

AN APPLICATION OF K-MEANS CLUSTERING ON TOURIST ACTIVITY IN MAUI USING MOBILE LOCATION DATA

by

Anna Meg Shields Brooker,
HBA in Environment and Urban Sustainability,
Ryerson University, 2018

A Major Research Paper presented to Ryerson University

in partial fulfillment of the
requirements for the degree of
Master of Spatial Analysis (MSA)
in the program of Spatial Analysis

Toronto, Ontario, Canada, 2019

© Anna Meg Shields Brooker, 2019

Author's declaration for electronic submission of an MRP

I hereby declare that I am the sole author of this MRP. This is a true copy of the MRP, including any required final revisions. I authorize Ryerson University to lend this MRP to other institutions or individuals for the purpose of scholarly research. I further authorize Ryerson University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research. I understand that my MRP may be made electronically available to the public.

Anna Meg Shields Brooker

An Application of K-Means Clustering On Tourist Activity in Maui Using Mobile Location Data
Master of Spatial Analysis (MSA), 2019
Anna Meg Shields Brooker
Spatial Analysis
Ryerson University

Abstract

Mobile location data are a major form of Big Data that hold many possibilities for study and insight into human behaviour. This research used mobile location data to investigate the differences in the activity patterns of tourists in Maui, Hawai'i. Mobile data used in this study were app-based location data collected as a stream of mobile phone locations with a timestamp. Tourists were clustered using K-Means based on time spent at attraction types. Different travel experiences were analyzed based on traveler's accommodation choices, the average distance travelled from accommodation to attraction, and vacation length, which all varied significantly between the tourist clusters. This work provided a new use for K-means clustering with mobile location data to provide insightful information to marketing professionals and tourism management bodies.

Acknowledgements

This research would not have been possible without the help of Peter Miron, Layne Newhouse, and Dr. Siyue Tian at Environics Analytics. Thank you for the extensive data cleaning, database management, organization, and magic, as well as the endless well of support, suggestions, and inspiration. Many thanks to Dr. Joseph Aversa for his excitement and guidance throughout this work, and to Dr. Richard Shaker for his recommendations. Finally, thank you to my parents for supporting me and to Sally and Jasper for carrying me through the year.

Table of Contents

Author's declaration for electronic submission of an MRP	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	vii
List of Figures	viii
CHAPTER 1: Introduction	1
CHAPTER 2: Theoretical Framework	3
2.1 Mobile Big Data	3
2.1.1 Applications of mobile Big Data	4
2.2 Travel and Tourism	4
2.2.1 Modelling tourist experiences	5
2.2.2 Points of interest (POIs) & regions of interest (ROIs)	6
2.2.3 Applications of mobile data in tourism research	7
2.3 Methodology used in literature	7
2.3.1 Data preprocessing and filtering	8
2.3.2 Identify tourist devices	8
2.3.3 Identify tourist attractions/points of interest	9
2.3.4 Clustering & results analysis	10
CHAPTER 3: Data & Methodology	11
3.1 Study data	11
3.2 Data preprocessing and filtering	12
3.3 Identifying tourist devices	13
3.4 Identify tourist attractions/points of interest & accommodations	13
3.5 Trip identification	14
3.6 K-Means clustering & cluster summaries	17
CHAPTER 4: Results	19
4.1 Identify tourists	19
4.2 Identify points of interest and accommodations	19
4.3 K-Means clustering & cluster summaries	22

4.3.1 Time spent at attraction types	26
4.3.2 Distance travelled accommodation to attraction	26
4.3.3 Cluster descriptions	28
4.3.4 Spatial distribution of cluster accommodation locations.....	29
CHAPTER 5: Discussion & Conclusion	33
5.1 Segmentation of tourists using time spent at point of interest types & differences between segments	33
5.2 Relationship between point of interest choices and summary variables	34
5.3 Future research opportunities	35
5.4 Limitations	36
5.5 Conclusion.....	38
References	39

List of Tables

Table 1 – Classification parameters of point of interest types per class, for geofence annotation	14
Table 2 – Example of trip identification, where suitable trips are highlighted with red boxes	17
Table 3 – Data filtering results.....	19
Table 4 – Geofence summary of point of interest types	20
Table 5– Number of cases in each cluster	22
Table 6 – ANOVA test results	22
Table 7 – Final cluster centres: average proportion of time spent at each activity type, per cluster	23
Table 8– Based on the percentage of devices staying at each accommodation type	25
Table 9 – Summary of cluster vacation lengths and travelling range	25
Table 10 – Average distance between attraction and accommodation, by attraction type	27

List of Figures

Figure 1 – Flowchart of study methodology	11
Figure 2 – Geofences of attractions and accommodations in Maui, Hawai'i. Based on imagery and business information from Google Maps, 2019	21
Figure 3 – Homogeneity of cases within each cluster, based on their distance from the cluster centre	24
Figure 4 – Density distribution of accommodation locations by cluster	31
Figure 5 – Density distribution of accommodation locations by cluster	32

CHAPTER 1: Introduction

Information on lodging types and tourist attractions was previously available mainly using surveys (Gursoy, Jurowski, and Uysal 2002; Collins, Rose, and Hess 2012; Phithakkitnukoon et al. 2015; Chen et al. 2018; Hasnat and Hasan 2018; Kirilenko, Stepchenkova, and Hernandez 2019), however this form of data collection did not have the capability of capturing the nuances in travel experiences. Using mobile data in a tourism experience application may allow marketing professionals to pinpoint consumer behaviour and establish determining factors at a level of granularity that has been previously inaccessible. The use of mobile data in this study provided greater insight into tourist behaviour and motivation based on quantitative techniques rather than relying on qualitative surveys used to determine what might impact activity motivation. This work demonstrated the ability of mobile data to better inform tourism management bodies to better strategize their advertising and marketing based on the results – with the idea that this model could be replicated in other tourism regions.

Objectives:

1. Use mobile data to investigate tourism experiences and the potential determining factors for how tourists choose to spend their time. A desired outcome is also a segmentation of tourists based on their activity choices.
2. Provide insight on the effectiveness of the methodology and of mobile data for this purpose.

Questions:

1. Can tourists be segmented based solely on the frequency of their visits to different points of interest, and are these segments significantly different from one another?
2. Is there a relationship between points of interest choices and distance from accommodation to attraction, accommodation type, and/or tourists' vacation lengths? Is this consistent across the segments or does the relationship vary?

To answer these questions, tourist accommodations were identified using mobile data to determine the most common nighttime location and that location's corresponding accommodation-type geofence. The proportion of visit minutes to each POI type were calculated for each visitor and tourists were segmented using K-means clustering based on these POI visit

statistics. Distance from accommodation to attraction and vacation lengths were summarized along with accommodation types for each device to allow for an investigation of the impacts of these variables on tourist activity choices. The clusters were also described and compared to one another to illuminate what information can be gathered for marketing purposes from this type of mobile location data.

This paper first summarizes existing research in the field with the purpose of contextualizing this study and chosen methodologies. The methodology used in this study is then described and justified. The results are presented, and their significance discussed. The paper is ended with a summary of the study limitations and the future research opportunities, followed by a conclusion.

CHAPTER 2: Theoretical Framework

2.1 Mobile Big Data

In the last two decades, the influx of data being collected from a variety of devices such as smart devices, streaming, and app use, has exploded. The size of Big Data may vary depending on the means by which they are collected, but the consistent theme is that the size of this data type is too large to be able to interpret or decipher the way one might a traditional database (Cuzzocrea, 2016; Manyika et al., 2011). Big Data might refer to a database holding several terabytes of information, or may hold several petabytes (Manyika et al. 2011). Mobile location data are a major form of Big Data that hold many possibilities for study and insight into human behaviour (Gudmundsson, Laube, and Wolle, 2008). Most mobile location data have three components: latitude and longitude coordinates; a timestamp from when the location was visited by the user; and an identifier unique to the device (Manyika et al. 2011; Qiao et al. 2018). Mobile location data are typically difficult to manage because of their large size and less than ideal organization, in which a database contains many rows of instances and repeating unique device identifiers. For this reason, a large body of research is beginning to form around the effective management and analysis of Big Data, as well as on their collection and storage (Zhao, Ma, and Yang 2011; Xu et al. 2017; Qiao et al. 2018).

Location-based data have been used for analysis by scraping data from geotagged online photo posts and Tweets (Kawakubo and Yanai 2009; Li et al. 2013), using machine-learning to create geotags for photos of distinct landmarks (Chen and Grauman 2011), geotagged online reviews (Kirilenko, Stepchenkova, and Hernandez 2019), or users declaring their origin and travel destination tied to a social media posting (Hasnat and Hasan, 2018). This type of geolocation data lacks the level of detail now available through mobile location data, which have a fine scale of temporal and spatial geography. Mobile location data present an opportunity for fine grained analysis of movement and travel trajectories (Zhao, Ma, and Yang 2011; Phithakkitnukoon et al. 2015; Chen et al. 2018; Hasnat and Hasan 2018, 2018; Li et al. 2019), as well as information on habitual movement and space use at a variety of temporal scales. (Chen and Grauman 2011; Phithakkitnukoon et al. 2015; Guo, Zhang, and Zhang 2016; Shen and Cheng 2016; Xu et al. 2017; Kirilenko, Stepchenkova, and Hernandez 2019)

2.1.1 Applications of mobile Big Data

Mobile data have been used to analyse population movement after a major event such as an earthquake in order to plan resource allocation in catastrophe response (Bengtsson et al. 2011; Hu et al. 2011; Urata, Sasaki, and Iryo 2018). The frequent time stamps and small level of geography available through mobile data have allowed for researchers to track the number of patients visiting a hospital emergency room to best recommend a nearby hospital to ambulances based on crisis severity (Chao et al. 2018). Major uses of these data in academic literature focus on travel trajectories, regions of interest (ROIs), and places of interest (POIs). Research around ROIs/POIs use different validation techniques of known ROIs using mobile data (Fei Hu, Qingquan Sun, and Qi Hao 2010; Guo, Zhang, and Zhang 2016; Xu et al. 2017, 2017; Luo et al. 2018). Travel trajectories have been studied for both local residents (YuFang Dan and Zhongshi He 2010; Xu et al. 2017; Urata, Sasaki, and Iryo 2018) and tourists (Guo, Zhang, and Zhang 2016; Chen et al. 2018; Zhao et al. 2018) to derive information such as place of residence and place of work (Phithakkitnukoon et al. 2015; Guo, Zhang, and Zhang 2016), temporal population density (YuFang Dan and Zhongshi He 2010; Urata, Sasaki, and Iryo 2018), and movement habits (Phithakkitnukoon et al. 2015; Qiao et al. 2018; Urata, Sasaki, and Iryo 2018) for purposes such as city planning (YuFang Dan and Zhongshi He 2010), transportation allocation planning (Qiao et al. 2018; Urata, Sasaki, and Iryo 2018; Zhao et al. 2018), tourism analysis (Phithakkitnukoon et al. 2015; Zhao et al. 2018), and more (Gudmundsson, Laube, and Wolle 2006; YuFang Dan and Zhongshi He 2010; Phithakkitnukoon et al. 2015; Guo, Zhang, and Zhang 2016; Chen et al. 2018; Cheng et al. 2018; Qiao et al. 2018; Zhao et al. 2018).

2.2 Travel and Tourism

There is a huge body of research surrounding tourism, which can be an important source of commerce and an important part of the human experience in many cultures. People from many different fields study tourism and travel experiences from a variety of angles and with unique research goals. Travel experiences, tourist activities and related business opportunities, as well as the impacts of local tourism and sustainable tourism have been studied extensively. The economic impacts of tourism have been thoroughly investigated for possible business potential and marketing insights (Zhou et al. 1997; Latzko 2005; Rinne and Saastamoinen 2005; Balmford et al.

