234.

Liu, Y., Sui, Z., Kang, C. and Gao, Y. 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE*. **9**(1).

Liu, Z., Zhou, X., Shi, W. and Zhang, A. 2019. Recommending attractive thematic regions by semantic community detection with multi-sourced VGI data. *International Journal of Geographical Information Science*. **33**(8), pp.1520–1544.

Lloyd, A. and Cheshire, J. 2017. Deriving retail centre locations and catchments from geo-tagged Twitter data. *Computers, Environment and Urban Systems*. **61**, pp.108–118.

London and Partners 2017. *A tourism vision for London* [Online]. [Accessed 12 September 2020]. Available from: http://files.londonandpartners.com/l-and-p/assets/london_tourism_vision_aug_2017.pdf.

Long, X., Jin, L. and Joshi, J. 2013. Understanding Venue Popularity in Foursquare *In*: *Proceedings of the 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing* [Online]. ICST, pp.409–418. Available from: http://eudl.eu/doi/10.4108/icst.collaboratecom.2013.254258.

Longley, P. a. and Adnan, M. 2016. Geo-temporal Twitter demographics. *International Journal of Geographical Information Science*. **30**(2), pp.369–389.

Longley, P. a., Adnan, M. and Lansley, G. 2015. The geotemporal demographics of twitter usage. *Environment and Planning A*. **47**(2), pp.465–484.

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., Malleson, N., Harland, K. and Birkin, M. 2014. Geotagged tweets to inform a spatial interaction model: a case study of museums.

arXiv preprint arXiv:1403.5118.

Luo, F., Cao, G., Mulligan, K. and Li, X. 2016. Explore spatiotemporal and demographic characteristics of human mobility via Twitter: A case study of Chicago. *Applied Geography*. **70**, pp.11–25.

Ma, D., Sandberg, M. and Jiang, B. 2017. A Socio-Geographic Perspective on Human Activities in Social Media. *Geographical Analysis*. **49**(3), pp.328–342.

Maeda, T.N., Yoshida, M., Toriumi, F. and Ohashi, H. 2018. Extraction of tourist destinations and comparative analysis of preferences between foreign tourists and domestic tourists on the basis of geotagged social media data. *ISPRS International Journal of Geo-Information*. **7**(3).

Maitland, R. 2013. Backstage Behaviour in the Global City: Tourists and the Search for the 'Real London'. *Procedia - Social and Behavioral Sciences*. **105**, pp.12–19.

Maitland, R. 2007. Cultural tourism and the development of new tourism areas in London *In*: G. Richards, ed. *Cultural tourism: global and local perspectives*. New York: Haworth Hospitality Press, pp.113–129.

Maitland, R. 2019. Extending the Frontiers of City Tourism: Suburbs and the Real London *In*: A. Smith and A. Graham, eds. *Destination London: The Expansion of the Visitor Economy* [Online]. University of Westminster Press, pp.15–35. Available from: https://www.jstor.org/stable/j.ctvhrd0t9.5.

Maitland, R. and Newman, P. 2009. London – Tourism Moving East? *In*: R. Maitland and P. Newman, eds. *World tourism cities: Developing tourism off the beaten track*. Abingdon: Routledge.

Majid, A., Chen, L., Mirza, H.T., Hussain, I. and Chen, G. 2015. A system for mining interesting tourist locations and travel sequences from public geo-tagged photos. *Data and Knowledge Engineering*. **95**, pp.66–86.

Malik, M.M., Lamba, H., Nakos, C. and Pfeffer, J. 2015. Population bias in geotagged tweets *In*: *AAAI Workshop - Technical Report* [Online]., pp.18–27. [Accessed 16 January 2022]. Available from: https://ojs.aaai.org/index.php/ICWSM/article/download/14688/14537/18206.

Malleson, N. and Birkin, M. 2014. New Insights into Individual Activity Spaces using Crowd-Sourced Big Data *In*: *2014 Academy of Science and Engineering BIGDATA/SOCIALCOM/CYBERSECURITY Conference*. Stanford University, pp.1–10.

Manikonda, L., Vamsikrishna, V. and Kambhampati, M.S. 2016. Tweeting the mind and Instagramming the heart: Exploring differentiated content sharing on social media *In*: *Proceedings of the 10th International Conference on Web and Social Media*. AAAI Press, pp.639–642.

