

Article

Identification of Mobility Patterns of Clusters of City Visitors: An Application of Artificial Intelligence Techniques to Social Media Data

Jonathan Ayebakuro Orama ^{1,*} , Assumpció Huertas ² , Joan Borràs ¹ , Antonio Moreno ³ 
and Salvador Anton Clavé ^{1,4} 

- ¹ Eurecat, Centre Tecnològic de Catalunya, C/Joanot Martorell, 15, 43480 Vila-Seca, Spain; joan.borras@eurecat.org (J.B.); salvador.anton@urv.cat (S.A.C.)
² Department of Communication, Universitat Rovira i Virgili, Av. Catalunya, 35, 43002 Tarragona, Spain; sunsi.huertas@urv.cat
³ Intelligent Technologies for Advanced Knowledge Acquisition (ITAKA) Research Group, Escola Tècnica Superior d'Enginyeria, Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans, 26, 43007 Tarragona, Spain; antonio.moreno@urv.cat
⁴ Department of Geography, Universitat Rovira i Virgili, C/Joanot Martorell, 15, 43480 Vila-Seca, Catalonia, Spain
* Correspondence: jonathan.orama@eurecat.org; Tel.: +34-674-885-859

Abstract: In order to enhance tourists' experiences, Destination Management Organizations need to know who their tourists are, their travel preferences, and their flows around the destination. The study develops a methodology that, through the application of Artificial Intelligence techniques to social media data, creates clusters of tourists according to their mobility and visiting preferences at the destination. The applied method improves the knowledge about the different mobility patterns of tourists (the most visited points and the main flows between them within a destination) depending on who they are and what their preferences are. Clustering tourists by their travel mobility permits uncovering much more information about them and their preferences than previous studies. This knowledge will allow DMOs and tourism service providers to offer personalized services and information, to attract specific types of tourists to certain points of interest, to create new routes, or to enhance public transport services.

Keywords: mobility patterns; social media data; artificial intelligence; tourist clusters; tourist flows



Citation: Orama, J.A.; Huertas, A.; Borràs, J.; Moreno, A.; Anton Clavé, S. Identification of Mobility Patterns of Clusters of City Visitors: An Application of Artificial Intelligence Techniques to Social Media Data. *Appl. Sci.* **2022**, *12*, 5834. <https://doi.org/10.3390/app12125834>

Academic Editors: Dionisis Margaritis and Stefanos Ougiaroglou

Received: 6 May 2022

Accepted: 6 June 2022

Published: 8 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Technological evolution has brought changes in tourists and their behavior [1–3], and it has catalyzed new information challenges for destinations. In this evolution, some destinations have started to become smart by integrating technological infrastructures and end-user devices [4–6]. However, the fulfillment of two of the main objectives of smart destinations, namely the enhancement of tourists' experiences and the improvement of their management, is still far from complete, and results from the destinations' efforts remain largely unreported [7]. Progress in this area requires Destination Management Organizations (DMOs) to know who their tourists are, what needs and travel preferences they have, what they visit the most, and which are their mobility patterns and flows around the destination [8].

Technological evolution has allowed some Destination Management Organizations to start to maintain personalized and real-time exchange of information with tourists [9] and, at the same time, to collect vast amounts of information from them (big data). The analysis of these data permits offering them even more personalized services [10,11] at the moment and place they need it [12,13]. This two-way communication between tourists and

DMOs helps to generate more satisfactory tourist experiences [5,14,15] and to improve the destination's management [16].

Many studies have focused on knowing who are the visitors of a destination [17]. These studies have classified tourists in clusters or profiles depending on their travel preferences and behaviors [18,19]. Some studies have analyzed the acceptance of technology by tourists [20], while others have focused on their degree of connectivity during the trip [21,22]. To find out their preferences and behaviors, some authors have analyzed their information searches during the trip [23–25] and others their movements at the destination (flow analysis) through GPS or other smart technologies [26,27]. Knowing the mobility patterns of tourists is one of the most important issues for the development of tourism planning and destination management [28,29].

Social media has brought a great transformation in the related tourism research [8,30,31]. Social media analytics use natural language processing and machine learning techniques to analyze social media content [32]. Geotagged Social Media Data (GSMD) offer information about tourist behavior, travel route choices, emotions, and satisfaction level [33]. Nowadays, technology and analytical tools are evolving and social media analytics allow us to know the movements or flows of tourists [34]. Twitter is a platform with global coverage, very useful for tourist mobility analysis [35] that even allows us to determine the most visited POIs (Points of Interest) at destinations [36].

Existing research in spatial data or mobility/flow analysis through GSMD still offers little information about mobility patterns of different sub-groups or clusters of tourists [37]. Nevertheless, the more DMOs know about the mobility, preferences and behavior of tourists, the more they will be able to offer personalized services, packages and information [11,38]. Additionally, the more they can automatize processes to obtain information from open social media data, the more efficient they will be when providing personalized information throughout travel recommender systems [39].

Therefore, the aim of the study is to develop a methodology that, through the application of Artificial Intelligence techniques to geotagged social media data, creates clusters of tourists according to their mobility patterns and visiting preferences at the destination. The analytical and managerial goal is to know the most visited points of interest and the main flows between them within a destination for each tourist cluster. This will allow the DMOs to uncover their tourists' profiles, their preferences, and their mobility patterns at the destination. This will also let them enhance the visitors experience through the personalization of tourist packages, services, public transport and also the information offered on each attraction or spot. Finally, this will also enhance the usefulness and efficiency of DMOs to mitigate the array of problems caused by the mobility of visitors towards and around the main points of interest through a better experience design management and interaction with the environment [40].

2. Theoretical Framework

2.1. Tourist Flow Analysis and Social Media Analytics

Prior to the technological developments, studies on tourist mobility were based on tourists' surveys [35], and they were rather limited. Tourist flow analysis grew enormously with the development of tracking mechanisms like GPS, cell-tower identification or Wi-Fi positioning [41], which have made it possible to obtain big data from the movement of tourists at destinations. Several studies have used GPS [27,42,43] to find out which were the most visited places in a particular destination and when they were visited. In destinations, the proliferation of sensor networks and portable devices like smartphones has also made it possible to obtain big data from tourists and to know their movements or flows [26]. In general, the increasing effectiveness and reliability of GPS data and the mobile positioning data have increased the possibilities of analyzing spatial-temporal behaviors, widening the research objectives beyond the initial aim to know where and when visitors went. In this vein, they have been used to identify seasonal demand patterns by Ahas et al. [44] or to improve the management of destination marketing by Kuusik et al. [45]. Other authors

have considered data obtained from mobility services, such as the subway smart card [46] or bike sharing systems [47].

Social media also has very useful platforms for knowing the movements of tourists. Social media analytics (SMA) is a research field that has advanced heavily since 2014, but it is still in an early stage of development [32]. The potential of social media as sources for big data research in the field of tourism has increased in the last decade [8,30], and studies on big data, social media, UGC (User-Generated Content), and online reviews have proliferated in hospitality and tourism [30,32,34,48,49]. Moreover, the innumerable footprints that millions of tourists leave online using the technological platforms constitute an interesting source for knowing the tourists' movements and flows [49,50], although big data-based theoretical studies still remain limited [34].

Text analytics and data mining studies try to find out tourists' interests and to predict their decisions and behaviors [51]. Trend analysis studies also try to predict tourists' behaviors [52] or future trends in tourist behavior at destinations [53]. Nevertheless, one of the best ways to know the behavior of tourists during the trip is through spatial data analysis.

Spatial data analysis is a stream of studies within SMA that, through the analysis of GSMD [8], aim to ascertain the spatial distribution of tourists on a place or a destination [54] and even to know tourists' movements or flows [26,41,55]. Flow is the collective movement of people [26], and flow analysis shows the movement of tourists in a location [38]. GSMD are key sources of information to analyze tourists flows [56] in order to uncover their travel preferences and behaviors [26] or the tourists' experiences through a spatial analysis [33]. Provenzano et al. [50] compared GSMD results with the UNWTO record-based network demonstrating the usefulness of GSMD to discover tourist flows. However, there are still few publications that analyze Geotagged Big Data to examine the spatial distribution and flows of different types of tourists in the destinations [35].

GSMD analysis allows for knowing the dispersion of tourists and also the routes and activities they carry out in the destination [57–62], their density of movements [63], their flows [26,38,64,65], and the most popular resorts, attractions, or points of interest in the destinations [36,62,66,67].

Most studies that analyze GSMD have focused on Twitter because it is a platform that has global coverage and its data are available for free on the Internet from the moment the tweets are published [35]. Twitter is, along with Tripadvisor, one of the two most used platforms by SMA. It also allows the analysis of multi-modal data such as User Generated Content (including text, images, and even videos), and geotagged information [32,68]. However, studies based on geotagged photos and other social media like Flickr [38,69], Foursquare [60], or Instagram [70] have also proliferated showing tourists' flows, movements, and behaviors in destinations.