2009), and for use in customer segmentation models (Arrowsmith and Ntuwah 2002; Asan and Emeksiz 2016). Asan and Emeksiz (2016) found that the investigation of vacation activities equated to an investigation of vacation behaviour, which was a meaningful factor in market segmentation. The environmental and social impact of tourism has been researched through the lens of sustainability (Gursoy, Jurowski, and Uysal 2002; Balmford et al. 2009; Saito 2013). Travel experiences have also been modelled through the study of locations visited or used by tourists.

2.2.1 Modelling tourist experiences

Several studies focus on identifying travel POIs (Kawakubo and Yanai 2009; Chen and Grauman 2011; Li et al. 2013; Phithakkitnukoon et al. 2015; Chen et al. 2018; Hasnat and Hasan 2018), ROIs (Guo, Zhang, and Zhang 2016; Xu et al. 2017; Chen et al. 2018; Hasnat and Hasan 2018; Luo et al. 2018; Kirilenko, Stepchenkova, and Hernandez 2019) and travel patterns between POIs (Phithakkitnukoon et al. 2015; Xu et al. 2017; Chen et al. 2018; Zhao et al. 2018; Kirilenko, Stepchenkova, and Hernandez 2019) with the use of mobile location data. POIs and ROIs were used to categorize types of attractions and make inferences about tourist experiences or tourist profiling based on the POI/ROI visitation activity findings (Phithakkitnukoon et al. 2015; Luo et al. 2018). Kirilenko, Stepchenkova, and Hernandez (2019) discussed different tourist types identified throughout literature with particular regard to tourist motivations as a means of explaining travel choices. Accommodation type and availability may also be an important determinant of tourist behaviour (Keogh 1984; Chen et al. 2018).

Other studies focus on one type of POI and the motivations for visiting as well as the impact of the visit experience upon the tourist. Quan and Wang (2004) focused on the concepts of “peak” experiences vs. extensions of the daily routine as a means of quantifying the motivations for engaging in different food-related travel activities. Accommodation choices, (Keogh 1984), gastronomical attractions (Quan and Wang 2004; Kivela and Crofts 2009; Mkono, Markwell, and Wilson 2013), and nature-related attractions or activities (Rinne and Saastamoinen 2005; Balmford et al. 2009; Asan and Emeksiz 2016; Urata, Sasaki, and Iryo 2018) are all POIs that have been studied to investigate tourist motivations.

2.2.2 Points of interest (POIs) & regions of interest (ROIs)

Points of interest (POIs) and regions of interest (ROIs) have been used extensively in tourism research. Sometimes in the context of identifying tourist attractions, and other times using known tourist attractions to aid in the analysis of other patterns. Hasnat and Hasan (2018) investigated tourist POIs (referred to as “tourist spots”) as a way of determining travel trajectories for the purposes of informing transportation planning. K-Means, Mean Shift, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) methods were used to identify POIs (Hasnat and Hasan 2018). Kirilenko, Stepchenkova, and Hernandez (2019) identified POIs (“tourist attractions”) as any attraction that received a high percentage of the online reviews, and then used Getis-Ord G_i^* hot spot analysis to identify the key areas where these attractions were located, identifying ROIs. They found the following types of attractions to be significantly popular (referred to as “communities of attractions-interests”): theme parks, museums and historical places, and nature-based attractions. Chen et al. (2018) found that non-nature related POIs belonged to one of the following categories: lodging, dining, shopping, recreation (such as a cinema or museum), and transfer (such as an airport or cruise ship port). Quan and Wang (2004) highlight the essential nature of eating and food experiences to the tourist experience. Extensions of the daily routine, such as eating familiar foods, have been found to make up most of a tourist experience and are essential in the overall travel experience (Quan and Wang 2004; Mkono, Markwell, and Wilson 2013). Gastronomy is an essential part of cultural tourism, and many tourists may plan an entire trip around food experiences (Quan and Wang 2004; Kivela and Crofts 2009; Mkono, Markwell, and Wilson 2013). Restaurant experiences as well as visits to grocery stores then, may be reasonably considered essential POIs when modelling a tourist experience. Outdoor recreation has also been identified as an important tourist activity (Rinne and Saastamoinen 2005; Asan and Emeksiz 2016; Shiaty, Taalab, and Osama 2016). Hiking in particular is a common form of outdoor recreation amongst tourists, motivated by a desire for “learning and sociality, independence and personal development, and relaxation” (Asan and Emeksiz 2016, pp 8). Nature walks, cycling trips, mountain climbing, nature photography, fishing, water sports (like diving), spelunking, and air sports (like bungee jumping), were all identified as being significantly frequented outdoor activities that tourists tended to participate in (Asan and Emeksiz 2016). Zhao et al. (2018) found that in addition to nature-based attractions, POIs could be categorized as cultural heritage, landscapes, and theme parks.

2.2.3 Applications of mobile data in tourism research

In tourism research, mobile data have been used principally to investigate traveller behaviour. Mobile data may be considered to provide a reliable account of tourists' activity choices, movement patterns and chronology, and be useful in identifying POIs (Xu et al. 2017). Previously, travel activity and trajectory information was typically gathered through survey-style data collection methods (Gursoy, Jurowski, and Uysal 2002; Collins, Rose, and Hess 2012). The influx of geolocation data has created new opportunities for tourism analysis. Geotagged posts of different kinds (Twitter, reviews, Instagram) have been used in tourism research to identify popular POIs and tourist sentiments (Rinne and Saastamoinen 2005; Kawakubo and Yanai 2009; Chen et al. 2018; Hasnat and Hasan 2018; Kirilenko, Stepchenkova, and Hernandez 2019). Mobile location data have been used similarly (Fei Hu, Qingquan Sun, and Qi Hao 2010; Phithakkitnukoon et al. 2015; Guo, Zhang, and Zhang 2016; Chen et al. 2018; Luo et al. 2018; Zhao et al. 2018; Kirilenko, Stepchenkova, and Hernandez 2019). Zhao et al. (2018) uses mobile location data to investigate travel behaviour and how it is impacted by party size. They found that large parties were likely to travel to attractions closer to their accommodations, whereas small groups were likely to be more adventurous and travel further. Chen et al. (2018) also investigated the effect of party size on travel trajectories, but used a combination of manual annotation, geofencing, and DBSCAN to identify POIs and study the frequent daily travel patterns to identify the most popular attractions.

2.3 Methodology used in literature

Methodologies used in tourism research and the modelling of human behaviour using mobile location data typically utilize the same steps, achieved through different means. These steps are as follows:

1. Preprocess and filter data
2. Identify tourist devices
3. Identify tourist attractions or points of interest
4. Results analysis

Methodologies used to achieve each of these steps in existing literature will be explored in the following sections.

2.3.1 Data preprocessing and filtering

The nature of mobile location data is untidy and noisy. It can be difficult to analyze tourist travel patterns because of the non-repetitive nature of movement patterns (Hasnat and Hasan 2018). Filtering these data prior to analysis can help mitigate these issues and increase the reliability of the results. Some device users produce less consistent data due to app use differences, subscription differences and changes, or having their phones turned off while abroad (Phithakkitnukoon et al. 2015). This issue can be addressed by only including devices that meet a set of criteria for study. Phithakkitnukoon et al. (2015) used only devices that had been observed for 350 days out of 365 in the previous year, in order to introduce some consistency into the dataset. Similarly, Guo, Zhang, and Zhang (2016) only included devices that had data from two or more latitude, longitude coordinates. This removes devices that are only being observed at one location, and can mitigate resting device issues, such as a home tablet that pings frequently and may skew a dataset even though it has not changed locations and is therefore not producing movement data. Xu et al. 2017 also found that single datum points such as single coordinate points are not helpful because they do not indicate any trajectory in movement and therefore cannot illuminate any patterns.

2.3.2 Identify tourist devices

Tourist devices can be identified through a detailed filtering process and may include a clustering technique, but typically includes the identification of a home location (Phithakkitnukoon et al. 2015; Guo, Zhang, and Zhang 2016; Hasnat and Hasan 2018; Zhao et al. 2018). This may be identified by assuming that devices will be in their home at night, and therefore nighttime coordinates can be used as a general home location (Hasnat and Hasan 2018). The limitation presented by this approach is that not every night is necessarily spent at home. Zhao et al. (2018) overcame this limitation using a general home region, in which any device present in a given area for the majority of days in the year is likely a resident of that area (65/90 days in their study). Another approach defines a home region as any location that a device resides in prior to entering a central business district (Guo, Zhang, and Zhang 2016). This approach is very limited and does

not consider the consistency of a home location over time. Phithakkitnukoon et al. (2015) addressed this issue by performing K-means clustering on identified temporal stops and having home locations confirmed by a sample of the survey subjects. In the absence of survey subjects, Zhao et al. (2018) performed ground-truthing on home locations by manually investigating 200 unique devices to confirm the assumed home regions.

2.3.3 Identify tourist attractions/points of interest

Chen et al. (2018) defined tourist attractions as frequently visited areas that overlap with attraction-type regions such as beaches or shopping outlets. A manual geofencing method with a spatial threshold of 500m was also successfully implemented by Zhao et al. (2018) in identifying tourist attractions. The methods proposed by Guo, Zhang, and Zhang (2016) may also be used for this purpose. They used hierarchical clustering on count data of how frequently a certain location was visited in a day. Their methodology identified particularly popular spots, rather than just all potential attractions as would be found using the methods from Chen et al. (2018). It may be that the distance to an attraction is a defining factor, as was found in Zhao et al. (2018), particularly in larger travel parties (Chen et al. 2018).

Chen et al. (2018) identified ROIs by categorizing them as either scenic regions or non-scenic regions. The ROIs were extracted from traveller “stay-points” identified by mobile data behaviour, in which a stay-point was defined as a series of consecutive data points within a 500m area and occurring within 30 minutes. Phithakkitnukoon et al. (2015) used a similar tactic, but employed an area of 196m and a temporal area of 14 minutes.

DBSCAN has often been used to identify POIs, particularly those that cannot be otherwise easily categorized (Phithakkitnukoon et al. 2015; Chen et al. 2018; Hasnat and Hasan 2018). Chen et al. (2018) used a minimum points value of 10,000 and an epsilon value of 5m. In contrast, Phithakkitnukoon et al. (2015) used a minimum points value of 5 and an epsilon value of 30m. Hasnat and Hasan (2018) used a minimum points value of 25 and an epsilon value of 1000m, however this was applied to geotagged online posts, rather than mobile location data. K-Means was also employed as a cluster mechanism to identify POIs. It was more effective than DBSCAN when applied to geotagged photos, and was consistent with tourism statistics (Hasnat and Hasan 2018), but DBSCAN was more effective in clustering accurately when applied to mobile location

data (Phithakkitnukoon et al. 2015). Kirilenko, Stepchenkova, and Hernandez (2019) used network analysis to identify attraction clusters.

Once the clusters are identified, several methods have been used to categorize the type of POI represented by each cluster of points. Some studies employed hand drawn or manually annotated geofences (Phithakkitnukoon et al. 2015; Chen et al. 2018), whereas Hasnat and Hasan (2018) produced a cluster centroid, buffered it by 3km, and then generated a list of attractions within that buffer. Zhao et al. (2018) manually geofenced the known scenic areas, applied DBSCAN to the non-scenic attraction clusters, and extracted a list of the top 29 attractions as rated by the China National Tourism Administration. Attraction popularity was measured by the number of return visits by the tourist device.