Martí, P., García-Mayor, C. and Serrano-Estrada, L. 2019. Identifying opportunity places for urban regeneration through LBSNs. *Cities*. **90**, pp.191–206.

Martí, P., García-Mayor, C. and Serrano-Estrada, L. 2020. Taking the urban tourist activity pulse through digital footprints. *Current Issues in Tourism*., pp.1–20.

Martí, P., Serrano-Estrada, L. and Nolasco-Cirugeda, A. 2019. Social Media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*. **74**, pp.161–174.

Martin, D., Cockings, S. and Leung, S. 2010. *Progress report: 24-hour gridded population models Exploiting crowd-sourced annotation and semantic indexing to improve search results View project* [Online]. Available from: https://www.researchgate.net/publication/268428802.

Mayer-Schönberger, V. and Cukier, K. 2013. *Big Data: A Revolution that Will Transform how We Live, Work, and Think* [Online]. Houghton Mifflin Harcourt. Available from: https://books.google.co.uk/books?id=uy4lh-WEhhIC.

McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M. and Fisher, P. 2007. The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology*. **7**(1), p.30.

McKenzie, G., Janowicz, K., Gao, S. and Gong, L. 2015. How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems*. **54**(October), pp.336–346.

Md Khairi, N.D., Ismail, H.N. and Syed Jaafar, S.M.R. 2019. Tourist behaviour through consumption in Melaka World Heritage Site. *Current Issues in Tourism*. **22**(5), pp.582–600.

Mellon, J. and Prosser, C. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of british social media users. *Research and Politics*. **4**(3).

Memon, I., Chen, L., Majid, A., Lv, M., Hussain, I. and Chen, G. 2015. Travel Recommendation Using Geo-tagged Photos in Social Media for Tourist. *Wireless Personal Communications*. **80**(4), pp.1347–1362.

Miah, S.J., Vu, H.Q., Gammack, J. and McGrath, M. 2017. A Big Data Analytics Method for Tourist Behaviour Analysis. *Information and Management*. **54**(6), pp.771–785.

Miller, H.J. and Goodchild, M.F. 2015. Data-driven geography. *GeoJournal*. **80**(4), pp.449–461.

Morales, Jose, Flacke, J., Morales, Javier and Zevenbergen, J. 2019. Mapping Urban Accessibility in Data Scarce Contexts Using Space Syntax and Location-Based Methods. *Applied Spatial Analysis and Policy*. **12**(2), pp.205–228.

Munar, A.M. and Jacobsen, J.K.S. 2014. Motivations for sharing tourism experiences through social media. *Tourism Management*. **43**, pp.46–54.

National Travel Survey 2017. Average trip length by main mode, for eat/drink purposes only: England, 2002 to 2017. *Department for Transport statistics*. [Online]. [Accessed 16 July 2020]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/858316/ntsq03006.ods.

Newing, A. 2013. *Incorporating seasonal visitor demand in retail location modelling*. University of Leeds.

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

Newing, A., Clarke, G.P. 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.P. and Clarke, M. 2013a. 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. 2013b. Visitor expenditure estimation for grocery store location planning: a case study of Cornwall. *The International Review of Retail, Distribution and Consumer Research*. **23**(3), pp.221–244.

Nielsen 2017. *Outbound Chinese Tourism and Consumption Trends* [Online]. Available from: http://www.nielsen.com/content/dam/nielsenglobal/cn/docs/Outbound Chinese Tourism and Consumption Trends.pdf.

Noulas, A. 2013. *Human urban mobility in location-based social networks: analysis, models and applications*. University of Cambridge.

Noulas, A., Mascolo, C. and Frias-Martinez, E. 2013. Exploiting foursquare and cellular data to infer user activity in urban environments *In*: *Proceedings - IEEE International Conference on Mobile Data Management*., pp.167–176.

Noulas, A., Scellato, S., Lathia, N. and Mascolo, C. 2012. A random walk around the city: New venue recommendation in location-based social networks *In*: *Proceedings 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust*. IEEE, pp.144–153.

Noulas, A., Scellato, S., Mascolo, C. and Pontil, M. 2010. An Empirical Study of Geographic User Activity Patterns in Foursquare *In*: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*., pp.570–573.