It has also been shown that analyzing different social media sources or platforms is useful because they provide complementary information and enrich the knowledge of different tourists' movements [35]. In this line, Dietz et al. [71] analyzed tourists' movements at destinations through three social media (Twitter, Foursquare, and Flickr) and identified different types of trips according to the origin of the tourists. A study by Sugimoto et al. [72] combined tracking technologies and surveys to study the relationship between visitor mobility and urban spatial structures. Salas-Olmedo et al. [35] analyzed the digital footprint of urban tourists through photos, check-ins, and tweets from three social media (Panoramio, Foursquare, and Twitter). In addition, they used a clustering methodology to identify certain areas of the destinations according to the tourist activities that visitors carried out in them. However, they did not cluster or segment tourists to know their different preferences, movements, and behaviors.

Many studies on GSMD have focused on tourists' mobility patterns [59,73–75]. However, the difference in mobility patterns between sub-groups or clusters of tourists has not been fully researched [37].

2.2. Uncovering the Mobility Patterns of Clusters of Tourists

Previous studies have shown that different types, sub-groups, or clusters of tourists may present different travel behaviors [76–79]. Domènech et al. [80], for instance, identified that cruise passengers with different expenditure levels have different mobility patterns in port destination cities. However, this kind of studies usually apply an ad-hoc combination of analytic techniques that is not easy to generalize.

From a complementary perspective, many researchers have followed the digital footprint of tourists [35] to know their mobility in destinations, but the current research shows that it is difficult to analyze all these data by segmenting tourists. In fact, many studies have analyzed tourists' mobility patterns [73,74] without taking into account the diversity of tourists [37] because of the difficulty of obtaining their socio-demographic data. This aspect can be considered only in those cases in which user information is available, such as the one described by Massimo and Ricci [81]. In that case, they use information produced by a recommender system to define patterns of visitors depending on the similarities in their observed visit trajectories.

Previous GSMD-based studies have focused on the clustering of tourists according to diverse factors. Following Liu et al. [37], studies that segment visitors according to their mobility patterns can be based on non-spatial factors (socio-economic status, gender, age, income, education, race) or on spatial factors. For instance, Manca et al. [82] focused on spatial data and Jin et al. [41] on spatial and temporal data. Nevertheless, very few studies have focused on the analysis of mobility patterns according to the socio-demographic data in order to segment visitors in a destination because, with the currently applied methods of analysis, very little socio-demographic data can be obtained from users. In this vein, several SMA studies based on GSMD analysis have claimed to obtain demographic data from tourists to better understand who they are and to be able to classify them [26,69,83]. However, the available information is still very limited [60] and, in some cases, it is even reduced to the country of origin [56].

To name some of those studies, Chua et al. [26] focused on spatial, temporal, and also demographic data from Twitter to discover the tourists flows in a destination, creating tourist profiles and segmenting them by country of origin. Similarly, Vu et al. [77] analyzed the different mobility patterns, popular locations, and routes in Hong Kong of Western and Asian tourists. In the same line, Paldino et al. [84] analyzed geo-tagged picture data from Flickr, segmenting domestic and foreign tourists, and Ma et al. [18] also analyzed the mobility of tourists in destinations and their most visited attractions by classifying tourists into foreign tourists and domestic tourists. Van der Zee and Bertocchi [85] analyzed the spatial behavior of visitors at a destination through a relational approach and Trip Advisor data, classifying visitors as local, national, European, and non-European. Vu et al. [86] analyzed the activities carried out in a destination by different groups of tourists also segmenting them by their country of origin. Xu et al. [87] analyzed mobility patterns of tourists in a country, South Korea, and its diverse destinations, according to their nationality or country of origin. Liu et al. [37] analyzed mobility via GSMD from Twitter by considering homogeneous segments of users (state visitors, national visitors, and international visitors), created according to their past visits.

However, despite the difficulty of obtaining other socio-demographic data than the origin of users from the GSMD, some mobility studies have tried to take a step further in segmenting visitors. Huang and Wong [88], for example, analyzed their mobility segmenting them by their socio-economic status through Twitter's GSMD. They identified this status from the home and work location of the users. In addition, they showed that socio-economic status and urban spatial structure are the factors that have a stronger influence on the mobility of visitors. On the other hand, Han et al. [89] analyzed the mobility patterns of visitors from the analysis of social media check-in. They used a deep learning method to try to classify tourists by the purpose of their travel.

Studies of mobility patterns that employ Artificial Intelligence techniques are still emerging, and very few of them try to identify meaningful clusters of tourists. Liao [90]

obtained trajectory data from different location-based services and tourism applications, and then they applied cluster analysis to identify the most popular tourist attractions. DBSCAN clustering was used to identify spatial clusters of trajectories at the points of greatest interest, but not to identify clusters of tourists. Xu et al. [87] analyzed a mobile positioning data set in order to know the nationality and movement patterns of foreign tourists in South Korea. They used network analysis to identify the structure of tourism destinations based on patterns of travel flow, and clustering analysis to identify similar patterns. They identified areas of destinations with different visit patterns of tourists according to their nationality, but not clusters of tourists. Instead, Giglio et al. [91] used cluster analysis to identify automatically clusters of tourists around points of interest at destinations. They studied the relationship between human mobility and tourist attractions through geo-located images of Italian destinations provided by Flickr users. The results showed that social media data are a valuable source to understand the behavior of tourists in a destination. However, the study did not define or specify the different clusters of tourists and their different mobility flows between the most popular attractions.

Considering that, despite the difficulties and limitations, GSMD data can be analyzed in order to make segmentations more precise than the ones based on the country of origin, and also considering that mobility patterns are a key issue for DMOs; this study aims to make a contribution to the current challenges of analyzing tourism flows in destinations. Hence, its goal is to provide and test a methodology to create tourists' clusters (groups of visitors with similar characteristics, preferences, and patterns of travel) taking into account not only their personal and cultural interests, leisure activities, and the context, but also their mobility patterns on the destination. The authors' perception is that, from an academic perspective, this is a fundamental issue in order to better understand tourists spatial and temporal behavior. Additionally, from the point of view of the developers of tourists' experiences, this methodology is a new tool that can help to offer services to tourists in a more personalized way, enhancing the communication to the appropriate targets or the attraction of new ones [37]. Finally, it can help to improve the management of critical flows in certain points of interest of the tourism destinations, helping to minimize social stresses, built environment management difficulties, transportation issues, and frictions between tourists and local population in situations of congestion and overcrowding [40].

3. Methodology

3.1. Data Collection and Processing

Tourist mobility patterns can be analyzed by building user profiles, which group individuals with common visits to some points of interest or with similar travel behavior. These profiles can be discovered via a clustering process, which uncovers the common interests and habits of the visitors. The data required to cluster individuals based on their visited POIs for tourist mobility analysis can be obtained from various sources, including social media and mobile carriers (GPS). Social media data have become popular for this purpose in recent years.

The data used in this research were collected from the popular micro-blogging platform Twitter, using their application interface (API). The API allows developers to stream live tweets published at a specific location, which is provided as a JavaScript Object Notation (JSON) file that includes timestamps, texts, tweet IDs, user details, tweet language, coordinates, and place details. We streamed geo-located tweets published in the city of Barcelona in the year 2019, which amounted to over 1.5 million tweets from more than 100,000 users.

After the data collection, a cleaning process was applied. Users that had sent less than three tweets were discarded, along with their tweets. It was necessary to distinguish residents from visitors, as we were interested only in detecting visitor profiles. A tweet was considered to be from a tourist if Barcelona is not among his/her home locations. These home locations are specified explicitly in the individual's Twitter profile, or they are inferred from his/her tweets. Concretely, a place is considered to be a user's home location

if he/she has posted daily tweets in a 20-day period (i.e., it is assumed that someone residing in a location for more than 20 consecutive days is not a tourist). Afterwards, tweets from residents were discarded.

In order to build the user profiles, it was crucial to identify the points of interest experienced by the tourists from their published tweets. The geographical coordinates of each tweet were sent to Overpass, a query engine for requesting specific features on the Open Street Map (OSM) server, in order to obtain points of interest in their proximity. The POIs returned from Overpass were then assigned to the tweets. OSM classifies physical features (buildings, parks, attractions, etc.) under certain tags that describe their type (e.g., nature: beach). We used these tags to build an activity tree, which has a hierarchical structure that categorizes activities. The activity tree contained 175 OSM tags, classified under nine main categories (Routes, Sports, Gastronomy, Leisure, Accommodation, Transportation, Nature, Events, and Culture) and 32 subcategories.

The activity tree allows us to categorize tweets based on an assigned POI visited in the tweet. This POI is chosen from the POIs returned from Overpass, by giving priority to POIs explicitly mentioned in the tweet and to POIs with OSM tags that belong to categories that are of higher interest to a tourist. For instance, we consider a museum to be more interesting to a tourist than a coffee shop, so the museum is prioritized over the coffee shop. The tweets that could not be assigned to any POI and activity were removed from the analysis. After the filtering process, the dataset had 37,302 tweets from 6066 individuals. The summary of the dataset is shown in Table 1.