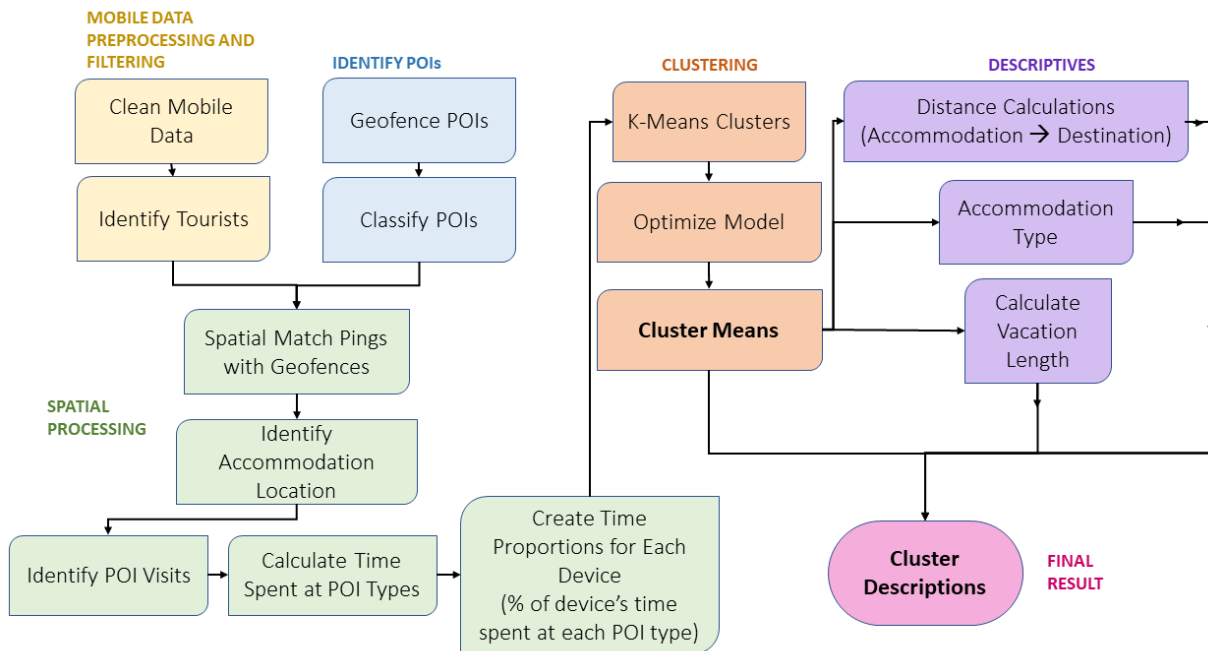
2.3.4 Clustering & results analysis

Methodological investigation is a theme in research that analyzes results typically by testing accuracy or comparing the accuracy of a variety of methods (Guo, Zhang, and Zhang 2016; Hasnat and Hasan 2018). Another theme in results analysis includes calculating the frequency of use or visits of POIs by tourists (Chen and Grauman 2011; Li et al. 2013; Kirilenko, Stepchenkova, and Hernandez 2019). This may involve first finding commonality between tourist devices to compare the differences, which can be accomplished by first segmenting the tourist devices. Party size (Zhao et al. 2018), home region (Phithakkitnukoon et al. 2015), and time of day (Chen et al. 2018) are some segmentation variables that have been used. K-Means clustering, DBSCAN, and other similar clustering algorithms may be used to identify these segments (Phithakkitnukoon et al. 2015; Zhao et al. 2018). The differences between traveller segments have been compared by measuring dissimilarities between trips and using the Mann-Whitney U test to see if the differences are statistically significant (Chen et al. 2018), and by using mean and standard deviation of different variables such as distance between POIs visited to compare their relationship to the traveller segment (Phithakkitnukoon et al. 2015; Zhao et al. 2018). Common variables identified as impacting travel behaviour are accommodation (Keogh 1984; Chen et al. 2018), party size (Zhao et al. 2018), home region (Phithakkitnukoon et al. 2015; Zhao et al. 2018), and travel motivation (Mkono, Markwell, and Wilson 2013; Asan and Emeksiz 2016).

CHAPTER 3: Data & Methodology

The methodology used in this study followed the steps identified in literature. This process began with the cleaning of the mobile data and the creation of the geofences (Figure 1). A large-scale spatial match was then performed between the geofences and mobile pings to allow for trip quantification. K-Means clustering was performed based on the proportion of each device's time spent at each type of POI, and descriptive statistics were created. The cluster means combined with the descriptives were the foundation upon which the cluster summaries were created (Figure 1).

Figure 1 – Flowchart of study methodology



3.1 Study data

The data used in this study were retrieved consensually and anonymously through app use by a third-party data provider. Users of said apps consented to the collection of their location data in exchange for using the app. These apps are embedded with a Software Development Kit (SDK) that collects a stream of times and locations from devices during app use along with a 25-digit identifier that is unique to each device. This is an identifier that is consistent over time for each device and is not linked to any other form of personal information. Phone numbers, IP addresses, credit card information, and any other kind of personal information are not collected, accessed, or

accessible through this data stream. The times, locations (in the form of a latitude and longitude) and unique device IDs are collected each time the app is accessed and continuously while the app is in constant use. There is up to a 10m error on all the location data collected, so while these data are being produced at a very fine scale, they are only reliable on a slightly more general level. For the purposes of this study, a device registered in the database was considered to be a person, or member of the study population.

3.2 Data preprocessing and filtering

Due to limitations of computing power, the dataset was generalized to reduce its size. All data points had the latitude and longitude floored to 4 decimal degrees. This reduced the accuracy of the data but also significantly reduced the number of data points to be assessed as the data points exist only to 10m accuracy, rather than 1m or smaller. The data were also aggregated based on the time stamps, with each data point having the time generalized to a ten-minute interval. For example, mobile data points occurring at 10:32am, 10:35am, and 10:39am were all considered to be in the 10:30am ten-minute block, comprised of all time points between 10:30:00am and 10:39:59am. Data points occurring at 10:40am, 10:47am, and 10:49am were considered to be in the 10:40am ten-minute block of 10:40:00am to 10:49:59am. These ten-minute blocks were referred to as 10:30am and 10:40am respectively. Each ten-minute block was associated with a single latitude/longitude coordinate, which was the mean coordinate occurring during the given ten-minute block. Several filters were applied to the data in order to identify only the desired devices for study. Some devices were eliminated due to their ping frequency. Preliminary data investigation found that there were some devices that had a ping frequency that far exceeded that of other devices, as well as devices that had a particularly low ping frequency. “Ping frequency” is defined as the number of times a device’s time and location is registered in the database. Devices with a ping frequency in the top 1% and bottom 1% were filtered out to introduce some consistency into the dataset. Additionally, devices that pinged only at one location were removed from the dataset, like Guo, Zhang, and Zhang (2016), to mitigate the presence of a device that does not provide any travel trajectory information and to eliminate resting devices such as home tablets that produce a high frequency of pings but only at one location. In order to be included in the dataset, the device had to be observed at a minimum of two locations at least 100m apart. A study period of December 1st, 2017 to January 31st, 2018 was selected as this is typically the high season of

tourism in Maui and therefore was likely to have the largest tourist dataset compared to any other time of year. Due to differences in seasonality of tourism experiences, only this short period of time was selected in order to reduce the variability in other external factors that might affect the proportionate representation of tourists throughout the year.

3.3 Identifying tourist devices

Tourist devices were identified using their most common yearly night time location, as devices were assumed to be at home most nights (Hasnat and Hasan 2018; Zhao et al. 2018). This home location was based on the dominant 3 decimal degree latitude and longitude most commonly observed (via ping frequency) during night time hours 6:00pm to 8:00am throughout the entire year of 2017. Devices with a home location outside of the state of Hawai'i were considered non-residents of the state. Tourists to Maui were then identified as non-resident devices that were observed in Maui, Hawai'i within the study period (December 2017 and January 2018). To further ensure that only tourist devices were being detected, only devices that had been present in Maui for 150 days or fewer in 2017 were kept in the dataset. This helped to exclude devices that may have lived in Maui for some of 2017 and may not have been picked up in the filter for non-residents.

3.4 Identify tourist attractions/points of interest & accommodations

Like Chen et al. (2018), geofences were created manually and annotated to create attraction POIs and accommodation POIs. The geofences were created using the most current (2019) data available on Google Maps. Satellite imagery was inspected and each location that either appeared to be likely a business or POI of some type or was listed as such by Google was investigated. This investigation took the form of finding a business website with an address matching the location on the map, and/or investigating the location using Google Street View to confirm the presence of the POI in that location and confirm the extent of the property. Then, each POI was geofenced and annotated with the name of the POI, the class (attraction or accommodation), and the type (Figure 1). The POI types used for classification in this study were based on combined attraction types found in literature (Chen et al. 2018; Kirilenko, Stepchenkova, and Hernandez 2019) and include 17 attraction types and 9 accommodation types (Table 1). Large areas consisting of the same

attraction type, such as an outlet shopping centre, or power centre were geofenced as one location. All accommodation geofences except residential ones were made to specifically delineate the businesses in order to allow for accurate distance to destination calculations. Residential accommodation geofences were made around what appeared from the satellite imagery to be neighbourhood areas. Like Hasnat and Hasan (2018), accommodation location was identified by selecting the most dominant coordinates during nighttime hours of 6:00pm to 8:00am. It was assumed that a tourist in Maui would stay at the same accommodation for the entire trip, and likely return there at some point each night. Accommodation type for each device was classified based on the spatial match of those coordinates with an annotated accommodation geofence.

3.5 Trip identification

Once tourists and POIs had been identified, attraction visits were quantified. First, the speed of each device per data point was calculated. This was accomplished by measuring the distance between each data point per device and calculating the time passed. Distance was then divided by time to retrieve the km/h for each data point. Points that had moved at a speed of over 3km/h were removed. This is because 3km/h is a typical walking speed. This filter did not remove devices, it only removed the data points that occurred when the device was moving faster than 3km/h. This ensured that when the sequential points were identified for a visit, it only included walking speed points. This removed road traffic (devices in cars moving faster than walking speed) and added assurance that attraction visits were being calculated using data points that suggest the person was engaged in an activity related to the attraction type, and not driving past it. An attraction visit was defined as a sequence of time exceeding 30 minutes and not surpassing 480 minutes within a geofence. An upper limit of 480 minutes was chosen because an attraction visit exceeding this amount of time is likely a case of a resting device, due to errors in the data or the geofences, or some other issue that is outside the scope of this analysis.

Table 1 – Classification parameters of point of interest types per class, for geofence annotation

Class	Type	Description	Example Key Words
Attraction Types	Adventure	Land-based outdoor activities that involve hiking, deep forest, and/or risk of danger.	Hiking, Water Falls, Crater, Ziplining, Safari, Lava Fields
	Arts Recreation	Any activity that is focused around arts or music.	Luau, Gallery, Cinema, Music, Theatre
	Athletic Club	Any physical recreation outdoor or indoor, land or water, that is somewhat casual in nature but is associated with a membership to a club, marina, etc., and is typically associated with higher activity costs.	Country Club, Golfing, Tennis Club, Yacht Club
	Basic Retail	Stores offering basic amenities.	Grocery, Pharmacy
	Boating	Pleasure craft boating - not fishing.	Harbour, Pier, Port
	Casual Nature	Outdoor experiences that involve walking, strolling, in outdoor areas that are organized in some way - a garden, park, etc.	Park, Plantation, Botanical Garden
	Casual Waterfront	Waterfront experiences associated with lounging (beach), and casual walking to waterfront scenic areas such as lighthouses, points, bays.	Beach, Boardwalk, Lookout, Bay, Lighthouse
	Gastronomic	Any establishment associated with a culinary experience or service.	Restaurant, Food Truck, Eatery, Fast Food
	Historic Educational	Activities that include any kind of historical or cultural learning, or any educational activities.	Library, Museums, Aquarium, History, Veterans
	Private Events	Venues for private events.	Events, Weddings
	Retail	Shopping amenities not including basic amenities such as grocery stores.	Boutiques, Outlet Shops, Clothing Store
	Rural Activity	Farm or animal related activities that take place in a rural setting.	Farm, Trail Rides, Horseback Riding, Ranch
	Spa	Facilities that offer spa services.	Manicure, Massage, Spa Treatment
	Spiritual Recreation	Attractions related to faith, spirituality, and religion.	Chapel, Cathedral, Temple, Cemetery, Memorial, Shrine, Psychic
	Sports & Fitness	Public/community-based fitness complexes and centres, pools, gyms, skate park, rental bikes, etc.	Aquatic Centre, Gym, Community Centre, Local Arena, Recreation Centre, Skate Park, Rental Bikes
	Transit	Transportation hubs.	Airports, Car Rental Facilities
	Water Sports	Any sport conducted in water including kitesurfing, surfing, rafting, fishing. Yachts/sailing are not included due to the different nature of these activities and the typical cost/membership requirements.	Surfing, Snorkeling, Kayaking, Diving, Fishing, Paragliding
Accommodation Types	Apartment Rental	Apartments of any type available for rent.	Condominium, Airbnb, Townhouse Rental, Apartment
	Hostel	A lodging business self identifying as a hostel.	
	Hotel	A typical hotel business not offering any resort type amenities.	Hotel, Inn
	Private House	Private rental home, typically that has its own staff.	Villa, Estate, Beach House
	Residential	Neighbourhood areas of residential homes.	
	Resort	Any largescale hotel offering extensive on-site amenities including shopping, spa, pool, cuisine, etc.	Resort, Retreat
	Small-scale Lodging	An accommodation that may offer some staff services but is a smaller and more personable size.	Guest House, Bed And Breakfast
	Camp	An accommodation that is mostly outside.	Campgrounds, Cabins
	Cruise Ship	Cruise ships that are docked in Maui with a visible location on satellite images.	