Noulas, A., Scellato, S., Mascolo, C. and Pontil, M. 2011. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks *In*: *Fifth International AAAI Conference on Weblogs and Social Media*. AAAI Press, pp.32–35.

Novy, J. 2018. ‘Destination’ Berlin revisited. From (new) tourism towards a pentagon of mobility and place consumption. *Tourism Geographies*. **20**(3), pp.418–442.

Novy, J. and Colomb, C. 2019. Urban Tourism as a Source of Contention and Social Mobilisations: A Critical Review. *Tourism Planning and Development*.

Novy, J. and Huning, S. 2009. New tourism areas in the new Berlin *In*: R. Maitland and P. Newman, eds. *World tourism cities: Developing tourism offthe beaten track*. London: Routledge, pp.87–108.

Office for National Statistics (ONS) 2017. Living Costs and Food Survey (LCF). [Accessed 7 February 2020]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/expenditure/datasets/familyspendingworkbook3expenditurebyregion.

Office for National Statistics (ONS) 2014. *Workplace Zones : A new geography for workplace statistics* [Online]. Available from: http://www.demographicsusergroup.co.uk/objectives.html.

Oh, S. and Syn, S.Y. 2015. Motivations for sharing information and social support in social media: A comparative analysis of Facebook, Twitter, Delicious, YouTube, and Flickr. *Journal of the Association for Information Science and Technology*. **66**(10), pp.2045–2060.

Olteanu, A., Castillo, C., Diaz, F. and Kıcıman, E. 2019. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*. **2**(13).

ONS 2011. 2011 residential-based area classifications. [Accessed 28 September 2020]. Available from: https://www.ons.gov.uk/methodology/geography/geographicalproducts/areaclassifications/2011areaclassifications.

ONS 2016. International Passenger Survey (IPS) - Office for National Statistics. [Accessed 15 January 2022]. Available from: https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/internationalpassengersurvey.

ONS 2020. *International Passenger Survey (IPS) Methodology* [Online]. [Accessed 15 January 2022]. Available from: http://www.ons.gov.uk/ons/guide-method/method-quality/specific/population-and-migration/international-migration-methodology/international-passenger-survey-quality-information-in-relation-to-migration-flows.pdf.

Oshan, T.M. 2021. The spatial structure debate in spatial interaction modeling: 50 years on. *Progress in Human Geography*. **45**(5), pp.925–950.

Panczak, R., Charles-Edwards, E. and Corcoran, J. 2020. Estimating temporary populations: a systematic review of the empirical literature. *Palgrave Communications*. **6**(1).

Pappalepore, I., Maitland, R. and Smith, A. 2014. Prosuming creative urban areas. Evidence from East London. *Annals of Tourism Research*. **44**(1), pp.227–240.

Patel, N.N., Stevens, F.R., Huang, Z., Gaughan, A.E., Elyazar, I. and Tatem, A.J. 2017. Improving Large Area Population Mapping Using Geotweet Densities. *Transactions in GIS*. **21**(2), pp.317–331.

Payntar, N.D., Hsiao, W.L., Covey, R.A. and Grauman, K. 2021. Learning patterns of tourist movement and photography from geotagged photos at archaeological heritage sites in Cuzco, Peru. *Tourism Management*. **82**, p.104165.

Piovani, D., Zachariadis, V. and Batty, M. 2016. Quantifying Retail Agglomeration using Diverse Spatial Data. *Scientific Reports*. **7**(1), pp.1–8.

Qu, Y. and Zhang, J. 2013. Trade area analysis using User Generated Mobile Location Data *In*: *Proceedings of the 22nd International Conference on World Wide Web*., pp.1053–1063.

Rashidi, T.H., Abbasi, A., Maghrebi, M., Hasan, S. and Waller, T.S. 2017. Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*. **75**, pp.197–211.

Rizwan, M., Wan, W., Cervantes, O. and Gwiazdzinski, L. 2018. Using location-based social media data to observe check-in behavior and gender difference: Bringing weibo data into play. *ISPRS International Journal of Geo-Information*. **7**(5), p.196.

Roick, O. and Heuser, S. 2013. Location based social networks - definition, current state of the art and research agenda. *Transactions in GIS*. **17**(5), pp.763–784.