Table 1. Dataset summary.

Statistics	Value
Total number of tweets in Barcelona	1,523,801
Total number of users in Barcelona	108,515
Statistics after filtering	
Total number of tweets in Barcelona	37,302
Total number of users in Barcelona	6066

Source: Authors, from a previous work [39].

3.2. Feature Engineering and Data Clustering

In order to cluster the users, each of them was represented by a numerical vector of features, which represents their interests and travel habits. Four types of features were considered:

- Activity interest features (25¹): these features represent different levels in the activity tree, which were chosen to show the categories more represented in the dataset. All tweets published by a user are assigned to activity categories as described in Section 3.1. The ratio of tweets in the highlighted activity interest features (Routes, Sports, Accommodation, Transportation, Nature, Food, Enotourism, AmusementParks, RecreationFacilities, Beach, Health&Care, NightLife, Shopping, Viewpoint, CulturalAmenities, Historic, Religious, Events, tourism_museum, amenity_arts_center, tourism_gallery, artwork_type_sculpture, artwork_type_architecture, artwork_type_statue, and other_artwork.) to the total number of tweets published by the user is taken to represent the user's degree of interest in those activities. They are in the range [0, 1].
- Travel features (3): these features give an idea of the degree of mobility of the user inside the city. They encode the length of the stay of the user in Barcelona (maximum consecutive days that the user sent tweets from this city), and the maximum and average distances between the location of published tweets. These values are computed by counting the days the user published tweets in Barcelona, eliminating gaps to find the maximum stretch of days with published tweets. They are normalized in the range [0, 1].

- Popularity features (5): these features show if the user is interested in the most popular places or if he/she prefers to visit places off the beaten track. They represent the percentage of tweets sent from the user from the top 10 most visited locations in Barcelona, the top 10–20, the top 20–50, the top 50–100, or from other POIs. These top visited locations are ranked by number of user visits in the dataset, and split into bins 1–10, 20–50, 50–100, and the rest (each bin represents a feature in this category). Then, the ratio of a user's tweets in each bin to his/her total number of tweets is used to represent the user in these features. They are in the range [0, 1].
- Temporal features (4): these features give an idea on the time of the day in which the user prefers to tour. They represent the percentages of tweets sent from the user at dawn (12:00 a.m.–7:00 a.m.), morning (7:00 a.m.–12:00 p.m.), afternoon (12:00 p.m.–8:00 p.m.), or night (8:00 p.m.–12:00 a.m.). As in the popularity features, bins are created to represent different time periods in the day (each bin represents a feature in this category). Then, the ratio of a user's tweets in each bin to his/her total number of tweets is used to represent the user in these features. They are in the range [0, 1].

Thus, each of the 6066 users was represented by a vector of 37 numbers, which codifies his/her leisure interests, moving ability, interest in popular places, and the preferred period of the day for visiting points of interest. All features were standardized using the *Z-score* scaler, and the k-means clustering algorithm was used to cluster this dataset into 25 clusters (this number was empirically decided after some experiments). Five of the 25 clusters had less than 50 users, so they were not considered to be very relevant, and they were dismissed from the posterior mobility analysis. Thus, at the end of the clustering process, there were 20 clusters with a minimum of 50 visitors. By averaging the values of the 37 features for the users in each cluster, we obtained a general description of the preferences of the users in that group.

In a previous work [39], you can find a more detailed technical explanation of the data collection, filtering, activity identification, and clustering steps.

4. Results

4.1. Characterization of the Visitors in Each Cluster

The analysis identified different clusters with a wide range of visit preferences. Figure 1 highlights the main characteristics shared by tourists in each of the 20 clusters, including their origin, the kind of leisure activities they visit, their preferred time of the day, and their interest in different kinds of activities. All clusters are associated with one or more types of activities, which match the activity features used in the clustering process.

As it can be seen in Figure 1, the historic and religious activities (columns A and B) are the ones associated with most clusters (0, 2, 5, 8, 11, 16, 17, and 19). In most of these clusters, users enjoy visiting the most popular POIs, revealing that the historic and religious attractions are among the most popular in the city of Barcelona. This also proves a direct correlation between tourist interests and their popularity, as it could be expected.

It can also be seen that the individuals in clusters 2 and 17 visit a larger number of popular POIs and have a wider spread of different kinds of activities. These clusters contain mostly non-Spanish tourists, who visit and experience many varieties of POIs that the city offers. On the other hand, clusters characterized mainly by Spanish nationals (clusters 1, 3, 9, 10, 15, 21, 23, and 24) focus on a small set of features and on unpopular POIs, indicating a higher specificity in their favorite POIs and their visits. People in cluster 1, for example, prefer to do scenic routes within the city. However, those in cluster 3 prefer to go to the beach or places near the beach, and some historic sites. Tourists on cluster 9 visit Barcelona mainly for health and care, although they also visit religious sites and cafes, whereas those in cluster 15 prefer food and enotourism, those in cluster 23 are mainly interested in visiting Camp Nou, the football stadium, and people in cluster 24 mainly visit museums.

Cluster	Origin					Activity Interests																	Temporal Interests				Popularity Interests							
	ID	USA	Spain	UK	Italy	France	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	Dawn	Morning	Afternoon	Night	Popular	Unpopular
0																																		
1																																		
2																																		
3																																		
5																																		
6																																		
7																																		
8																																		
9																																		
10																																		
11																																		
12																																		
13																																		
15																																		
16																																		
17																																		
19																																		
21																																		
23																																		
24																																		

Figure 1. Cluster characteristics highlighting tourists' origin, preferred tour time, orientation towards popular POIs, and interests in different activities. * Note: A—Religious, B—Historic, C—Routes, D—Nature, E—Art Gallery, F—Recreational Facilities, G—Beach, H—Cultural Amenities, I—Shopping, J—Nightlife, K—Statues, L—Health & Care, M—Viewpoints, N—Sculptures, O—Accommodation, P—Food, Q—Enotourism, R—Amusement Parks, S—Museums, T—Architecture, U—Other Artworks, V—Art Centers.

The mobility analysis based on visitor clusters that was carried out also allows for knowing the specific POIs visited by the tourists of each cluster. Figure 2 shows the percentage of tourists from each cluster who visit the top 20 most visited POIs of the destination, highlighting in each row the values above the average. For example, Camp Nou, the Barcelona football stadium, is visited mostly by tourists from cluster 23, the religious architecture (as the Sagrada Família) and historic monuments (as the Palau de la Generalitat or the Casa Batlló) are mainly visited by tourists of clusters 0, 2, or 17, and beaches (like Barceloneta) or places near them by tourists of cluster 3. This figure also clearly identifies those clusters that are focused on the most popular POIs, and to what degree. Clusters 0 and 2 are heavily focused on popular POIs as they have an above average representation of 55% and 80% (respectively) in the popular POIs. In contrast, clusters 7, 12, 13, 15, 21, and 23 have very little representation in the popular POIs. This shows that the 20 clusters do not favor only popular or unpopular POIs.