For a data point to be included in an attraction visit, each proceeding data point needed to occur within the same geofence as the first and within 1 hour of the previous data point (Table 2). The succession of qualifying points was then assigned a trip identification number for data organization. Each trip was then classified as an attraction type trip based on the geofence type it occurred within. The length of each trip was calculated in minutes by subtracting the first date and time of the trip from the last qualifying date and time of the trip. Devices were only included in further analysis if they had more than one attraction visit (including own accommodation visits). Additionally, average trips per detection day were calculated for each device. This was calculated by first counting the number of unique days that the device had produced data records, and counting the total number of trips taken. Total trips were then divided by total detected days to retrieve the average trips per detected day. Only devices with 1 or more trips taken per detected day on average were included in the study population. This was to introduce more consistency and robustness into the dataset and produce a study population with travel habits that could be reasonably compared to one another, as in Phithakkitnukoon et al. (2015). Furthermore, average trips per detected day were used because it was not affected by the actual vacation length for each device. The device's vacation could be of any length so long as there was at least one trip being detected per day. It was also required that the device register an attraction visit that was in a location further than 30m from its accommodation location. This was to further ensure that resting devices were not included, and to include only devices with varied trip movement in the dataset.

Accommodation location visits were also included as POI visits, but a device could only log an accommodation visit at its own accommodation, rather than at any accommodation location. The accommodation visit was also only counted if it occurred during the day (between 8:00am and 6:00pm) to distinguish nights spent sleeping at an accommodation versus spending actual day time there. It was important to include accommodation visits to distinguish between tourists who spent most of their time at their accommodation (such as at a resort) and those who only used their accommodation as lodging, spending the majority of their consecutive trip minutes out at other attractions.

Table 2 – Example of trip identification, where suitable trips are highlighted with red boxes

Device ID	Geofence ID	Date/Time Stamp
12481	62	2018-01-08 0:40
12481	62	2018-01-08 1:40
12481	62	2018-01-08 2:20
12481	62	2018-01-08 2:50
12481	62	2018-01-08 4:20
12481	909	2018-01-08 4:20
12481	62	2018-01-08 8:30
15405	810	2017-12-04 8:40
15405	810	2017-12-04 9:40
10568	588	2018-01-18 20:10
10568	451	2018-01-19 6:20
10568	452	2018-01-19 7:40
10568	120	2018-01-19 14:20
10568	120	2018-01-19 14:30
10568	120	2018-01-19 15:20
10568	120	2018-01-19 16:20
10568	120	2018-01-19 16:30
10568	120	2018-01-19 16:40

3.6 K-Means clustering & cluster summaries

Tourists were first segmented to illuminate underlying patterns of travel behaviour (Phithakkitnukoon et al. 2015; Chen et al. 2018; Zhao et al. 2018). Tourists were segmented using K-Means clustering based on the proportion of their time spent at different attraction types as well as daytime spent at their own accommodation. Distance to attraction from lodging (Zhao et al. 2018), lodging type (Keogh 1984; Chen et al. 2018), and total vacation length (Chen et al. 2018) were all variables that were summarized as a part of the descriptive investigation of these travel groups. Total vacation length was calculated by subtracting the last day the device was detected from the first day the device was detected within the study period.

K-Means clustering was performed in SPSS using the proportion of each device's total visit minutes being spent at different attraction types. Total visit minutes were defined as the sum of the length of all observed qualifying visits. Visit minutes were then also summed by POI type, and each was divided by the total visit minutes to produce the proportion of time spent at each POI type. The data were calculated into proportions as a means of normalizing the data, particularly to mitigate the differences in vacation lengths affecting the total number of visit minutes observed. Variables that did not meet significance in the K-Means model were not removed, as the values for each device were proportions of a whole. Even if a variable was insignificant in the division of

the cluster groups, it was still representing a portion of the devices time and for that reason was left in. The model was run beginning at 10 clusters and 100 iterations. After inspecting the homogeneity of the boxplots and inspecting the number of cases in each cluster, clusters with fewer than 10 cases were removed, and the remaining 7 clusters centres were readjusted. The model optimization was ended when the boxplots were suitably homogenous based on a visual inspection. The dispersion of the accommodation locations in each cluster were inspected to determine differences in spatial distribution between the clusters.

CHAPTER 4: Results

4.1 Identify tourists

Data filtering resulted in a significant reduction in study population size. Overall, the population was reduced from 29,407 to 428 (Table 3). The most significant filter was that which removed devices that had been observed making fewer than one POI visits per day that the device was observed in Maui (Table 3). The filter that removed devices observed at only one three decimal latitude and longitude location also removed many devices (Table 3).

Table 3 – Data filtering results

Data Filter Applied	Device Count
Starting population, Maui December 2017 to January 2018, devices that have a home state outside of Hawai'i and have been observed in Maui for <=150 days in 2017	29,407
Devices observed at >1 latitude and longitude, rounded to 3 decimals (i.e. observed at >1 100m block)	19,619
Devices with a common night time location falling within an accommodation geofence	13,408
After removing the top and bottom 1% of high and low ping frequency devices	13,139
After removing records that fell within 30m of the device's accommodation location (not all devices had pings outside of this location)	12,708
After removing records during times when the device was moving at >3km/h	12,708
Devices that had POI visits of any length of time	9,976
Devices that had POI visits >30mins and daytime accommodation visits >30mins	7,755
Devices that visited non-transit related POIs	6,997
Devices that were observed making a minimum of one POI or accommodation visit per day in Maui	428
Devices that made a total of >1 POI visits	428
Final device count	428

4.2 Identify points of interest and accommodations

In total, 893 geofences were created (Table 4). Gastronomic POIs were the largest category, with 152 geofences. Spiritual Recreation, Retail, Residential, and Casual Waterfront were the next largest POI categories with between 70 and 85 geofences each. Cruise ship was the smallest category with only 1 geofence, followed by Hostel with 3 geofences, and Transit and Camp with 6 geofences each (Table 4). There were no devices with a Camp geofence as an accommodation and so these geofences were excluded from further summaries. Similarly, there were no trips taken

within Spa or Rural Activity geofences that met trip criteria, and for this reason they were omitted from further summaries. Transit geofences were also excluded from the analysis as they contained airports and car rental facilities which were not a part of the focus of this study. Some geofences covered large areas in cases such as a hiking activity occurring in a forest. In this case, the whole forest area where the hiking might be taking place was geofenced as one point of interest (Figure 1). Attraction geofences did not overlap with one another so that a device's presence in a geofence is associated with only one distinct activity (Figure 1). Accommodation geofences were allowed to overlap in the cases of rental apartments or rental houses within a residential area (Figure 1). The most delineated accommodation geofence was always selected to ensure that the correct accommodation was being selected for each device. For example, if a device's most common nighttime coordinates fell within a listed rental apartment that was within a residential area, the rental apartment geofence was selected rather than the residential area geofence.

Table 4 – Geofence summary of point of interest types

Geofence POI Type	Count
Adventure	29
Apartment Rental	53
Arts Recreation	27
Athletic Club	19
Basic Retail	18
Boating	7
Camp	6
Casual Nature	24
Casual Waterfront	71
Cruise Ship	1
Gastronomic	152
Historic Educational	20
Hostel	3
Hotel	17
Private Events	7
Private House	45
Residential	84
Resort	42
Retail	77
Rural Activity	13
Small-Scale Lodging	43
Spa	13
Spiritual Recreation	82
Sports & Fitness	13
Transit	6
Water Sports	21
Total	893

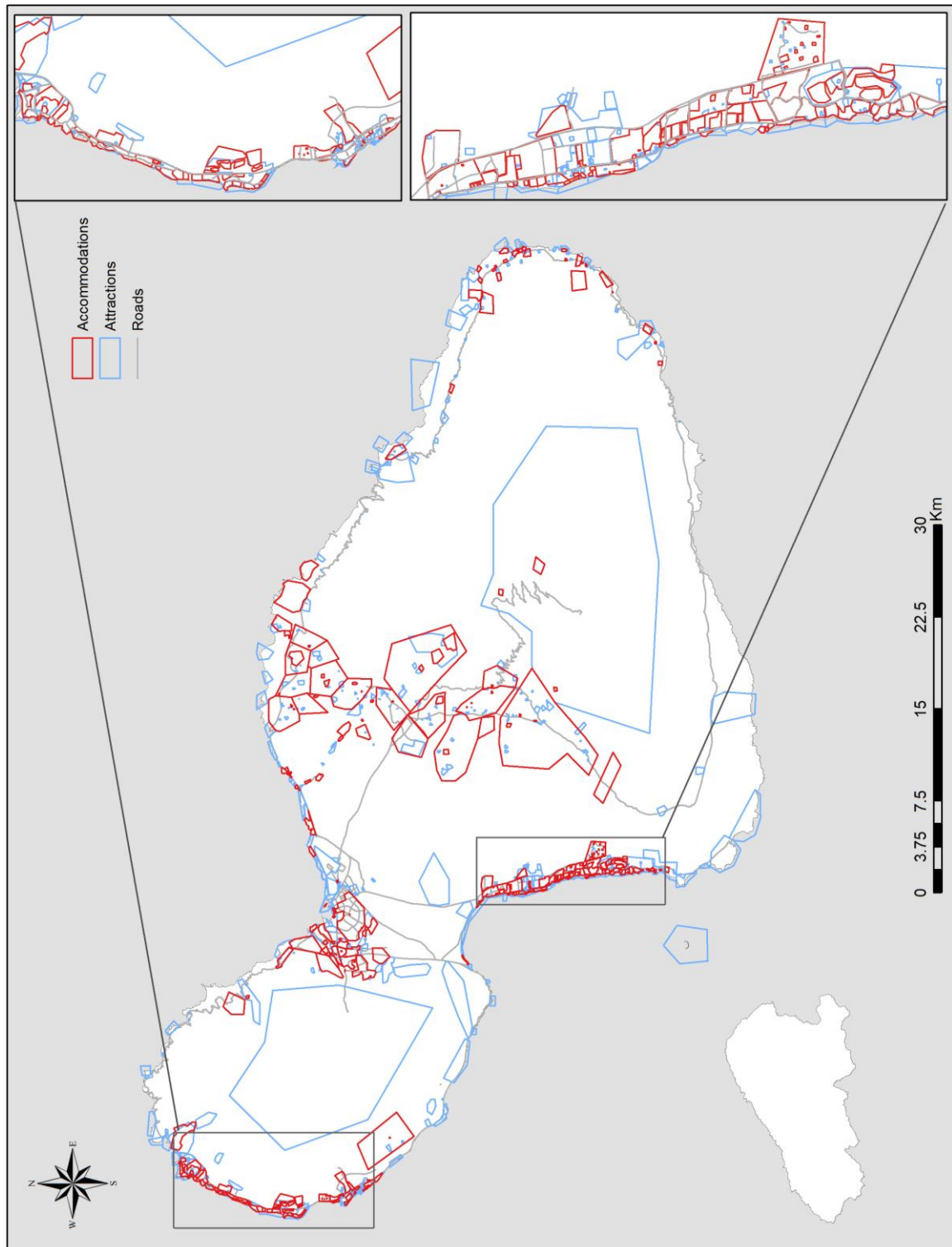


Figure 2 – Geofences of attractions and accommodations in Maui, Hawai'i. Based on imagery and business information from Google Maps, 2019

4.3 K-Means clustering & cluster summaries

On average, 61 devices were present in each cluster (Table 5). Cluster 1 and cluster 4 both contained a significantly higher number of devices, with 148 and 127 devices respectively. Cluster 3 contained the fewest devices, with only 20 (Table 5).