Sainaghi, R. 2012. Tourist expenditures: The state of the art. *Anatolia*. **23**(2), pp.217–233.

Salas-Olmedo, M.H., Moya-Gómez, B., García-Palomares, J.C. and Gutiérrez, J. 2018. Tourists’ digital footprint in cities: Comparing Big Data sources. *Tourism Management*. **66**, pp.13–25.

Sánchez-Galiano, J.C., Martí-Ciriquián, P. and Fernández-Aracil, P. 2017. Temporary population estimates of mass tourism destinations: The case of Benidorm. *Tourism Management*. **62**, pp.234–240.

Scellato, S., Noulas, A., Lambiotte, R. and Mascolo, C. 2011. Socio-Spatial Properties of Online Location-Based Social Networks *In*: *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* [Online]., pp.329–336. Available from: www.aaai.org.

Schuckert, M., Liu, X. and Law, R. 2015. Hospitality and Tourism Online Reviews: Recent Trends and Future Directions. *Journal of Travel and Tourism Marketing*. **32**(5), pp.608–621.

Seresinhe, C.I., Moat, H.S. and Preis, T. 2018. Quantifying scenic areas using crowdsourced data. *Environment and Planning B: Urban Analytics and City Science*. **45**(3), pp.567–582.

Sessions, C., Wood, S.A., Rabotyagov, S. and Fisher, D.M. 2016. Measuring recreational visitation at U.S. National Parks with crowd-sourced photographs. *Journal of Environmental Management*. **183**, pp.703–711.

Sevtsuk, A. and Kalvo, R. 2018. Patronage of urban commercial clusters: A network-based extension of the Huff model for balancing location and size. *Environment and Planning B: Urban Analytics and City Science*. **45**(3), pp.508–528.

Shabrina, Z. 2020. *The impact of the platform economy in cities: the case of Airbnb*.

Shabrina, Z., Buyuklieva, B. and Ming, M.N.K. 2019. Airbnb, hotels, and saturation of the food industry: A multi-scale GWR approach Available from: http://arxiv.org/abs/1905.12543.

Shao, H., Zhang, Y. and Li, W. 2017. Extraction and analysis of city's tourism districts based on social media data. *Computers, Environment and Urban Systems*. **65**, pp.66–78.

Shen, Y. and Karimi, K. 2016. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities*. **55**, pp.9–21.

Shen, Y., Karimi, K., Law, S. and Zhong, C. 2019. Physical co-presence intensity: Measuring dynamic face-to-face interaction potential in public space using social media check-in records. *PLoS ONE*. **14**(2).

Shi, B., Zhao, J. and Chen, P.J. 2017. Exploring urban tourism crowding in Shanghai via crowdsourcing geospatial data. *Current Issues in Tourism*. **20**(11), pp.1186–1209.

Shoval, N. and Ahas, R. 2016. The Use of Tracking Technologies in Tourism Research : A Review of the First Decade. *Tourism Geographies*. **6688**(September), pp.1–20.

Shoval, N., McKercher, B., Ng, E. and Birenboim, A. 2011. Hotel location and tourist activity in cities. *Annals of Tourism Research*. **38**(4), pp.1594–1612.

Siła-Nowicka, K. and Fotheringham, A.S. 2019. Calibrating spatial interaction models from GPS tracking data: An example of retail behaviour. *Computers, Environment and Urban Systems*. **74**, pp.136–150.

Sklar, M., Mascolo, C., Noulas, A., D'Silva, K. and Musolesi, M. 2017. If I build it, will they come? Predicting new venue visitation patterns through mobility data *In*: *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. Los Angeles Area, CA, USA, pp.1–4.

Sloan, L., Morgan, J., Burnap, P. and Williams, M. 2015. Who tweets? deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *PLoS ONE*. **10**(3), p.e0115545.

Smith, A. 2019a. Destination London: An Expanding Visitor Economy *In*: A. Smith and A. Graham, eds. *Destination London: The Expansion of the Visitor Economy* [Online]. University of Westminster Press Chapter, pp.1–13. Available from: https://www.jstor.org/stable/j.ctvhrd0t9.4.