Attractions	0	1	2	3	5	6	7	8	9	10	11	12	13	15	16	17	19	21	23	24
0 Basílica de la Sagrada Família	25.04%	2.26%	11.48%	4.74%	7.61%	3.57%	1.42%	5.67%	4.02%	3.89%	6.63%	3.45%	1.43%	2.27%	5.08%	7.77%	5.20%	0.67%	1.50%	2.38%
1 Camp Nou	2.88%	1.23%	11.62%	2.42%	4.70%	2.80%	0.95%	3.72%	3.72%	3.01%	1.89%	3.79%	1.04%	1.08%	3.32%	4.57%	2.39%	0.33%	44.55%	1.19%
2 Park Güell	10.05%	1.29%	5.29%	1.43%	3.29%	2.51%	0.27%	3.01%	1.86%	2.22%	5.11%	2.41%	0.33%	0.67%	3.13%	3.75%	1.54%	0.33%	0.34%	0.56%
3 Palau de la Generalitat de Catalunya	6.02%	0.78%	1.89%	0.77%	1.13%	0.68%	0.27%	0.35%	1.55%	0.95%	2.27%	0.34%	0.46%	0.59%	0.59%	1.76%	1.47%	0.00%	0.31%	0.94%
4 Casa Batlló	5.37%	0.26%	1.95%	0.77%	0.85%	0.87%	0.27%	0.53%	0.31%	0.32%	0.95%	0.69%	0.59%	0.33%	1.37%	2.56%	1.90%	0.00%	0.17%	0.63%
5 Casa Milà	5.37%	0.39%	1.74%	0.11%	1.13%	0.48%	0.07%	0.71%	0.31%	0.56%	1.14%	0.34%	0.33%	0.59%	0.39%	2.59%	2.39%	0.33%	0.07%	1.00%
6 Arc de Triomf	3.78%	1.03%	1.66%	2.20%	1.60%	0.97%	0.20%	1.60%	0.93%	1.90%	2.08%	1.38%	0.52%	0.52%	1.95%	1.66%	1.12%	0.67%	0.20%	0.50%
7 Catedral de la Santa Creu i Santa Eulàlia	3.20%	0.45%	1.91%	0.55%	0.75%	0.58%	0.20%	1.77%	0.31%	0.40%	0.76%	0.69%	0.26%	0.41%	2.15%	1.52%	0.98%	0.00%	0.20%	0.25%
8 Platja de la Barceloneta	0.59%	0.26%	1.66%	15.97%	0.85%	0.29%	0.34%	1.60%	0.31%	0.56%	0.19%	0.69%	0.39%	0.33%	1.17%	1.19%	0.28%	0.33%	0.20%	0.19%
9 Font Màgica de Montjuïc	0.59%	0.00%	0.90%	0.44%	20.86%	0.29%	0.07%	0.35%	0.00%	0.79%	0.76%	1.03%	0.13%	0.19%	1.95%	0.93%	0.07%	0.00%	0.10%	0.00%
10 Plaça de Catalunya	1.15%	4.46%	1.26%	0.66%	0.66%	0.97%	0.14%	0.53%	1.24%	0.48%	0.57%	0.69%	0.39%	0.52%	0.00%	1.50%	0.98%	0.33%	0.31%	0.19%
11 Museu Nacional d'Art de Catalunya	0.48%	0.45%	1.36%	0.22%	1.88%	0.39%	0.34%	0.00%	0.93%	6.03%	0.57%	0.00%	0.00%	0.22%	0.39%	1.30%	0.28%	0.00%	0.14%	1.88%
12 Lilló	0.04%	0.39%	0.48%	1.87%	0.09%	1.16%	0.81%	0.00%	0.00%	0.56%	0.00%	0.00%	0.13%	8.84%	0.00%	0.01%	0.70%	0.00%	0.31%	0.25%
13 Palau Sant Jordi	0.10%	0.39%	0.25%	0.22%	0.66%	0.48%	0.14%	0.18%	0.31%	18.64%	1.14%	0.34%	0.26%	0.19%	0.00%	0.17%	0.28%	1.33%	0.41%	0.38%
14 Al actor Iscle Soler	1.36%	0.58%	0.65%	0.99%	0.66%	0.39%	0.14%	0.71%	0.31%	0.08%	0.19%	0.34%	0.39%	0.56%	1.17%	1.55%	1.40%	0.00%	0.03%	1.44%
15 Arenas de Barcelona	0.38%	0.19%	1.11%	0.33%	0.75%	11.29%	0.20%	0.71%	0.62%	1.27%	0.95%	0.34%	0.00%	0.19%	0.59%	0.82%	0.49%	0.00%	0.10%	0.31%
16 Parc de la Ciutadella	0.65%	0.52%	0.69%	0.88%	0.75%	0.48%	0.07%	0.71%	1.24%	6.58%	0.95%	1.72%	0.07%	0.37%	0.78%	1.14%	0.28%	0.00%	0.27%	0.44%
17 Castell de Montjuïc	0.48%	0.19%	1.26%	0.11%	1.22%	0.29%	0.07%	0.35%	0.00%	4.36%	0.57%	0.00%	0.07%	0.19%	0.39%	1.70%	0.14%	0.00%	0.00%	0.13%
18 Font de la Plaça d'Espanya	0.61%	0.58%	1.11%	0.33%	1.13%	1.74%	0.14%	0.53%	0.93%	5.47%	0.38%	0.00%	0.26%	0.26%	0.78%	0.87%	0.28%	0.00%	0.14%	0.50%
19 Museu Picasso	0.42%	0.00%	0.71%	0.44%	0.28%	0.29%	0.07%	0.35%	0.31%	3.25%	0.38%	0.34%	0.26%	0.30%	0.39%	1.48%	0.28%	0.00%	0.10%	2.19%

Figure 2. Top 20 most visited attractions and the percentage of their tweets per cluster, highlighting values above average in green.

In summary, the clusters present unique groups of tourists with different characteristics and preferences derived from the clustering features and allows knowing which are the visitors of the most popular tourist POIs. This information can be very useful for the marketing managers of destinations and tourist attractions of the place because it allows them to know the visiting preferences of their visitors. Moreover, these data can also be exploited for mobility analysis, as described in the next section.

4.2. Tourist Mobility/Flow Analysis

To analyze the mobility of tourists within each cluster, bigrams (A, B) were extracted from their sequences of visits. These bigrams are n -grams of size two that represent the movement of a tourist from A to B (i.e., the user sent a tweet from A and the next one from B). Given a sequence of items, n -grams are unique sets of n directly adjacent items. For example, a sequence $S = \{a, b, c, d\}$ has the following 2-grams (popularly known as bigrams) $2grams = \{(a, b), (b, c), (c, d)\}$. N -grams must maintain the order in the original sequence and must be unique.

Bigrams were extracted from the sequences of visited places of each tourist in a certain cluster, and we counted their frequency of occurrence (i.e., the number of times a bigram appeared in the cluster). The clusters 0, 2, 3, 7, and 16 were selected to illustrate this analysis because of their diversity. The top 20 bigrams with a higher frequency in each cluster were taken as the most relevant. Figure 3 shows heat map plots of the tourist mobility within clusters, where the color intensity represents the movement between POIs measured by the frequency of occurrence.

As it can be seen from Figure 3, *Basílica de la Sagrada Família* acts as a hub in cluster 0, as all other POIs are directly connected to it. In most cases, tourists are visiting the other POIs after visiting *Basílica de la Sagrada Família* because its outflows exceed its inflows from other locations, except *Casa Batlló*, *Catedral de la Santa Creu i Santa Eulàlia*, and *Al actor Iscle Soler*, which might be a result of route preference. The strongest connection can be seen between *Basílica de la Sagrada Família* and *Park Güell* with almost equivalent inflow and outflow between them. In the case of cluster 2, *Basílica de la Sagrada Família* is once again the most interconnected attraction but with more inflows than in cluster 0, and the strongest connection is between *Basílica de la Sagrada Família* and *Camp Nou*. The other selected clusters (3, 7, and 16) show connections between various attractions with no clear hub.

Figures 4 and 5 help to further understand tourist mobility with network graphs plotted on the Barcelona city map. Nodes represent POIs, and edges are the bigrams that connect them (the wider is the edge, the more tourists travel between those two locations). It can be seen that proximity plays a role in why *Basílica de la Sagrada Família* acts as a hub in clusters 0 and 2. In cluster 3, the focus is the Mediterranean Sea as most POIs are near the beach, except in the case of *Basílica de la Sagrada Família*, *Park Güell*, and *Camp Nou*, as tourists are willing to travel out of the way to see these POIs. Clusters 7 and 16 show two different kinds of tourists. The former is focused on bars near the city center, whereas the latter visits many different kinds of places all around the city, including amusement parks, shop malls, and the beach, but also the most popular venues.

In summary, the analyzed clusters showed inter-connected POIs which are of interest to certain groups of tourists. In some cases, they focus on the popular attractions, but in others they also visit places off the beaten track. The graphic representation in the map of the mobility of different clusters of tourists in a destination provides new knowledge to DMOs, who could take them into account to define new tourist routes, to create targeted marketing campaigns or to optimize transport routes between heavily connected POIs for different types of tourists.