Table 5– Number of cases in each cluster

Cluster	Number of Cases
Avg # Dev per Cluster	61
1	148
2	24
3	20
4	127
5	36
6	44
7	29
Total	428

ANOVA test results found 13 of the 20 variables met significance with a significance value below the 0.05 threshold (Table 6). Arts recreation, historic educational, private events, spiritual recreation, sports and fitness, and small-scale lodging did not meet significance (Table 6).

Table 6 – ANOVA test results

Variable	Sig.
Adventure	0.000
Arts Recreation	0.516
Athletic Club	0.000
Basic Retail	0.000
Boating	0.012
Casual Nature	0.014
Casual Waterfront	0.000
Gastronomic	0.000
Retail	0.000
Water Sports	0.070
Historic Educational	0.372
Private Events	0.923
Spiritual Recreation	0.769
Sports & Fitness	0.862
Apartment Rental	0.000
Hotel	0.000
Private House	0.000
Residential	0.000
Resort	0.000
Small-scale Lodging	0.345

Final cluster centres show the mean of each variable per cluster. These values show what the typical values might be for any given device in the cluster, and in this case, the typical proportions of time being spent at different attraction types. Casual waterfront was the primary variable distinguishing Cluster 1 with a mean of 24.3% (Table 7). Retail defined Cluster 2 with a

mean of 70.76%. Cluster 3 was mainly defined by hotels with a mean of 57.19%. For Cluster 4, resorts were the distinguishing variable with a mean of 73.46%. Cluster 5 was heavily defined by private housing, with a mean of 68.57% (Table 7). Basic retail, with a mean of 45.85%, distinguished Cluster 6. Cluster 7 was most heavily defined by athletic clubs, with a mean of 64.59% (Table 7). Cruise ships were not included in the clustering model because people staying on cruise ships did not spend any extended time there during the daytime hours, and so there were no cases where time spent at a cruise ship was a part of the overall proportion of time spent during the day.

Table 7 – Final cluster centres: average proportion of time spent at each activity type, per cluster

Cluster	1	2	3	4	5	6	7
Arts Recreation	1.52	0.18	1.21	0.66	0.79	0.13	0.00
Athletic Club	1.90	1.58	2.17	1.66	2.13	1.40	64.59
Sports & Fitness	0.79	0.00	0.38	0.23	0.00	0.36	0.00
Historic Educational	0.86	0.16	1.54	0.33	0.22	0.48	0.10
Private Events	0.33	0.00	0.00	0.04	0.00	0.46	0.00
Spiritual Recreation	0.78	0.00	0.00	0.02	0.20	0.17	0.00
Gastronomic	15.68	7.37	11.56	5.48	6.96	8.07	7.02
Basic Retail	4.61	0.76	8.88	1.76	2.56	45.85	0.57
Retail	6.94	70.76	1.61	7.64	6.10	3.60	6.40
Boating	4.85	0.00	0.00	0.16	1.42	1.73	0.08
Adventure	13.42	2.26	6.18	2.84	3.20	3.35	2.02
Casual Waterfront	24.30	2.59	5.43	4.99	6.70	4.11	3.62
Casual Nature	1.84	0.06	3.40	0.44	0.33	3.94	1.58
Water Sports	2.67	0.08	0.44	0.28	0.83	2.28	0.93
Apartment Rental	9.69	0.00	0.00	0.00	0.00	1.04	0.00
Hotel	0.83	0.00	57.19	0.00	0.00	0.98	0.87
Private House	0.70	0.33	0.00	0.00	68.57	0.00	4.49
Residential	0.45	0.00	0.00	0.00	0.00	18.93	0.36
Resort	6.26	13.87	0.00	73.46	0.00	3.12	7.38
Small-scale Lodging	1.58	0.00	0.00	0.00	0.00	0.00	0.00

Box and whisker plots of the clusters and the case distance from cluster centre depict the homogeneity of each cluster. Cluster 1 is the least homogenous group with the largest overall range of distances from the cluster centre (about 18 to 95) and containing the most outliers above the upper quartile (Figure 2). Cluster 1 has the largest range in its upper quartile. It contains three outliers that are furthest from the cluster centre, slightly above 100. The interquartile range shows an even distribution above and below the mean, which has a value slightly under 50. Cluster 1 has the highest mean distance from its cluster centre than any of the other clusters. Cluster 2 has a smaller range in its upper and lower quartiles (Figure 2). The mean value is slightly above 30, and the interquartile range shows a higher proportion of cases in the cluster that fall above the mean, with a larger range of cases falling in the interquartile range below the mean. Cluster 2 is a fairly

homogenous group with distance values ranging from around 8 to 60, and containing no outliers. Cluster 3 has a similar range to Cluster 2, about 15 to 50, but has a slightly higher mean at about 35. Cluster 4 is quite homogenous, with the lowest mean of all the clusters at about 25, the shortest interquartile range, and overall range from about 6 to 42. Cluster 4 also includes one outlier slightly above 50 (Figure 2). Cluster 5 is similar to Cluster 4 in range and homogeneity, but has two outliers falling around 55 and 60. Cluster 5 has a mean of around 30. Cluster 6 is similar to Cluster 1 in range and homogeneity, with an overall range from 10 to 70. Cluster 6 has an interquartile range that spans a large distance with a mean around 45. The interquartile range shows a slight skew towards a more homogenous distance from cluster centre below the mean, with more of a range in values above the mean. Cluster 7 is the most homogenous cluster of them all, with a mean of 30, and range from 20 to 50, and a well balanced interquartile range with an equal number of cases above and below the mean. The upper and lower quartile ranges in Cluster 7 are quite small (Figure 2).

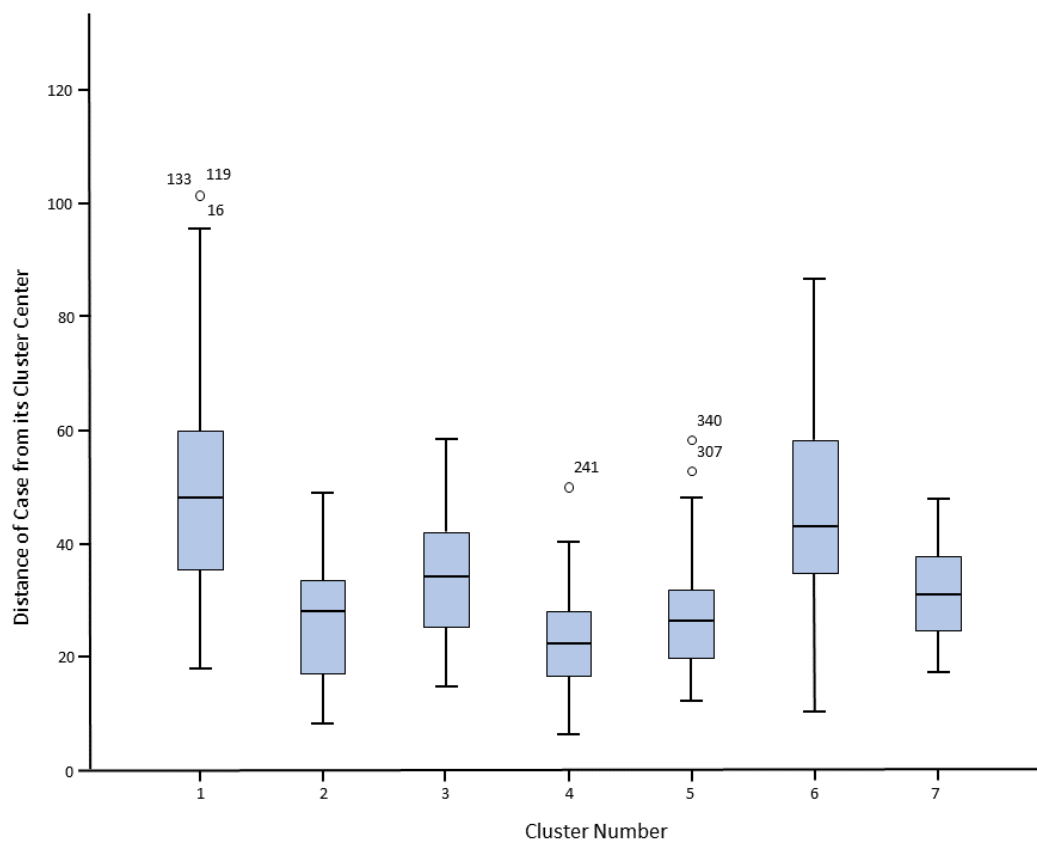


Figure 3 – Homogeneity of cases within each cluster, based on their distance from the cluster centre

Cluster 1 stayed predominantly in resorts and in apartment rentals (Table 8). The majority of Cluster 2 stayed in resorts. Cluster 3 stayed exclusively in hotels, and Cluster 4 stayed exclusively in resorts. Cluster 5 stayed only in private houses. Cluster 6 stayed mostly in residential areas and at resorts. Cluster 7 stayed mainly at resorts, with some staying at private houses and hotels (Table 8).

Table 8– Based on the percentage of devices staying at each accommodation type

Cluster	Apartment Rental	Hotel	Private House	Residential	Resort	Small- scale Lodging	Cruise ship	Residential Area
Base Avg (%)	51.6	11.2	12.4	14	3	0.9	0.2	6.5
1	35.1	10.1	5.4	6.8	33.8	2.7	6.1	0
2	4.2	8.3	8.3	4.2	75	0	0	0
3	0	100	0	0	0	0	0	0
4	0	0	0	0	100	0	0	0
5	0	0	100	0	0	0	0	0
6	9.1	13.6	0	36.4	29.5	0	9.1	2.3
7	10.3	17.2	24.1	3.4	44.8	0	0	0

Cluster 6 and 7 had the highest average vacation lengths at 11.7 and 12.2 days respectively (Table 9). Cluster 2 had the shortest average vacation length at 7.6 days. Cluster 2 also had the shortest distance between their accommodation and attraction destinations, at 3.8Km. By contrast, Cluster 6 had the highest distance between accommodation and attraction destination at 16.2Km. Cluster 2 also had the lowest average distance travelled between sequential attractions at 2.1Km, and Clusters 3 and 6 had the furthest average distance between sequential attractions at 6.9Km and 6.5Km respectively (Table 9).