Smith, A. 2019b. Event Takeover? The Commercialisation of London's Parks *In*: *Destination London: The Expansion of the Visitor Economy*., pp.205–223.

Smith, A. and Graham, A. 2019. *Destination London: The Expansion of the Visitor Economy*. University of Westminster Press.

Smith, G. and Fairburn, J. 2008. *Updating and improving the National Population Database to National Population Database 2*.

Al Sonosy, O.A., Rady, S., Badr, N.L. and Hashem, M. 2018. Toward efficient business behavior prediction using location-based social networks. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. **8**(4).

Spalding, M., Burke, L., Wood, S.A., Ashpole, J., Hutchison, J. and Zu Ermgassen, P. 2017. Mapping the global value and distribution of coral reef tourism. *Marine Policy*. **82**, pp.104–113.

Spyrou, E. and Mylonas, P. 2016a. A survey on Flickr multimedia research challenges. *Engineering Applications of Artificial Intelligence*. **51**, pp.71–91.

Spyrou, E. and Mylonas, P. 2016b. Analyzing Flickr metadata to extract location-based information and semantically organize its photo content. *Neurocomputing*. **172**, pp.114–133.

Statista 2018. Mobile social media - Statistics & Facts. [Accessed 16 March 2018]. Available from: https://www.statista.com/topics/1164/social-networks/.

Steiger, E., Albuquerque, J.P. and Zipf, A. 2015. An Advanced Systematic Literature Review on Spatiotemporal Analyses of Twitter Data. *Transactions in GIS*. **19**(6), pp.809–834.

Steiger, E., Resch, B. and Zipf, A. 2016. Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. *International Journal of Geographical Information Science*. **30**(9), pp.1694–1716.

Steiger, E., Westerholt, R., Resch, B. and Zipf, A. 2015. Twitter as an indicator for whereabouts of people? Correlating Twitter with UK census data. *Computers, Environment and Urban Systems*. **54**, pp.255–265.

Stock, K. 2018. Mining location from social media: A systematic review. *Computers, Environment and Urban Systems*. **71**, pp.209–240.

Stors, N. 2020. Constructing new urban tourism space through Airbnb. *Tourism Geographies*., pp.1–24.

Su, S., Wan, C., Hu, Y. and Cai, Z. 2016. Characterizing geographical preferences of international tourists and the local influential factors in China using geo-tagged photos on social media. *Applied Geography*. **73**, pp.26–37.

Sui, D. and Goodchild, M.F. 2011. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*. **25**(11), pp.1737–1748.

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L. and Toivonen, T. 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. *Scientific Reports*. **7**(1), pp.1–11.

Thompson, C., Clarke, G.P., Clarke, M. and Stillwell, J. 2012. Modelling the future opportunities for deep discount food retailing in the UK. *International Review of Retail, Distribution and Consumer Research*. **22**(2), pp.143–170.

Thompson, C.P. 2013. *Retail Spending and Store Location during a Recession: An Analysis of Changing Consumer Behaviour and Interaction Patterns*.

Timothy, D.J. 2005. *Shopping Tourism, Retailing and Leisure*. Channel View Publications.

Timothy, D.J. 2014. Trends in Tourism, Shopping, and Retailing *In*: A. A. Lew, C. M. Hall and A. M. Williams, eds. *The Wiley Blackwell Companion to Tourism*. Wiley-Blackwell, pp.378–388.

UNWTO 2014. Global Report on Shopping Tourism. . **8**, pp.1–66.

UNWTO 2008. *International Recommendations for Tourism Statistics*. New York / Madrid.

Vaca, C., Quercia, D., Bonchi, F. and Fraternali, P. 2015. Taxonomy-based discovery and annotation of functional areas in the city *In*: *Ninth international AAAI conference on web and social media.*, pp.445–453.

Venables, W.N. and Ripley, B.D. 2002. *Modern Applied Statistics with S* [Online]. Springer. [Accessed 17 January 2022]. Available from: http://barbra-coco.dyndns.org/yuri/R/Venables, WN and Ripley, BD(2002) Modern Applied Statistics with S, 4th Ed.pdf.

VisitBritain 2021a. 2021 tourism forecast. [Accessed 1 August 2021]. Available from: https://www.visitbritain.org/2021-tourism-forecast.