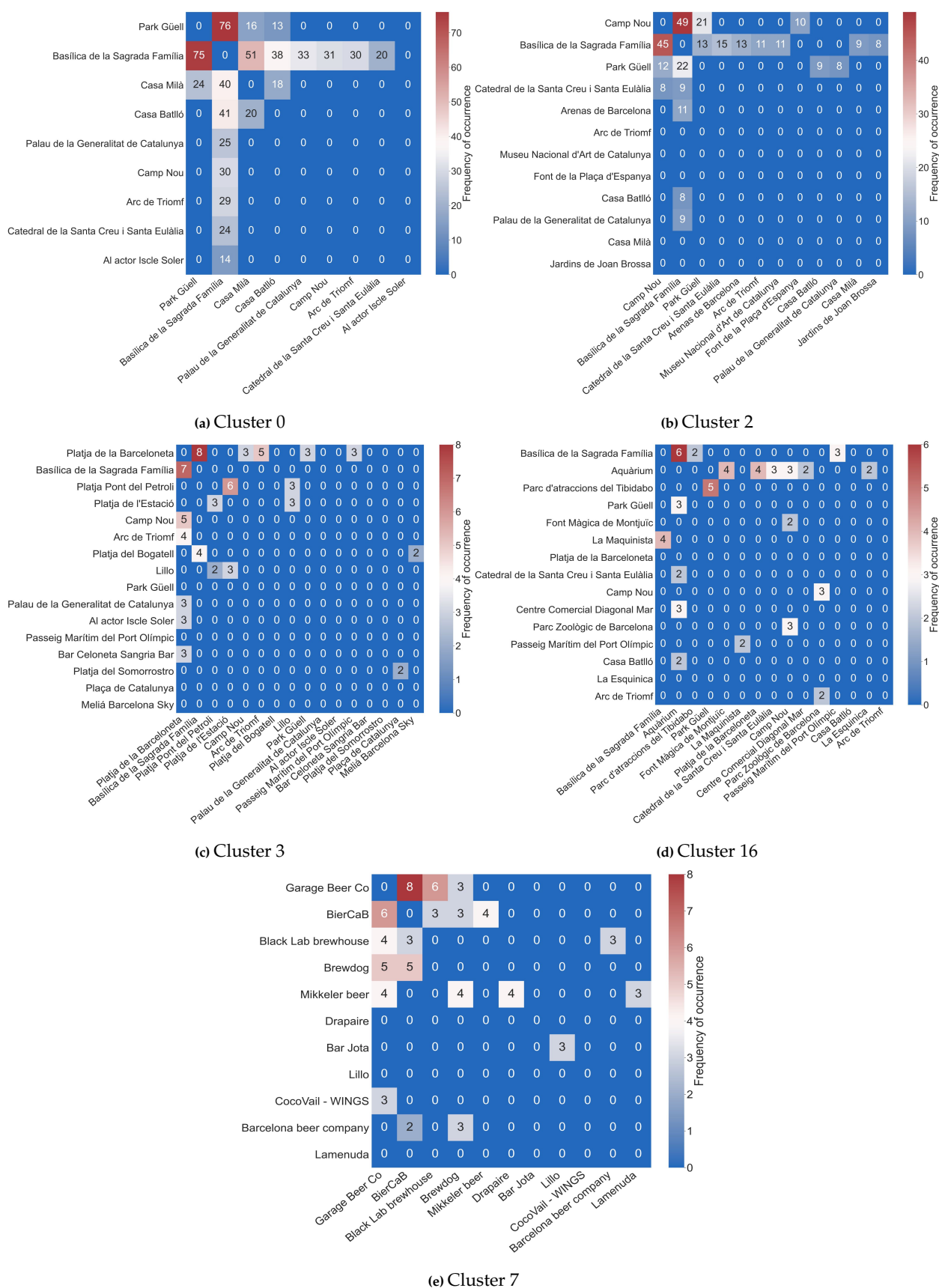


Figure 3. Mobility heat-maps for clusters 0, 2, 3, 16, and 7 representing bigrams with movement from left to bottom.

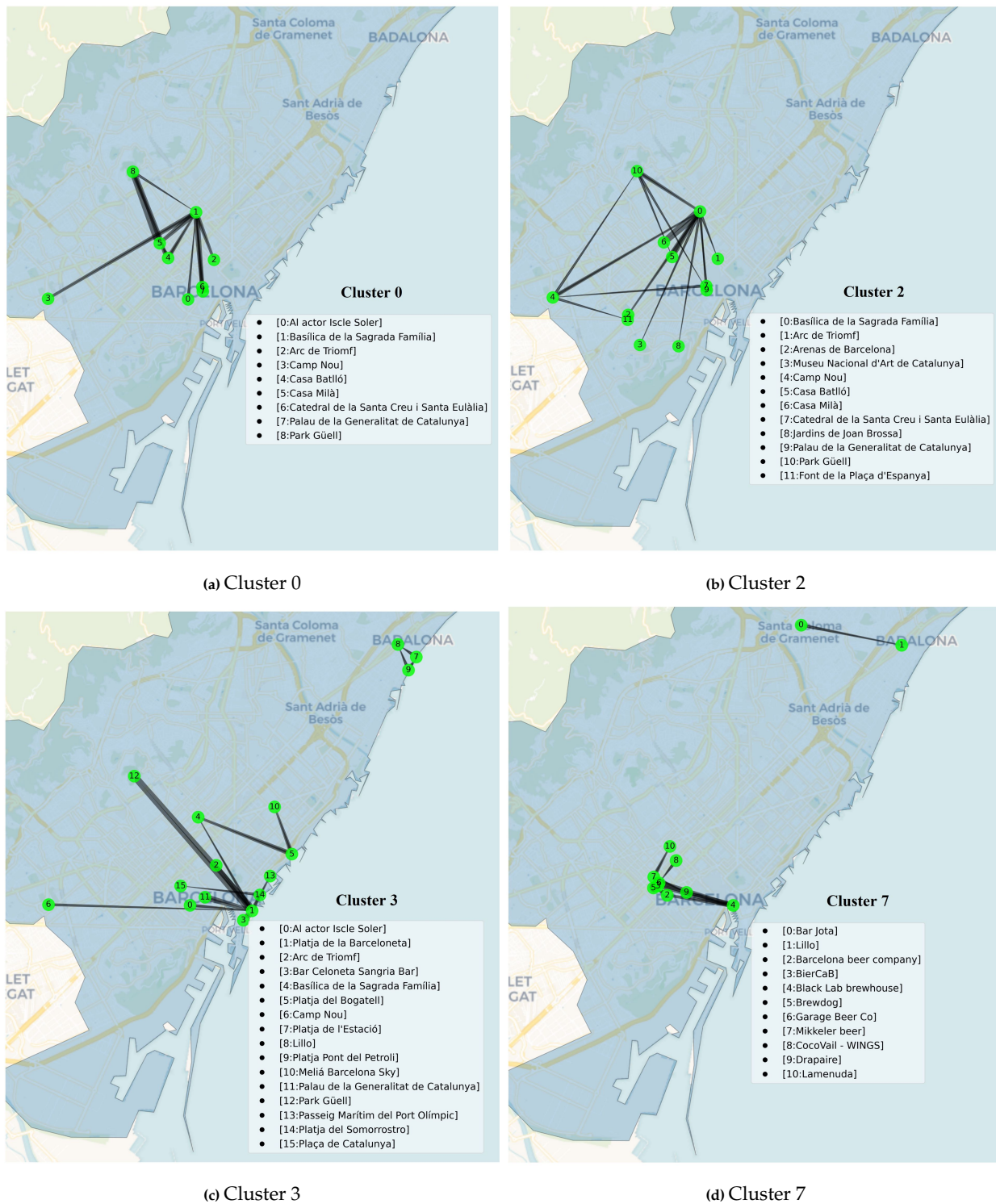


Figure 4. Tourist mobility pattern between attractions in Barcelona for clusters 0, 2, 3, and 7.

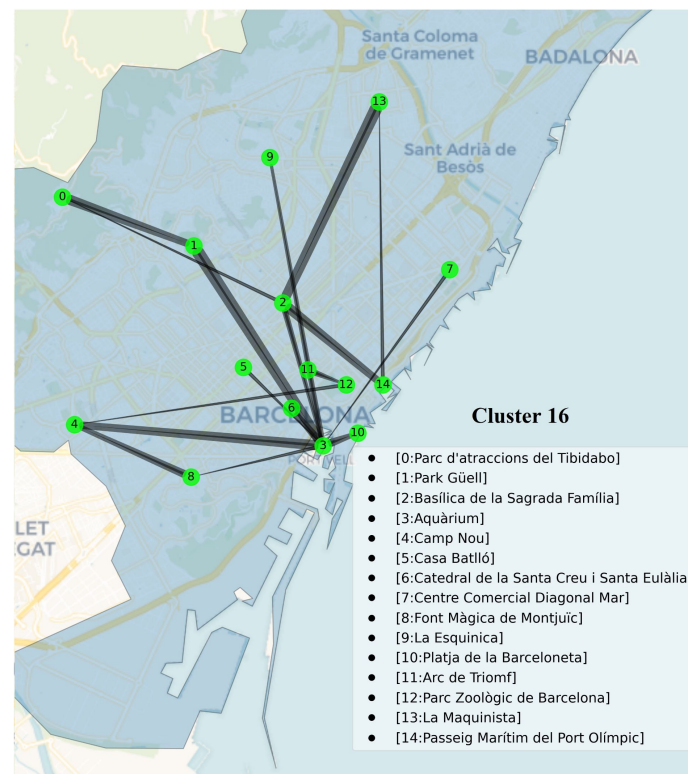


Figure 5. Tourist mobility pattern between attractions in Barcelona for cluster 16.

5. Discussion, Conclusions, and Implications

The main contribution of the study is the introduction of a method for analyzing social media data that create visitor profiles according to their travel preferences and mobility. This study corroborates that social media are very useful platforms as sources for big data research in the field of tourism [8,30], and they can also be used to study tourist mobility [26,35,41,55]. It also improves the knowledge about the different mobility patterns of tourists depending on who they are and what their preferences are [56]. Therefore, the study has shown that clustering visitors by their travel mobility permits uncovering much more information about visitors and their preferences than previous studies. It can also provide complementary information to DMOs, attraction operators, and developers of contextual and next-POI recommender systems [81].

Additionally, the study also reveals the most popular or the most visited points of interest at destinations. Previous studies had also found out the most popular tourist spots or the most visited routes by analyzing geo-tagged social media data [62,66], but their analysis did not allow for knowing which were the most visited POIs by the different tourists. Thus, interestingly, the applied method allows for knowing the percentages of tourists in each cluster who visit each attraction the most. This is crucial for the managers of the different tourist attractions that want to know who their majority visitors are as well as their interests; in that way, they will be able to adapt their service and information in an almost personalized way.