Table 9 – Summary of cluster vacation lengths and travelling range

Cluster	Vacation Length (Days)	Avg Distance Between Attraction and Accommodation (Km)
Base Avg	9.3	7.5
1	9.1	12.9
2	7.6	3.8
3	9.1	10.9
4	8.4	9.1
5	8.9	10.8
6	11.7	16.2
7	12.2	7.8

4.3.1 Time spent at attraction types

Cluster 1 was the most varied in terms of how time was spent on average, spending time at every attraction type (Table 7). Cluster 1 spent the most time at their accommodations, but also spent the most time at private events, at casual waterfront attractions, and athletic clubs. Cluster 2 spent most of their time at retail locations. Cluster 3 spent most of their time at their hotel accommodations, with some time spent at casual nature attractions. Cluster 4 spent most of their time at their resort accommodations (Table 7). Cluster 5 mainly spent time at their private house accommodations, and Cluster 6 spent the most time at their accommodations as well as basic retail locations. Cluster 7 spent the most time at athletic clubs, and some time at their accommodations (Table 7).

4.3.2 Distance travelled accommodation to attraction

The distance calculation from accommodation to attraction does not necessarily mean that the device travelled from their accommodation to their destination for that visit. It is simply the distance between these two points as a way of indicating how far on average they might be willing to travel from the accommodation as an overall travel range. All clusters travelled furthest when seeking out adventure activities (Table 10). Cluster 1 went furthest from their accommodations for water sport attractions. Cluster 2 travelled furthest from their accommodations for adventure attractions, as well as basic retail locations (Table 10). Cluster 3 travelled furthest for arts recreation attractions and water sports. Clusters 4 and 5 travelled furthest from their accommodations for adventure attractions and historic educational attractions (Table 10). Cluster 6 travelled furthest for arts recreation attractions and casual nature locations, and to a lesser degree, sports and fitness attractions, private events, and water sports. Cluster 7 travelled furthest from their accommodation for adventure attractions (Table 1).

Table 10 – Average distance between attraction and accommodation, by attraction type

Cluster	Arts Recreation	Athletic Club	Sports & Fitness	Historic Educational	Private Events	Spiritual Recreation	Gastronomic	Basic Retail	Retail	Boating	Adventure	Casual Waterfront	Casual Nature	Water Sports	Apartment Rental	Hotel	Private House	Residential Resort	Small-scale Lodging
Population Avg (km)	11.3	8.4	9.7	14.3	9.2	9.1	9.1	15.1	4.2	11.5	31.2	9.1	17	21	0	0	0	0	0
1	9.5	16.4	6.8	12.4	4.6	7.8	9.5	15.1	6.2	9.6	31.4	10	15.3	24.8	0	0	0	0	0
2	3.3	0.4	0	0.7	0	0	6	25.8	2.3	0	39.2	4.8	21.1	5.5	0	0	0	0	0
3	31.1	4.5	2.9	17	0	0	8.3	9.8	4.6	0	22.4	10.3	7.9	34.2	0	0	0	0	0
4	5.6	5.7	8.9	20.2	10.4	8.3	8.5	18	3	18.9	29.3	6	12.2	14.4	0	0	0	0	0
5	8.6	3.5	0	23.3	0	7	10.4	17.1	5	24.7	34	5.6	19	15.6	0	0	0	0	0
6	34.4	11.3	23.4	5	20.7	17.6	12.9	14.8	12	5.3	37.6	15	25.8	19.3	0	0	0	0	0
7	0	8.2	0	6.2	0	0	4.5	14.4	5.7	7.1	23.4	7.6	2.2	8.3	0	0	0	0	0

4.3.3 Cluster descriptions

Cluster 1: “*Adventurous Explorers*” – Stayed predominantly in apartments and in resorts. This cluster had low values across time spent at their accommodation, meaning that most of their prolonged time at POIs during the day occurred outside their accommodation. Adventurous Explorers were most likely to be spending time at the waterfront engaging in casual activities such as beach lounging and walking along ocean front boardwalks. This group was the most adventurous and spent more time than other clusters seeking out adventure activities, water sports, and finding places to eat. Overall, members of this group took longer vacations than average across the study population, and may have been more willing to venture further from their accommodations when seeking out POIs. They were willing to travel furthest for water sports and adventure activities. This group was the largest of all the clusters, with 148 members.

Cluster 2: “*Big Shoppers*” – Stayed predominantly in resorts. This cluster spent some of their time during the day at their accommodation, but mostly spent their days shopping at outlets and boutiques. They were willing to travel a fair distance to find places to shop at 6.2km, the second furthest range of all the groups. Going out to eat was the only other activity this group spent significant time doing, on average. They had the shortest vacations of any other cluster and travelled the shortest distances to their destinations.

Cluster 3: “*Staying In*” – Stayed exclusively in hotels and spent the majority of their time there during the day. When leaving their accommodations, it was to go out to eat or shop for necessities. The small amount of time they spent seeking out adventure activities, they were willing to travel far from their hotels. This group took longer vacations than average, typically staying over a week. This was the smallest cluster, with only 20 members.

Cluster 4: “*All Included*” – Stayed exclusively in resorts, and spent the majority of their time there during the day, likely engaging in resort-offered activities. When venturing out from their resorts, this group was shopping at the outlets and boutiques nearby. The rare times they went out to the beach or to have a meal away from the resort, they stayed close by and did not venture far. All Included tended to stay in Maui for a little over a week on average. This group was the second largest, with 127 members.

Cluster 5: “*Upscale Getaway*” - Stayed exclusively in private houses and spent most of their time there during the day, likely enjoying the offered amenities on-site. Sometimes this group would go out for a meal, shop at the outlets and boutiques, or to spend time near the ocean. This group travelled furthest for the very rare times they would seek out historic educational attractions or go boating.

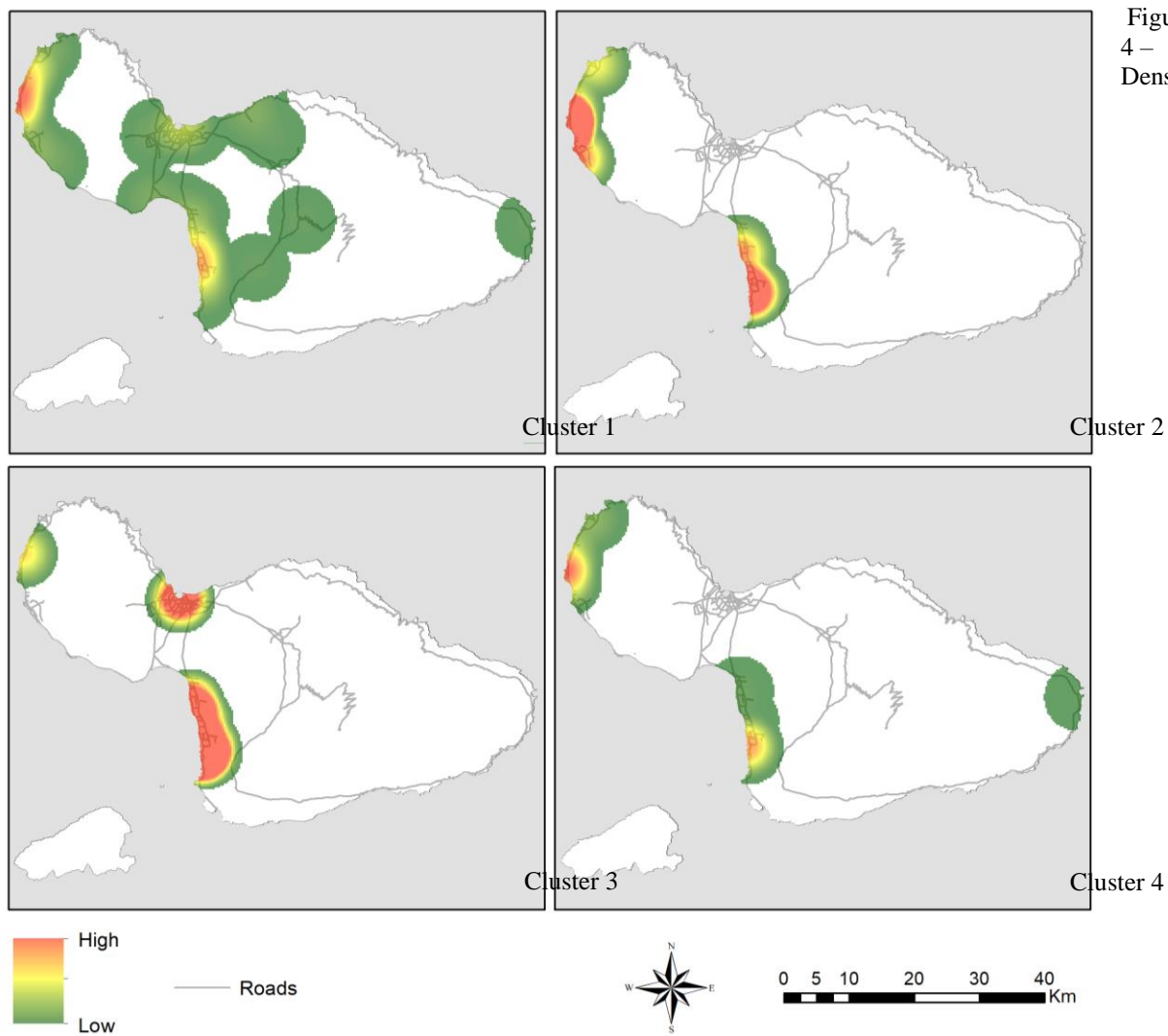
Cluster 6: “*Home Away From Home*” – This group contained members staying at a variety of accommodation types. Most stayed in homes in residential areas and spent a portion of their days there, but a similar proportion of members stayed in resorts. These accommodation types are typically associated with quite different travel experiences, and unsurprisingly this group is one of the most heterogeneous of the all the clusters (Figure 2). This group has the second longest vacations in Maui with an average of 11.7 days. They also had the largest travel range within the island, 16km on average. These higher averages may be the result of a larger range of traveller experience types contained within the same cluster. The majority of the day time activities included visits to areas providing basic shopping amenities. In the case of those staying in residential areas, this would be a part of the overall experience of staying in a home away from home. Other than staying in or shopping for basic amenities, this group spent some of their time travelling a fair distance of 12.9km on average to find new places to dine out.

Cluster 7: “*Country Clubbers*” – Stayed predominantly in resorts, with a smaller proportion staying in private houses, but spent little time there during the day. Most of their time was spent at athletic clubs, likely golfing, playing tennis, and enjoying other activities available only at private clubs. Some of their time was spent venturing to somewhere near their accommodation to find a place to eat or shop. This group was the most homogenous of all (Figure 2).