VisitBritain 2020. *China Inbound tourism overview* [Online]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/markets/china_fact_sheet_double_sided_sept-2020_with_alt_05112020.pdf.

VisitBritain 2021b. China Market and Trade Profile. , p.59.

VisitBritain 2019a. *China Outbound travel to the UK* [Online]. [Accessed 22 December 2021]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/markets/aviation_market_china_23062020.pdf.

VisitBritain 2017. International Market Advice: China. Available from: https://www.visitbritain.org/markets/china.

VisitBritain 2019b. Quarterly Inbound Update Full Year 2018. [Accessed 20 January 2020]. Available from: https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/2018_uk_and_regional_ips_summary.pdf.

VisitBritian 2020. 2019 snapshot. *VisitBritian*. [Online]. [Accessed 19 March 2021]. Available from: https://www.visitbritain.org/2019-snapshot.

VisitEngland 2016. Accommodation Stock Audit. [Accessed 1 February 2020]. Available from: https://www.visitbritain.org/accommodation-stock.

VisitEngland 2019. England Occupancy Survey. . (Nov). [Accessed 1 March 2020]. Available from: https://www.visitbritain.org/accommodation-occupancy-archive.

Vu, H.Q., Law, R. and Li, G. 2019. Breach of traveller privacy in location-based social media. *CURRENT ISSUES IN TOURISM*. **22**(15), pp.1825–1840.

Vu, H.Q., Li, G. and Law, R. 2019. Discovering implicit activity preferences in travel itineraries by topic modeling. *Tourism Management*. **75**, pp.435–446.

Vu, H.Q., Li, G., Law, R. and Ye, B.H. 2015. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*. **46**, pp.222–232.

Vu, H.Q., Li, G., Law, R. and Zhang, Y. 2018. Tourist Activity Analysis by Leveraging Mobile Social Media Data. *Journal of Travel Research*. **57**(7), pp.883–898.

Wachsmuth, D. and Weisler, A. 2018. Airbnb and the rent gap: Gentrification through the sharing economy. *Environment and Planning A: Economy and Space*. **50**(6), pp.1147–1170.

Waddington, T. 2017. *Modelling Spatiotemporal Fluctuationsof Consumer Demand in the UK Grocery Sector and their Impact on Retailers Store Sales*.

Waddington, T., Clarke, G.P., 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. 2017. Open all hours: spatiotemporal fluctuations in U.K. grocery store sales and catchment area demand. *The International Review of Retail, Distribution and Consumer Research*. **3969**(November), pp.1–26.

Wang, Y. and Davidson, M.C.G. 2010. A review of micro-analyses of tourist expenditure. *Current Issues in Tourism*. **13**(6), pp.507–524.

Wang, Y., Jiang, W., Liu, S., Ye, X. and Wang, T. 2016. Evaluating Trade Areas Using Social Media Data with a Calibrated Huff Model. *ISPRS International Journal of Geo-Information*. **5**(7), p.112.

Wang, Y., Wang, T., Tsou, M.H., Li, H., Jiang, W. and Guo, F. 2016. Mapping dynamic urban land use patterns with crowdsourced geo-tagged social media (Sina-Weibo) and commercial points of interest collections in Beijing, China. *Sustainability (Switzerland)*. **8**(11).

Ward, J.H. 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*. **58**(301), pp.236–244.

Weibo Corp. 2021. Weibo Reports First Quarter 2021 Unaudited Financial Results. [Accessed 26 May 2021]. Available from: http://ir.weibo.com/news-releases/news-release-details/weibo-reports-first-quarter-2021-unaudited-financial-results.

Weibo Data Centre 2021. *Annual Report on Weibo User Development* [Online]. [Accessed 22 December 2021]. Available from: https://data.weibo.com/report/reportDetail?id=456.

Weibo Data Centre 2016. *Weibo Travel Data Report*.

Wilson, A.G. 1971. A family of spatial interaction models, and associated developments. *Environment and Planning*. **3**, pp.1–32.

Wilson, A.G. 1974. *Urban and regional models in geography and planning*. London: Wiley.

Wojcik, S. and Hughes, A. 2019. How Twitter Users Compare to the General Public | Pew Research Center. *Pew Research Center*. [Online]. [Accessed 17 January 2022]. Available from: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/.