Another contribution of the study is to show the mobility of each cluster of tourists between POIs and to see graphically the movement that they make in the map of the destination. Previous studies have shown mobility with place maps and most visited points [26]. Many studies on GSMD only focused on tourists' mobility patterns [48,59,73]. However, the difference in mobility patterns between sub-groups or clusters of tourists has not been fully researched [37]. Therefore, GSMD analysis allows for knowing the dispersion and tourists' movements, the routes they follow, the activities they carry out in a territory [57–62], and their density of movements [63] and flows [38,64,65]. Hence, the resulting information is particularly valuable for city management, since it provides a

better knowledge of the connections between points of interest related to different clusters of tourists according to their preferences and behaviors. This information can help DMOs to define new tourist routes, to create targeted marketing campaigns, to optimize transport routes between heavily connected POIs for different types of tourists, or to improve the management of congestion or overcrowding situations.

To sum up, through this application of Artificial Intelligence techniques to social media data creating clusters of tourists, it has been possible to know how to segment them according to their visitor behavior and visit preferences. This information is key to the DMOs and the different service providers of the destinations. The interest of DMOs in analyzing big data and knowing the maximum information about their tourists is based on being able to anticipate their interests and preferences [38]. This is precisely the information that this study provides. Therefore, the study can have a major impact on the marketing and flows management of tourist destinations. The exploration of the relationship between tourists' profiles, points of interest, and tourist mobility allows for gaining further insight into really concerning debates on tourism pressure in specific locations, and destination carrying capacity. Accordingly, it can be used by local and regional authorities, as well as by planners and urban designers, to deal with urban complexity, especially in successful tourist cities with contradictions and conflicts generated by overtourism [92].

From this perspective, the managerial implications of the study are diverse. Using these kinds of analytical tools, DMOs and tourism service providers could be able to offer the most personalized services and information, to attract specific types of tourists to certain points of interest [37], to propose the visit of new under-visited attractions to certain market segments, to create new routes or to optimize the existing ones, to enhance public transport services, to develop new POIs or tourist services for the busiest routes [93] or to create tourism development plans for the least visited areas [60].

In addition, it will allow destinations to encourage smart development, overcoming some of the existing gaps in the level of achievement of their objectives [7]. This includes the improvement of smart tourism developments such as the creation of differentiated attractive travel packages [86], the adaptation of the marketing and communication tactics to the preferences of visitors [10,11], the improvement of the satisfaction of tourists [5,14,15], and the co-creation of a more positive tourist destination image [94]. It will even have a major impact on travel recommendation systems, as shown in a previous study [39]. Finally, from the place management perspective, the detection of visitor flow patterns would help to regulate the carrying capacity of the visitors' points of interest avoiding overcrowding, improving allocation of visitor services and reducing tensions produced by the different tourist and residential uses of the city areas, infrastructures, and services around the points of interest.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app12125834/s1>.

Author Contributions: Conceptualization, A.H. and S.A.C.; methodology, J.A.O., J.B. and A.M.; software, J.A.O.; validation, J.B. and A.M.; formal analysis, J.A.O.; investigation, J.A.O.; resources, J.B. and A.M.; data curation, J.A.O. and J.B.; writing—original draft preparation, J.A.O. and A.H.; writing—review and editing, A.H., A.M. and S.A.C.; visualization, J.A.O.; supervision, J.B. and A.M.; project administration, J.B. and A.M.; funding acquisition, J.B. and A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: 3rd Party Data Restrictions apply to the availability of these data. Data were obtained from Twitter and are available at <https://twitter.com> (accessed on 1 January 2019) with the permission of Twitter. Research is allowed to provide Tweet IDs to other researchers to download using Twitter’s API. The Tweet IDs of data presented in this paper are included in the supplementary material.

Acknowledgments: Jonathan Ayebakuro Orama is a fellow of Eurecat’s “Vicente López” PhD grant program.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DMO	Destination Management Organization
GPS	Global Positioning System
GSMD	Geotagged Social Media Data
POI	Point of Interest
SMA	Social Media Analytics
UGC	User Generated Content
UNWTO	United National World Tourism Organization
OSM	Open Street Map

References

1. Buhalis, D.; Law, R. Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. *Tour. Manag.* **2008**, *29*, 609–623. [\[CrossRef\]](#)
2. Hays, S.; Page, S.J.; Buhalis, D. Social media as a destination marketing tool: Its use by national tourism organisations. *Curr. Issues Tour.* **2013**, *16*, 211–239. [\[CrossRef\]](#)
3. Xiang, Z.; Gretzel, U. Role of social media in online travel information search. *Tour. Manag.* **2010**, *31*, 179–188. [\[CrossRef\]](#)
4. Buhalis, D.; Amaranggana, A. Smart Tourism Destinations Enhancing Tourism Experience Through Personalisation of Services. In *Information and Communication Technologies in Tourism*; Tussyadiah, I., Inversini, A., Eds.; Springer: Cham, Switzerland, 2015; pp. 377–389. [\[CrossRef\]](#)
5. Buonincontri, P.; Micera, R. The experience co-creation in smart tourism destinations: A multiple case analysis of European destinations. *Inf. Technol. Tour.* **2016**, *16*, 285–315. [\[CrossRef\]](#)
6. WTCF. WTCF Global Report on Smart Tourism in Cities. World Tourism Cities Federation. Beijing. 2019. Available online: <https://prefeitura.pbh.gov.br/sites/default/files/estrutura-de-governo/belotur/2020/wtcf-global-report-on-smart-tourism-in-cities.pdf> (accessed on 25 January 2022)
7. Femenia-Serra, F.; Ivars-Baidal, J.A. Do smart tourism destinations really work? The case of Benidorm. *Asia Pac. J. Tour. Res.* **2021**, *26*, 365–384. [\[CrossRef\]](#)
8. Xiang, Z.; Fesenmaier, D.R. Big Data Analytics, Tourism Design and Smart Tourism. In *The future of tourism: Innovation and Sustainability*; Xiang, Z., Fesenmaier, D., Eds.; Springer: Cham, Switzerland, 2017; pp. 299–307. [\[CrossRef\]](#)
9. Wang, K.; Lin, C. The adoption of mobile value-added services: Investigating the influence of IS quality and perceived playfulness. *Manag. Serv. Qual. Int. J.* **2012**, *22*, 184–208. [\[CrossRef\]](#)
10. Kotoua, S.; Ilkan, M. Tourism destination marketing and information technology in Ghana. *J. Destin. Mark. Manag.* **2017**, *6*, 127–135. [\[CrossRef\]](#)
11. Lamsfus, C.; Martín, D.; Alzua-Sorzabal, A.; Torres-Manzanera, E. Smart Tourism Destinations: An Extended Conception of Smart Cities Focusing on Human Mobility. In *Information and Communication Technologies in Tourism 2015*; Tussyadiah, I.; Inversini, A., Eds.; Springer: Cham, Switzerland, 2015; pp. 363–375. [\[CrossRef\]](#)
12. Choe, Y.; Fesenmaier, D.R. The Quantified Traveler: Implications for Smart Tourism Development. In *Analytics in Smart Tourism Design*; Xiang, Z., Fesenmaier, D., Eds.; Springer: Cham, Switzerland, 2017; pp. 65–77. [\[CrossRef\]](#)
13. Wang, D.; Xiang, Z.; Fesenmaier, D.R. Adapting to the mobile world: A model of smartphone use. *Ann. Tour. Res.* **2014**, *48*, 11–26. [\[CrossRef\]](#)
14. Boes, K.; Buhalis, D.; Inversini, A. Conceptualising smart tourism destination dimensions. In *Information and Communication Technologies in Tourism*; Tussyadiah, I., Inversini, A., Eds.; Springer: Cham, Switzerland, 2015; pp. 391–403. [\[CrossRef\]](#)
15. Molinillo, S.; Anaya-Sánchez, R.; Morrison, A.M.; Coca-Stefaniak, J.A. Smart city communication via social media: Analysing residents’ and visitors’ engagement. *Cities* **2019**, *94*, 247–255. [\[CrossRef\]](#)
16. Soares, J.C.; Domareski Ruiz, T.C.; Ivars Baidal, J.A. Smart destinations: A new planning and management approach? *Curr. Issues Tour.* **2021**, 1–16. [\[CrossRef\]](#)