4.3.4 Spatial distribution of cluster accommodation locations

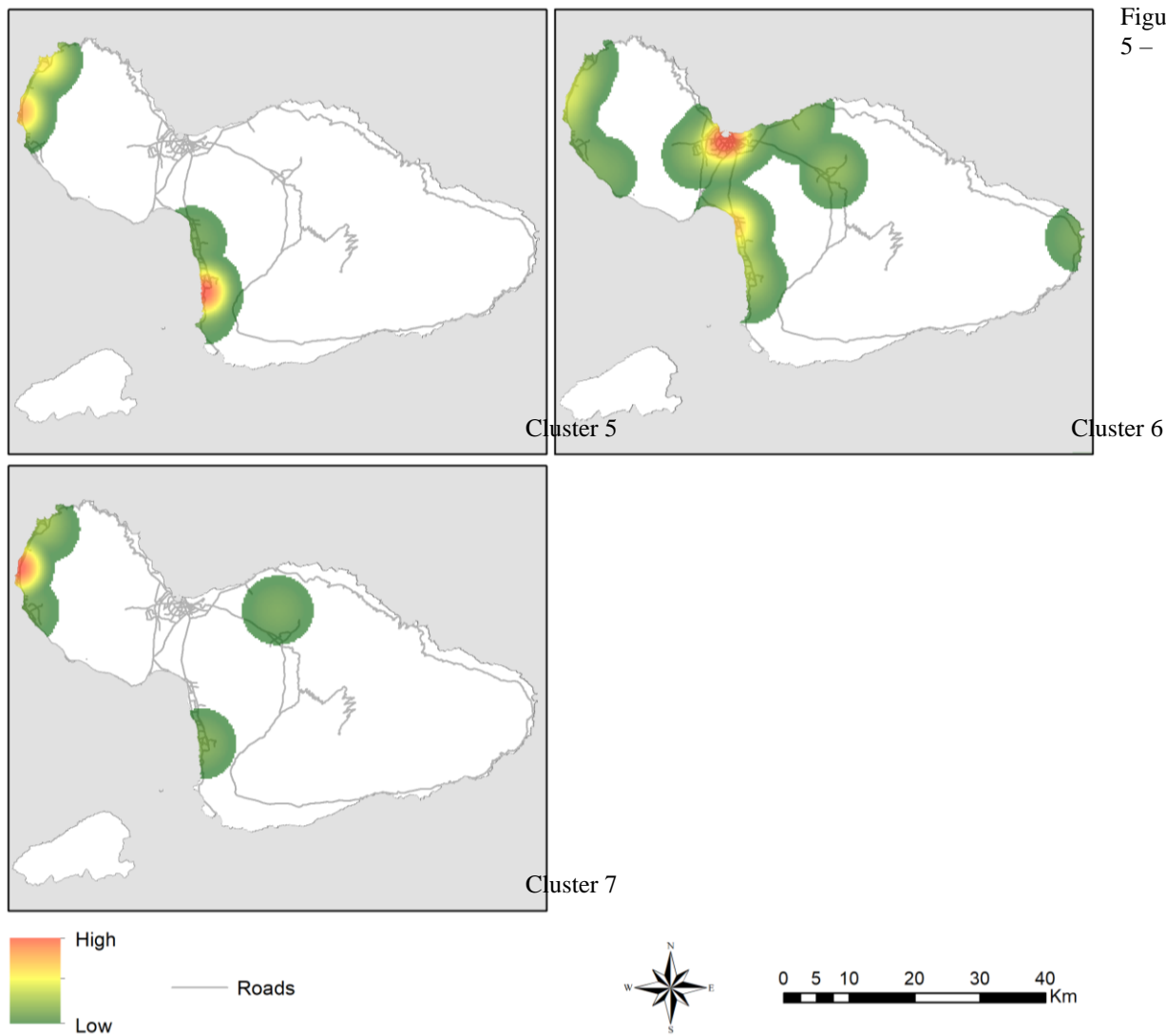
The accommodation locations of the devices in each cluster were mapped to display the differences in spatial distribution of lodging choices. Figures 4 and 5 display the particularly dense areas of accommodation locations per clusters. In this visualization, red areas were most densely populated with accommodation locations, followed by yellow areas of medium density. Areas without any colour did not have any accommodation locations in the cluster, and green areas had some presence of cluster accommodation locations but at a very low density. All clusters showed

some degree of presence along the two western coastlines. Clusters 1 and 6 had the most varied dispersion across the island, with only a few small areas of high density (Figures 4 and 5). Cluster 1 had the highest density in small areas on the western shorelines, and cluster 6 had the highest density on the west and northern shorelines. Clusters 2 and 3 both had a few very dense areas with little variation in density across the island (Figure 4).



distribution of accommodation locations by cluster

Figure 5 –



Density distribution of accommodation locations by cluster

CHAPTER 5: Discussion & Conclusion

5.1 Segmentation of tourists using time spent at point of interest types & differences between segments

Tourists were successfully segmented based on the proportion of time spent at different POI types. The segments had distinct characteristics and illuminated a variety of travel experiences that are defined by how consecutive time was spent, including time spent at the travel accommodation. Some clusters showed more of a variety of activity choices while others had seemingly singular interests. The purpose of describing the clusters is to demonstrate that even with so little information, a lot can be gathered about a group simply based on how they spend their time. There is a major opportunity here for future study to tie these devices back to their home regions and back to whatever local marketing segmentation system exists there. Comparative analysis could then be performed between the characteristics of the home versus tourist segment the device belongs to, as well as investigate the diversity of home segments within a tourist segment. Are tourist segments less varied than home segments? Can everyone fit into one of a few tourism segments? The answers to these questions would unveil a wealth of knowledge and insight into consumer behaviour at home and on holiday – a time that is necessarily a luxury and typically a time when people have increased spending habits.

The density visualisation in Figures 4 and 5 highlights the heterogeneity of clusters 1 and 6 in the wide and less concentrated dispersion of accommodation locations across the island. Inspection of the geofences present in the high-concentration accommodation areas of each cluster reveals that the spatial distribution correlates with the accommodation types of each cluster. The most concentrated areas have a higher proportion of the main accommodation type of the cluster. It is unsurprising that that two western beaches are most commonly highlighted across all seven maps as the geofences in these areas reveal a high concentration of accommodation locations used by tourists available to book or rent. This visualisation raises questions that are outside the scope of this study, but that present opportunities for future research. For instance, are devices in these clusters staying at the same accommodation or same few accommodations? If so, this may suggest a flaw in geofencing, clustering, or may simply indicate the popularity of certain locations. It may also point to co-travelling devices that are

being clustered due to the inherent similarity in their travel pattern. Identifying co-travellers could be further incorporated into a similar study to provide more details on the patterns of travel behaviour.

5.2 Relationship between point of interest choices and summary variables

In this study, the patterns in POI choices were quantified by segmenting the tourist devices. Some of the clusters seemed to exhibit more extreme differences in activity choice than others, as well as difference in distance from accommodation, accommodation type, and/or tourists' vacation lengths. The two clusters that varied most in terms of activity type, Clusters 1 and 6, also had longer vacation lengths than other clusters. This supports the theory proposed by Chen et al. (2018), that groups of tourists on vacation for longer than 7 days tend to have greater activity variety due to having more time to try different things. The inverse of this theory was supported by Cluster 2, which had the shortest vacation length and very little variety in their activity types. However, Cluster 7 had the longest vacation time, and yet was a highly homogenous cluster with an unvaried activity participation. In the case of Cluster 7 this may be more impacted by the socioeconomic status of the group having the funds to vacation for longer and participate in private clubs, golfing, etc., than it is impacted solely by the length of the vacation.

Accommodation types were quite segmented between the clusters, and the clusters that stayed exclusively in one accommodation type also seemed to spend most of their time there. These clusters, Clusters 3, 4 and 5, were all staying in accommodations that are associated with higher cost and more on-site amenities, activities, and entertainment – hotels, resorts, and private houses, respectively. It may be that those who are paying more for amenities at their accommodation plan to get the most out of the cost by utilizing what is available there as much as possible. The clusters that had the lower cost, lower amenity accommodations also seemed to have more variety in their activity choice. There may be a relationship here that those that are interested in more variety and adventurousness opt for accommodation types that are more similar to what they have at home – an apartment, a house in a residential area, or something of the like. It would seem as though this variable had the most effect on the cluster population's POI choices.

While it is unclear if the distance from accommodation to attraction destination has any meaningful trends between the clusters, it did highlight what was common across all versus what differed between groups. All clusters demonstrated a willingness to travel further for adventure

activities than for any other activity type. The other activity types did truly vary across the clusters, indicating that there may be more there to analyze with regards to travel interests. A willingness to travel far for one activity type, especially in a group unwilling to travel far in general, might have significant implications for their travel motivations. It may also be best to use the averages of these data per cluster. While the accommodation to attraction type distance metrics are not immediately illuminating, the differences in average distance range per cluster is quite significant with regards to a group's willingness to travel for any type of activity, further to Chen et al. (2018). This type of data and analysis provides significant opportunities for marketing professionals, tourism management bodies, as well as other decision-making bodies to create pseudo customer loyalty data for businesses or activities that are not structured to collect loyalty data in any other manner. Using this type of data analysis in conjunction with survey research could provide insight into how much can be inferred from these basic calculations using mobile data. Further to that, does it matter what a survey subject says about their activity, if what is actually observed (with mobile data) tells a different story?

5.3 Future research opportunities

This study, along with the works cited in the literature review, are only the beginning of the possibilities of research in this field. In a short study with relatively crude information, a clustering system was created that is capable of providing insight into the travel patterns of this study population. If this type of analysis were to be tied in to an analysis of a home segmentation system, the opportunities for further understanding and insight into the variations of consumer behaviour and economic impact would be endless. Further research opportunities are also present with regards to Quan and Wang (2004) and Mkono, Markwell, and Wilson (2013). Where previously this type of tourism research was limited to online reviews, there is an opportunity with these data and this type of methodology to truly compare habitual home behaviour to travel behaviour. There is a major research opportunity in investigating the extent to which people divert from their regular routines when they are on vacation away from home, and it may be that the variation is truly only present depending on the tourist cluster to which a person belongs.

The methodology used in this study provides a framework that could be used to focus in on several different types of results: identifying popular POIs, identifying the relationship between

time of day and POI type popularity, using demographic info from home states to further interpret the cluster groups, and more. This approach could be the groundwork for tying together “sibling” activities, or pairs of activities that are typically engaged in by a certain group, POI use frequency, investigating POI diversity within an attraction type (i.e. how many different POIs of the same type did the device visit, or was it just one location visited multiple times), and more.

5.4 Limitations

Geofencing is a manual and subjective practice, it raises many potential issues with inconsistency and irreproducibility. It is a time consuming and laborious task that may make this methodology inaccessible to others. It would not be feasible for a study at a larger spatial scale. Manual geofencing was only possible in this case because of the relatively small size of the island of Maui, as well as the limited presence of high-rise and multi-use buildings. There was a further limitation to the geofence creation in that it was based on the Google Maps data that were available for the location being investigated. The satellite imagery was current, but was being used to analyze data from 2018 and 2017. The use of Google Street View as a means of ground-truthing business locations also produces an unreliable geofencing atmosphere as some of the imagery can be as old as 2013, which in some cases could result in ground-truthing that is completely incorrect. The success of the clustering method in this paper was highly affected by the classification of the geofences. This likely affected the activity type participation statistics. For example, gastronomic POIs was the largest category containing the most geofences. This category was also had participation levels in similar degrees across all the segmentation groups. While most people may enjoy going out to eat regardless of their other travel interests, it may be that the number of possible geofences also influenced this. POIs were aggregated into type categories that were difficult to differentiate. The differences between the categories in some cases were marginal. What is the difference between Athletic Club attractions and Sports & Fitness attractions? These point of interest types were separated to create a distinction between private, membership-based locations typically associated with a more luxurious experience and regular community recreation centres and gyms. However, gyms also require a membership and depending on the location, may be considered to be a luxurious experience. They were separated because there is a sense of difference between these two experiences, but this difference is subjective. There are several similar cases

within the geofence classification, making the categories challenging to replicate for another researcher.

The clustering required the inclusion of time spent during the day at each device's own accommodation in order to fully model the division of time between activity types and segment the devices into somewhat homogenous groups. This complicates the results because the accommodation types were included in the clustering, whereas a general "time at accommodation" variable rather than using specific accommodation type may have produced results that contain fewer potential confounding factors. This would present its own limitations in that, for instance, spending time at a resort during the day speaks to a very different experience than spending time at an apartment rental or camp site during the day. These experiences would not be a part of separating the clusters if "time spent at accommodation" was a single agglomerated variable.

There were many limitations in the data used. The inconsistency and incompleteness of the dataset led to a significant reduction of sample population. It may be that the filtering thresholds used in this methodology were stricter than necessary. Helpful results could be yielded at a much larger scale by slightly reducing the rigidity of the filter parameters. Incorporating multiple data streams from different data providers would also help to mitigate the issue of incomplete data and small sample size. These data streams would likely be gathered from different apps because they would be from different data providers. This could help to overcome bias in the data due to user demographics of certain apps by broadening the range of app types. It is also possible to deduplicate the repeated devices across datasets, allowing for an increasingly granular database to be created.