Wood, S. and Reynolds, J. 2012. Leveraging locational insights within retail store development? Assessing the use of location planners' knowledge in retail marketing. *Geoforum*. **43**(6), pp.1076–1087.

Wood, S. and Reynolds, J. 2011. The intrafirm context of retail expansion planning. *Environment and Planning A*. **43**(10), pp.2468–2491.

Wood, S.A., Guerry, A.D., Silver, J.M. and Lacayo, M. 2013. Using social media to quantify nature-based tourism and recreation. *Scientific Reports*. **3**.

Wu, W., Wang, J. and Dai, T. 2016. The Geography of Cultural Ties and Human Mobility: Big Data in Urban Contexts. *Annals of the American Association of Geographers*. **0**(0), pp.1–19.

Xiang, Z., Du, Q., Ma, Y. and Fan, W. 2017. A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*. **58**, pp.51–65.

Xie, R., Chen, Y., Xie, Q., Xiao, Y. and Wang, X. 2018. We Know Your Preferences in New Cities : Mining and Modeling the Behavior of Travelers. *IEEE Communications Magazine*. **56**(11), pp.1–8.

Xue, L. and Zhang, Y. 2020. The effect of distance on tourist behavior: A study based on social media data. *Annals of Tourism Research*. **82**, p.102916.

Yao, X.A., Huang, H., Jiang, B. and Krisp, J.M. 2019. Representation and analytical models for location-based big data. *International Journal of Geographical Information Science*. **33**(4), pp.707–713.

Yu, Z., Tian, M., Wang, Z., Guo, B. and Mei, T. 2016. Shop-Type Recommendation Leveraging the Data from Social Media and Location-Based Services. *ACM Transactions on Knowledge Discovery from Data*. **11**(1), pp.1–21.

Yu, Z., Zhang, D. and Yang, D. 2013. Where is the largest market: Ranking areas by popularity from location based social networks. *2013 IEEE 10th International Conference on Ubiquitous Intelligence and Computing and 2013 IEEE 10th International Conference on Autonomic and Trusted Computing*., pp.157–162.

Yuan, Y. and Wang, X. 2018. Exploring the effectiveness of location-based social media in modeling user activity space: A case study of Weibo. *Transactions in GIS*.

van der Zee, E., Bertocchi, D. and Vanneste, D. 2020. Distribution of tourists within urban heritage destinations: a hot spot/cold spot analysis of TripAdvisor data as support for destination management. *Current Issues in Tourism*. **23**(2), pp.175–196.

Zervas, G., Proserpio, D. and Byers, J.W. 2017. The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*. **54**(5), pp.687–705.

Zhang, F., Zu, J., Hu, M., Zhu, D., Kang, Y., Gao, S., Zhang, Y. and Huang, Z. 2020. Uncovering inconspicuous places using social media check-ins and street view images. *Computers, Environment and Urban Systems*. **81**, p.101478.

Zhao, P., Liu, X., Shi, W., Jia, T., Li, W. and Chen, M. 2020. An empirical study on the intra-urban goods movement patterns using logistics big data. *International Journal of Geographical Information Science*. **34**(6), pp.1089–1116.

Zhen, F., Cao, Y., Qin, X. and Wang, B. 2017. Delineation of an urban agglomeration boundary based on Sina Weibo microblog ‘check-in’ data: A case study of the Yangtze River Delta. *Cities*. **60**, pp.180–191.

Zhong, Y., Yuan, N.J., Zhong, W., Zhang, F. and Xie, X. 2015. You Are Where You Go: Inferring Demographic Attributes from Location Check-ins. *Wsdm*., pp.295–304.

Zhou, X. and Zhang, L. 2016. Crowdsourcing functions of the living city from Twitter and Foursquare data. *Cartography and Geographic Information Science*. **43**(5), pp.393–404.

Zickuhr, K. 2013. Location-based services. *Pew Internet and American Life Project*. [Online]. [Accessed 23 February 2021]. Available from: http://www.pewinternet.org/2013/09/12/ location-based-services/.

Zook, M., Kraak, M.J. and Ahas, R. 2015. Geographies of mobility: applications of location-based data. *International Journal of Geographical Information Science*. **29**(11), pp.1935–1940.