17. Gazley, A.; Watling, L. Me, My Tourist-Self, and I: The Symbolic Consumption of Travel. *J. Travel Tour. Mark.* **2015**, *32*, 639–655. [\[CrossRef\]](#)
18. Ma, A.T.H.; Chow, A.S.Y.; Cheung, L.T.O.; Lee, K.M.Y.; Liu, S. Impacts of Tourists' Sociodemographic Characteristics on the Travel Motivation and Satisfaction: The Case of Protected Areas in South China. *Sustainability* **2018**, *10*, 3388. [\[CrossRef\]](#)
19. Oh, J.Y.J.; Cheng, C.K.; Lehto, X.Y.; O'Leary, J.T. Predictors of tourists' shopping behaviour: Examination of socio-demographic characteristics and trip typologies. *J. Vacat. Mark.* **2004**, *10*, 308–319. [\[CrossRef\]](#)
20. Kim, D.Y.; Park, J.; Morrison, A.M. A model of traveller acceptance of mobile technology. *Int. J. Tour. Res.* **2008**, *10*, 393–407. [\[CrossRef\]](#)
21. Fan, D.X.F.; Buhalis, D.; Lin, B. A tourist typology of online and face-to-face social contact: Destination immersion and tourism encapsulation/decapsulation. *Ann. Tour. Res.* **2019**, *78*, 102757. [\[CrossRef\]](#)
22. Kirillova, K.; Wang, D. Smartphone (dis)connectedness and vacation recovery. *Ann. Tour. Res.* **2016**, *61*, 157–169. [\[CrossRef\]](#)
23. Almeida-Santana, A.; Moreno-Gil, S. New trends in information search and their influence on destination loyalty: Digital destinations and relationship marketing. *J. Destin. Mark. Manag.* **2017**, *6*, 150–161. [\[CrossRef\]](#)
24. Pan, B.; Xiang, Z.; Law, R.; Fesenmaier, D.R. The Dynamics of Search Engine Marketing for Tourist Destinations. *J. Travel Res.* **2011**, *50*, 365–377. [\[CrossRef\]](#)
25. Wang, D.; Xiang, Z.; Fesenmaier, D.R. Smartphone Use in Everyday Life and Travel. *J. Travel Res.* **2016**, *55*, 52–63. [\[CrossRef\]](#)
26. Chua, A.; Servillo, L.; Marcheggiani, E.; Moere, A.V. Mapping Cilento: Using geotagged social media data to characterize tourist flows in southern Italy. *Tour. Manag.* **2016**, *57*, 295–310. [\[CrossRef\]](#)
27. Orellana, D.; Bregt, A.K.; Ligtenberg, A.; Wachowicz, M. Exploring visitor movement patterns in natural recreational areas. *Tour. Manag.* **2012**, *33*, 672–682. [\[CrossRef\]](#)
28. Baggio, R.; Scaglione, M. Strategic visitor flows and destination management organization. *J. Destin. Mark. Manag.* **2018**, *18*, 29–42. [\[CrossRef\]](#)
29. Grinberger, A.Y.; Shoval, N. Spatiotemporal Contingencies in Tourists' Intradiurnal Mobility Patterns. *J. Travel Res.* **2019**, *58*, 512–530. [\[CrossRef\]](#)
30. Mariani, M.; Baggio, R.; Fuchs, M.; Höepken, W. Business intelligence and big data in hospitality and tourism: A systematic literature review. *Int. J. Contemp. Hosp. Manag.* **2018**, *30*, 3514–3554. [\[CrossRef\]](#)
31. Marine-Roig, E. Online travel reviews: A massive paratextual analysis. In *Analytics in Smart Tourism Design: Concepts and Methods*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer: Cham, Switzerland, 2017; pp. 179–202. [\[CrossRef\]](#)
32. Mirzaalian, F.; Halpenny, E. Social media analytics in hospitality and tourism: A systematic literature review and future trends. *J. Hosp. Tour. Technol.* **2019**, *10*, 764–790. [\[CrossRef\]](#)
33. Zhang, X.; Yang, Y.; Zhang, Y.; Zhang, Z. Designing tourist experiences amidst air pollution: A spatial analytical approach using social media. *Ann. Tour. Res.* **2020**, *84*, 102999. [\[CrossRef\]](#)
34. Li, X.; Law, R. Network analysis of big data research in tourism. *Tour. Manag. Perspect.* **2020**, *33*, 100608. [\[CrossRef\]](#)
35. Salas-Olmedo, M.H.; Moya-Gómez, B.; García-Palomares, J.C.; Gutiérrez, J. Tourists' digital footprint in cities: Comparing Big Data sources. *Tour. Manag.* **2018**, *66*, 13–25. [\[CrossRef\]](#)
36. Hu, F.; Li, Z.; Yang, C.; Jiang, Y. A graph-based approach to detecting tourist movement patterns using social media data. *Cartogr. Geogr. Inf. Sci.* **2019**, *46*, 368–382. [\[CrossRef\]](#)
37. Liu, Q.; Wang, Z.; Ye, X. Comparing mobility patterns between residents and visitors using geo-tagged social media data. *Trans. GIS* **2018**, *22*, 1372–1389. [\[CrossRef\]](#)
38. Miah, S.J.; Vu, H.Q.; Gammack, J. A big-data analytics method for capturing visitor activities and flows: The case of an island country. *Inf. Technol. Manag.* **2019**, *20*, 203–221. [\[CrossRef\]](#)
39. Orama, J.A.; Borràs, J.; Moreno, A. Combining Cluster-Based Profiling Based on Social Media Features and Association Rule Mining for Personalised Recommendations of Touristic Activities. *Appl. Sci.* **2021**, *11*, 6512. [\[CrossRef\]](#)
40. Anton Clavé, S. Urban Tourism and Walkability. In *The Future of Tourism: Innovation and Sustainability*; Fayos-Solà, E., Cooper, C., Eds.; Springer: Cham, Switzerland, 2019; pp. 195–211. [\[CrossRef\]](#)
41. Jin, C.; Cheng, J.; Xu, J. Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility. *J. Travel Res.* **2018**, *57*, 779–791. [\[CrossRef\]](#)
42. Edwards, D.; Griffin, T. Understanding tourists' spatial behaviour: GPS tracking as an aid to sustainable destination management. *J. Sustain. Tour.* **2013**, *21*, 580–595. [\[CrossRef\]](#)
43. Shoval, N.; McKercher, B.; Ng, E.; Birenboim, A. Hotel location and tourist activity in cities. *Ann. Tour. Res.* **2011**, *38*, 1594–1612. [\[CrossRef\]](#)
44. Ahas, R.; Aasa, A.; Mark, Ü.; Pae, T.; Kull, A. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tour. Manag.* **2007**, *28*, 898–910. [\[CrossRef\]](#)
45. Kuusik, A.; Tiru, M.; Ahas, R.; Varblane, U. Innovation in destination marketing: The use of passive mobile positioning for the segmentation of repeat visitors in Estonia. *Balt. J. Manag.* **2011**, *6*, 378–399. [\[CrossRef\]](#)
46. Roth, C.; Kang, S.M.; Batty, M.; Barthélemy, M. Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. *PLoS ONE* **2011**, *6*, e15923. [\[CrossRef\]](#)
47. Beecham, R.; Wood, J.; Bowerman, A. Studying commuting behaviours using collaborative visual analytics. *Comput. Environ. Urban Syst.* **2014**, *47*, 5–15. [\[CrossRef\]](#)