A major limitation to the model created in this study is its strength in practical applications. This model may be a helpful model to smaller, more isolated communities such as Maui to give some basic insight into types of tourists visiting the island. However it seems that even in this case it would be best applied when seeking specific information about visitors, rather than to create a static segmentation system. Clustering to identify similar types of tourists to better understand the relationship between activity choices may be useful, such as in the case of "sibling" activities previously mentioned. This model would not be applicable in areas with denser urban development, high rise buildings, and/or a high proportion of multi-use buildings. In that case, points of interest may not be possible to identify with any certainty, and the creation of regions of interest would be more appropriate.

5.5 Conclusion

Marketing professionals have been segmenting consumer populations for decades, and the tourism industry has been using surveys and case studies to determine travel behaviours and POI use patterns. More recently, mobile data have been used in an attempt to discover further insight into how POIs are used by tourists, but it has never been used to create a tourist segmentation system based on POI use. This work has demonstrated the ability of mobile data to provide granular information about tourist experiences and create a model of tourist behaviour that could be replicated in other study areas for use by marketing professionals and by tourism management and planning bodies. A segmentation system was created based on tourists' activity choices, and vacation length, distance to attraction from accommodation, and accommodation type were all summarized and compared across the clusters.

The study found that there is significant segmentation of accommodation type across the clusters and that activity diversity may be tied to the length of visitors' stays. Distance from accommodation to attraction was not easily interpretable as significant at the attraction type level but did illuminate some distinct differences between the groups when summarized at the cluster level. This work provides a new use case of K-Means and mobile location data for the purpose of creating pseudo-loyalty data and consumer segments for POIs that could not otherwise have data collected. This approach creates an opportunity for marketing professionals to capture the nuances in travel behaviour.

References

- Arrowsmith, C., and W. Ntuwah. 2002. Mapping Potential for Tourist Segments. *Cartography* 31 (1):55–61. <http://www.tandfonline.com/doi/abs/10.1080/00690805.2002.9714180>.
- Asan, K., and M. Emeksiz. 2016. Outdoor Recreation Participants' Motivations, Experiences and Vacation Activity Preferences. *Journal of Vacation Marketing* 24 (1):3–15.
- Balmford, A., J. Beresford, J. Green, R. Naidoo, M. Walpole, and A. Manica. 2009. A Global Perspective on Trends in Nature-Based Tourism ed. W. V. Reid. *PLoS Biology* 7 (6):e1000144. <http://dx.plos.org/10.1371/journal.pbio.1000144>.
- Bengtsson, L., X. Lu, A. Thorson, R. Garfield, and J. von Schreeb. 2011. Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti ed. P. W. Gething. *PLoS Medicine* 8 (8):e1001083. <https://dx.plos.org/10.1371/journal.pmed.1001083>.
- Chao, H., Y. Cao, J. Zhang, F. Xia, Y. Zhou, and H. Shan. 2018. Population Density-Based Hospital Recommendation with Mobile LBS Big Data. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 37–44. Shanghai: IEEE <https://ieeexplore.ieee.org/document/8367095/>.
- Chen, C.-Y., and K. Grauman. 2011. Clues from the beaten path: Location estimation with bursty sequences of tourist photos. In CVPR 2011, 1569–1576. Colorado Springs, CO, USA: IEEE <http://ieeexplore.ieee.org/document/5995412/>.
- Chen, Q., Z. Hu, H. Su, X. Tang, and K. Yu. 2018. Understanding Travel Patterns of Tourists from Mobile Phone Data: A Case Study in Hainan. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), 45–51. Shanghai: IEEE <https://ieeexplore.ieee.org/document/8367096/>.
- Cheng, Z., L. Jiang, D. Liu, and Z. Zheng. 2018. Density based spatio-temporal trajectory clustering algorithm. In IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 3358–3361. Valencia: IEEE <https://ieeexplore.ieee.org/document/8517434/>.
- Collins, A. T., J. M. Rose, and S. Hess. 2012. Interactive stated choice surveys: a study of air travel behaviour. *Transportation* 39 (1):55–79. <http://link.springer.com/10.1007/s11116-011-9327-z>.
- Cuzzocrea, A. 2016. Algorithms for managing, querying and processing big data. In *Algorithms* 9 (13). Trieste, Italy: <http://www.mdpi.com/1999-4893/9/1/13>.
- Fei Hu, Qingquan Sun, and Qi Hao. 2010. Mobile targets region-of-interest via

- distributed pyroelectric sensor network: Towards a robust, real-time context reasoning. In 2010 IEEE Sensors, 1832–1836. Kona, HI: IEEE
<http://ieeexplore.ieee.org/document/5690006/>.
- Gudmundsson, J., P. Laube, and T. Wolle. 2006. Movement Patterns in Spatio-Temporal Data. 1–10.
- Guo, Y., J. Zhang, and Y. Zhang. 2016. An Algorithm for Analyzing the City Residents' Activity Information through Mobile Big Data Mining. In 2016 IEEE Trustcom/BigDataSE/ISPA, 2133–2138. Tianjin, China: IEEE
<http://ieeexplore.ieee.org/document/7847211/>.
- Gursoy, D., C. Jurowski, and M. Uysal. 2002. Resident attitudes. *Annals of Tourism Research* 29 (1):79–105.
<https://linkinghub.elsevier.com/retrieve/pii/S0160738301000287>.
- Hasnat, M. M., and S. Hasan. 2018. Identifying tourists and analyzing spatial patterns of their destinations from location-based social media data. *Transportation Research Part C: Emerging Technologies* 96:38–54.
<https://linkinghub.elsevier.com/retrieve/pii/S0968090X18300433>.
- Hu, T., X. Zhu, K. Su, and L. Duan. 2011. Collection and remote management of disaster information with mobile positioning terminals. In 2011 19th International Conference on Geoinformatics, 1–5. Shanghai, China: IEEE
<http://ieeexplore.ieee.org/document/5980970/>.
- Kawakubo, H., and K. Yanai. 2009. An analysis of the relation between visual concepts and geo-locations using geotagged images on the web. In 2009 IEEE International Conference on Multimedia and Expo, 1644–1647. New York, NY, USA: IEEE
<http://ieeexplore.ieee.org/document/5202835/>.
- Keogh, B. 1984. The measurement of spatial variations in tourist activity. *Annals Of Tourism Research* 11:267–282.
- Kirilenko, A. P., S. O. Stepchenkova, and J. M. Hernandez. 2019. Comparative clustering of destination attractions for different origin markets with network and spatial analyses of online reviews. *Tourism Management* 72:400–410.
<https://linkinghub.elsevier.com/retrieve/pii/S0261517719300019>.
- Kivela, J. J., and J. C. Crotts. 2009. Understanding Travelers' Experiences of Gastronomy Through Etymology and Narration. *Journal of Hospitality & Tourism Research* 33 (2):161–192. <http://journals.sagepub.com/doi/10.1177/1096348008329868>.
- Latzko, D. A. 2005. Economic conditions and Japanese tourism to Hawai'i. *Asia Pacific*

- Journal of Tourism Research 10 (2):151–156.
<http://www.tandfonline.com/doi/abs/10.1080/10941660500135936>.
- Li, J., X. Qian, Y. Y. Tang, L. Yang, and T. Mei. 2013. GPS Estimation for Places of Interest From Social Users' Uploaded Photos. *IEEE Transactions on Multimedia* 15 (8):2058–2071. <http://ieeexplore.ieee.org/document/6588893/>.
- Li, Y., L. Yang, H. Shen, and Z. Wu. 2019. Modeling intra-destination travel behavior of tourists through spatio-temporal analysis. *Journal of Destination Marketing & Management* 11:260–269.
<https://linkinghub.elsevier.com/retrieve/pii/S2212571X17303839>.
- Luo, S., Y. Ng, T. Z. W. Lim, C. C. H. Tan, N. He, G. Manai, and Y. Li. 2018. Improved Localisation Using Spatio-Temporal Data from Cellular Network. In 2018 19th IEEE International Conference on Mobile Data Management (MDM), 56–65. Aalborg, Denmark: IEEE <https://ieeexplore.ieee.org/document/8411262/>.
- Manyika, J., M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers. 2011. Big Data: The next frontier for innovation competition and productivity.
- Mkono, M., K. Markwell, and E. Wilson. 2013. Applying Quan and Wang's structural model of the tourist experience: A Zimbabwean netnography of food tourism. *Tourism Management Perspectives* 5:68–74.
<https://linkinghub.elsevier.com/retrieve/pii/S2211973612000876>.
- Phithakkitnukoon, S., T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, and R. Shibasaki. 2015. Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan. *Pervasive and Mobile Computing* 18:18–39.
<https://linkinghub.elsevier.com/retrieve/pii/S1574119214001321>.
- Qiao, Y., Z. Xing, Z. Md. Fadlullah, J. Yang, and N. Kato. 2018. Characterizing Flow, Application, and User Behavior in Mobile Networks: A Framework for Mobile Big Data. *IEEE Wireless Communications* 25 (1):40–49.
<http://ieeexplore.ieee.org/document/8304390/>.
- Quan, S., and N. Wang. 2004. Towards a structural model of the tourist experience: an illustration from food experiences in tourism. *Tourism Management* 25 (3):297–305.
<https://linkinghub.elsevier.com/retrieve/pii/S0261517703001304>.
- Rinne, P., and O. Saastamoinen. 2005. Local Economic Role of Nature-based Tourism in Kuhmo Municipality, Eastern Finland. *Scandinavian Journal of Hospitality and Tourism* 5 (2):89–101. <http://www.tandfonline.com/doi/abs/10.1080/15022250510014363>.
- Saito, O. 2013. Resource Use and Waste Generation by the Tourism Industry on the Big

- Island of Hawai'i: Resource Use and Waste on Hawai'i. *Journal of Industrial Ecology* 17 (4):578–589. <http://doi.wiley.com/10.1111/jiec.12007>.
- Shen, J., and T. Cheng. 2016. A framework for identifying activity groups from individual space-time profiles. *International Journal of Geographical Information Science* 30 (9):1785–1805. <http://www.tandfonline.com/doi/full/10.1080/13658816.2016.1139119>.
- Shiaty, R. E., M. Taalab, and I. Osama. 2016. Evaluating the Performance of the Outdoor Spaces in Healing Eco-Tourism. *Procedia Environmental Sciences* 34:461–473. <https://linkinghub.elsevier.com/retrieve/pii/S1878029616300639>.
- Urata, J., Y. Sasaki, and T. Iryo. 2018. Spatio-Temporal Analysis for Understanding the Traffic Demand After the 2016 Kumamoto Earthquake Using Mobile Usage Data. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2496–2503. Maui, HI: IEEE <https://ieeexplore.ieee.org/document/8569411/>.
- Xu, G., S. Gao, M. Daneshmand, C. Wang, and Y. Liu. 2017. A Survey for Mobility Big Data Analytics for Geolocation Prediction. *IEEE Wireless Communications* 24 (1):111–119. <http://ieeexplore.ieee.org/document/7731599/>.
- YuFang Dan, and Zhongshi He. 2010. A dynamic model for urban population density estimation using mobile phone location data. In *2010 5th IEEE Conference on Industrial Electronics and Applications*, 1429–1433. Taichung, Taiwan: IEEE <http://ieeexplore.ieee.org/document/5514844/>.
- Zhao, M., M. Ma, and Y. Yang. 2011. Efficient Data Gathering with Mobile Collectors and Space-Division Multiple Access Technique in Wireless Sensor Networks. *IEEE Transactions on Computers* 60 (3):400–417. <http://ieeexplore.ieee.org/document/5487503/>.
- Zhao, X., X. Lu, Y. Liu, J. Lin, and J. An. 2018. Tourist movement patterns understanding from the perspective of travel party size using mobile tracking data: A case study of Xi'an, China. *Tourism Management* 69:368–383. <https://linkinghub.elsevier.com/retrieve/pii/S0261517718301377>.
- Zhou, D., J. F. Yanagida, U. Chakravorty, and P. Leung. 1997. Estimating economic impacts from tourism. *Annals of Tourism Research* 24 (1):76–89.