48. Li, J.; Xu, L.; Tang, L.; Wang, S.; Li, L. Big data in tourism research: A literature review. *Tour. Manag.* **2018**, *68*, 301–323. [\[CrossRef\]](#)
49. Lu, W.; Stepchenkova, S. User-Generated Content as a Research Mode in Tourism and Hospitality Applications: Topics, Methods, and Software. *J. Hosp. Mark. Manag.* **2015**, *24*, 119–154. [\[CrossRef\]](#)
50. Provenzano, D.; Hawelka, B.; Baggio, R. The mobility network of European tourists: A longitudinal study and a comparison with geo-located Twitter data. *Tour. Rev.* **2018**, *73*, 28–43. [\[CrossRef\]](#)
51. Sohrabi, B.; Raeesi Vanani, I.; Nasiri, N.; Ghassemi Rudd, A. A predictive model of tourist destinations based on tourists' comments and interests using text analytics. *Tour. Manag. Perspect.* **2020**, *35*, 100710. [\[CrossRef\]](#)
52. Del Vecchio, P.; Mele, G.; Ndou, V.; Secundo, G. Creating value from Social Big Data: Implications for Smart Tourism Destinations. *Inf. Process. Manag.* **2018**, *54*, 847–860. [\[CrossRef\]](#)
53. Pantano, E.; Priporas, C.V.; Stylos, N. 'You will like it!' using open data to predict tourists' response to a tourist attraction. *Tour. Manag.* **2017**, *60*, 430–438. [\[CrossRef\]](#)
54. Huang, A.; Gallegos, L.; Lerman, K. Travel analytics: Understanding how destination choice and business clusters are connected based on social media data. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 245–256. [\[CrossRef\]](#)
55. Önder, I. Classifying multi-destination trips in Austria with big data. *Tour. Manag. Perspect.* **2017**, *21*, 54–58. [\[CrossRef\]](#)
56. Hawelka, B.; Sitko, I.; Beinart, E.; Sobolevsky, S.; Kazakopoulos, P.; Ratti, C. Geo-located Twitter as proxy for global mobility patterns. *Cartogr. Geogr. Inf. Sci.* **2014**, *41*, 260–271. [\[CrossRef\]](#)
57. Li, Y.; Xiao, L.; Ye, Y.; Xu, W.; Law, A. Understanding tourist space at a historic site through space syntax analysis: The case of Gulangyu, China. *Tour. Manag.* **2016**, *52*, 30–43. [\[CrossRef\]](#)
58. Önder, I.; Koerbitz, W.; Hubmann-Haidvogel, A. Tracing Tourists by Their Digital Footprints: The Case of Austria. *J. Travel Res.* **2016**, *55*, 566–573. [\[CrossRef\]](#)
59. Orsi, F.; Geneletti, D. Using geotagged photographs and GIS analysis to estimate visitor flows in natural areas. *J. Nat. Conserv.* **2013**, *21*, 359–368. [\[CrossRef\]](#)
60. Vu, H.Q.; Li, G.; Law, R.; Zhang, Y. Tourist Activity Analysis by Leveraging Mobile Social Media Data. *J. Travel Res.* **2018**, *57*, 883–898. [\[CrossRef\]](#)
61. Wood, S.A.; Guerry, A.D.; Silver, J.M.; Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **2013**, *3*, 1–7. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Zhou, X.; Xu, C.; Kimmons, B. Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Comput. Environ. Urban Syst.* **2015**, *54*, 144–153. [\[CrossRef\]](#)
63. García-Palomares, J.C.; Gutiérrez, J.; Mínguez, C. Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Appl. Geogr.* **2015**, *63*, 408–417. [\[CrossRef\]](#)
64. Cheng, M.; Edwards, D. Social media in tourism: A visual analytic approach. *Curr. Issues Tour.* **2015**, *18*, 1080–1087. [\[CrossRef\]](#)
65. Miah, S.J.; Vu, H.Q.; Gammack, J.; McGrath, M. A Big Data Analytics Method for Tourist Behaviour Analysis. *Inf. Manag.* **2017**, *54*, 771–785. [\[CrossRef\]](#)
66. Chen, Z.; Shen, H.T.; Zhou, X. Discovering popular routes from trajectories. In Proceedings of the 2011 IEEE 27th International Conference on Data Engineering, Hannover, Germany, 11–16 April 2011; pp. 900–911. [\[CrossRef\]](#)
67. Zanker, M.; Fuchs, M.; Seebacher, A.; Jessenitschnig, M.; Stromberger, M. An Automated Approach for Deriving Semantic Annotations of Tourism Products based on Geospatial Information. In *Information and Communication Technologies in Tourism 2009*; Höpken, W., Gretzel, U., Law, R., Eds.; Springer: Vienna, Austria, 2009; pp. 211–221. [\[CrossRef\]](#)
68. Jurdak, R.; Zhao, K.; Liu, J.; AbouJaoude, M.; Cameron, M.; Newth, D. Understanding Human Mobility from Twitter. *PLoS ONE* **2015**, *10*, e0131469. [\[CrossRef\]](#)
69. Barchiesi, D.; Moat, H.S.; Alis, C.; Bishop, S.; Preis, T. Quantifying international travel flows using Flickr. *PLoS ONE* **2015**, *10*, e0128470. [\[CrossRef\]](#)
70. Ma, S.D.; Kirilenko, A.P.; Stepchenkova, S. Special interest tourism is not so special after all: Big data evidence from the 2017 Great American Solar Eclipse. *Tour. Manag.* **2020**, *77*, 104021. [\[CrossRef\]](#)
71. Dietz, L.W.; Sen, A.; Roy, R.; Wörndl, W. Mining trips from location-based social networks for clustering travelers and destinations. *Inf. Technol. Tour.* **2020**, *22*, 131–166. [\[CrossRef\]](#)
72. Sugimoto, K.; Ota, K.; Suzuki, S. Visitor Mobility and Spatial Structure in a Local Urban Tourism Destination: GPS Tracking and Network analysis. *Sustainability* **2019**, *11*, 919. [\[CrossRef\]](#)
73. Gabrielli, L.; Furlotti, B.; Trasarti, R.; Giannotti, F.; Pedreschi, D. City users' classification with mobile phone data. In Proceedings of the 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, USA, 29 October–1 November 2015; pp. 1007–1012. [\[CrossRef\]](#)
74. Li, D.; Zhou, X.; Wang, M. Analyzing and visualizing the spatial interactions between tourists and locals: A Flickr study in ten US cities. *Cities* **2018**, *74*, 249–258. [\[CrossRef\]](#)
75. Wu, Y.; Li, Z.; Wu, W.; Zhou, M. Response selection with topic clues for retrieval-based chatbots. *Neurocomputing* **2018**, *316*, 251–261. [\[CrossRef\]](#)
76. Batra, A. Senior pleasure tourists: Examination of their demography, travel experience, and travel behavior upon visiting the Bangkok metropolis. *Int. J. Hosp. Tour. Adm.* **2009**, *10*, 197–212. [\[CrossRef\]](#)
77. Vu, H.Q.; Li, G.; Law, R.; Ye, B.H. Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tour. Manag.* **2015**, *46*, 222–232. [\[CrossRef\]](#)

78. Ahn, M.J.; McKercher, B. The Effect of Cultural Distance on Tourism: A Study of International Visitors to Hong Kong. *Asia Pac. J. Tour. Res.* **2015**, *20*, 94–113. [[CrossRef](#)]
79. Phillips, W.J.; Jang, S. Destination image differences between visitors and non-visitors: A case of New York city. *Int. J. Tour. Res.* **2010**, *12*, 642–645. [[CrossRef](#)]
80. Domènech, A.; Gutiérrez, A.; Anton Clavé, S. Cruise Passengers' Spatial Behaviour and Expenditure Levels at Destination. *Tour. Plan. Dev.* **2020**, *17*, 17–36. [[CrossRef](#)]
81. Massimo, D.; Ricci, F. Clustering Users' POIs Visit Trajectories for Next-POI Recommendation. In *Information and Communication Technologies in Tourism 2019*; Pesonen, J., Neidhardt, J., Eds.; Springer: Cham, Switzerland, 2019; pp. 3–14. [[CrossRef](#)]
82. Manca, M.; Boratto, L.; Morell Roman, V.; Martori i Gallissà, O.; Kaltenbrunner, A. Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Soc. Net. Media* **2017**, *1*, 56–69. [[CrossRef](#)]
83. Fuchs, M.; Höpken, W.; Lexhagen, M. Big data analytics for knowledge generation in tourism destinations—A case from Sweden. *J. Destin. Mark. Manag.* **2014**, *3*, 198–209. [[CrossRef](#)]
84. Paldino, S.; Bojic, I.; Sobolevsky, S.; Ratti, C.; González, M.C. Urban magnetism through the lens of geo-tagged photography. *EPJ Data Sci.* **2015**, *4*, 1–17. [[CrossRef](#)]
85. Van der Zee, E.; Bertocchi, D. Finding patterns in urban tourist behaviour: A social network analysis approach based on TripAdvisor reviews. *Inf. Technol. Tour.* **2018**, *20*, 153–180. [[CrossRef](#)]
86. Vu, H.Q.; Li, G.; Law, R. Cross-Country Analysis of Tourist Activities Based on Venue-Referenced Social Media Data. *J. Travel Res.* **2020**, *59*, 90–106. [[CrossRef](#)]
87. Xu, Y.; Li, J.; Belyi, A.; Park, S. Characterizing destination networks through mobility traces of international tourists — A case study using a nationwide mobile positioning dataset. *Tour. Manag.* **2021**, *82*, 104195. [[CrossRef](#)]
88. Huang, Q.; Wong, D.W.S. Activity patterns, socioeconomic status and urban spatial structure: What can social media data tell us? *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 1873–1898. [[CrossRef](#)]
89. Han, S.; Ren, F.; Wu, C.; Chen, Y.; Du, Q.; Ye, X. Using the TensorFlow Deep Neural Network to Classify Mainland China Visitor Behaviours in Hong Kong from Check-in Data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 158. [[CrossRef](#)]
90. Liao, Y. Hot Spot Analysis of Tourist Attractions Based on Stay Point Spatial Clustering. *J. Inf. Process. Syst.* **2020**, *16*, 750–759. [[CrossRef](#)]
91. Giglio, S.; Bertacchini, F.; Bilotta, E.; Pantano, P. Machine learning and points of interest: Typical tourist Italian cities. *Curr. Issues Tour.* **2020**, *23*, 1646–1658. [[CrossRef](#)]
92. Lew, A.; McKercher, B. Modeling Tourist Movements: A Local Destination Analysis. *Ann. Tour. Res.* **2006**, *33*, 403–423. [[CrossRef](#)]
93. Chancellor, H.C. Applying travel pattern data to destination development and marketing decisions. *Tour. Plan. Dev.* **2012**, *9*, 321–332. [[CrossRef](#)]
94. Jabreel, M.; Huertas, A.; Moreno, A. Semantic analysis and the evolution towards participative branding: Do locals communicate the same destination brand values as DMOs? *PLoS ONE* **2018**, *13*, e0206572. [[CrossRef](#)] [[PubMed](#)]