



# Article Utilizing Urban Geospatial Data to Understand Heritage Attractiveness in Amsterdam

Sevim Sezi Karayazi \*<sup>10</sup>, Gamze Dane and Bauke de Vries <sup>10</sup>

Department of the Built Environment, Information Systems in the Built Environment, Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands; g.z.dane@tue.nl (G.D.); b.d.vries@tue.nl (B.d.V.) \* Correspondence: s.s.karayazi@tue.nl

Abstract: Touristic cities are home to historical landmarks and irreplaceable urban heritages. Although tourism brings financial advantages, mass tourism creates pressure on historical cities. Therefore, "attractiveness" is one of the key elements to explain tourism dynamics. User-contributed and geospatial data provide an evidence-based understanding of people's responses to these places. In this article, the combination of multisource information about national monuments, supporting products (i.e., attractions, museums), and geospatial data are utilized to understand attractive heritage locations and the factors that make them attractive. We retrieved geotagged photographs from the Flickr API, then employed density-based spatial clustering of applications with noise (DBSCAN) algorithm to find clusters. Then combined the clusters with Amsterdam heritage data and processed the combined data with ordinary least square (OLS) and geographically weighted regression (GWR) to identify heritage attractiveness and relevance of supporting products in Amsterdam. The results show that understanding the attractiveness of heritages according to their types and supporting products in the surrounding built environment provides insights to increase unattractive heritages' attractiveness. That may help diminish the burden of tourism in overly visited locations. The combination of less attractive heritage with strong influential supporting products could pave the way for more sustainable tourism in Amsterdam.

**Keywords:** location-based social media data; urban geospatial data; Flickr data; heritage; spatial analysis; DBSCAN; OLS; GWR

## 1. Introduction

Tourism is one of the fastest-growing industries for many cities, especially for cities with significant historical values. In many historical cities, certain areas become very attractive to visitors due to their significant heritage buildings or sites, over-promotion, and supporting facilities and services [1,2]. In contrast, some areas that have heritage significance may not attract many visitors. Although tourism creates economic benefits, the unbalanced distribution of visitors in cities can result in negative impacts, such as excessive crowds, loss of cultural and local values, and environmental degradation, both in overly and underly represented areas.

Historic cities consist of tangible and intangible heritages, monuments, and cultural landscapes [3]. The service providers such as hotels, restaurants, and tour guides shape functional places for tourists [4]. The combination of historical places and services creates an attractive historic area. Thus, in addition to the location of heritage places, the supporting services and facilities (i.e., public transport stops, eating/drinking facilities) play an important role in their attractiveness [5,6]. For instance, Vong and Ung's (2012) study found that history and culture, facilities and services at heritage sites, heritage interpretation, and heritage attractiveness are distinctive factors for heritage visitors in Macau. Other distinctive factors of heritage visitation are how tourists perceive and experience a place [7]. In that sense, visitation is influenced by heritage's surroundings, such as tram-metro stations, eating points, local products, attractions, museums, open markets,



Citation: Karayazi, S.S.; Dane, G.; Vries, B.d. Utilizing Urban Geospatial Data to Understand Heritage Attractiveness in Amsterdam. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 198. https://doi.org/10.3390/ijgi10040198

Academic Editors: Wolfgang Kainz and Maria Antonia Brovelli

Received: 31 December 2020 Accepted: 22 March 2021 Published: 25 March 2021

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



**Copyright:** © 2021 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/). and so on. Therefore, understanding the distribution of visitors at the city level and what built environment characteristics attract visitors to these places are important aspects for developing sustainable policies on visitor management. As a result of this, suggestions can be given for making under tourism areas more attractive to visitors and therefore, evenly distribute the visitors in cities.

Emerging technologies have dramatically changed the way of retrieving information in recent years. Big data from newly available sources, such as location-based social networks, sensors, that consist of a wide range of information and present data-driven evidence [8]. This type of data contains mainly three key concepts "3 V's"; volume represents a large amount of quantity; velocity is the measurement of how fast data arrives from sources and variety in the range of data types [9]. Afterward, this concept was updated by adding veracity [10], which represents the accuracy and applicability of data and value [11], representing the potential of big data. Such data influx comes with various information, such as where people visit, how they move in the city; therefore, newly available big data have become an essential source for urban studies and planning practices.

Urban tourism and heritage studies that utilized the newly available big data sets have focused on different subjects, such as destination management [3,12], tourist activities in historical cities [13], mapping historical values [14], and investigating historical places [15]. These studies conclude that such data are beneficial to investigate the relation between heritage sites and people's activities.

Recent studies indicate that a wide range of datasets from social media [13,16–22], global positioning systems (GPS) [23–25] and sensors can provide a better understanding of visitors' behavior and patterns compared to traditional methods, such as surveys. Such newly available datasets are also used for tourism and heritage studies in order to understand visitors' patterns and behavior. Koutras et al. [15] focused on tourist behavior using location-based social network data in Athens, and their study identified the temporal tourist concentration in every POI, weekly, monthly and yearly time intervals by using the spatiotemporal characteristic of Flickr data set. Devkota et al. [16] utilized Flickr and Twitter datasets to explain tourism areas of interest (TAOI) with density-based spatial clustering of applications with noise (DBSCAN). Their findings revealed that the top ten keywords also point out the most photo-shared locations and activities associated with these locations. In addition to that, they suggested that integrated data from multiple sources generate better scores for TAOI definition considering natural language processing (NLP) algorithm schemes. Garcia Palomerez et al. [17] focused on identifying tourists' hot spots based on social networks, and they revealed that uploaded photos are concentrated around monuments, tourist attractions, and museums. Ganzaroli et al. [18] analyzed the efficiency of TripAdvisor on the quality of restaurants as part of the cultural heritage of Venice, and it was concluded that the ranking of restaurants is strongly related to visitors' expected quality in Venice. It was found that tourists' photographs are clustered in the city center; however, locals' movements are extended, such as parks and recreational areas. Girardin et al. [19] carried out a study on quantifying urban attractiveness using Flickr photo tags and AT&T network data, digital footprints, and it was found that the attractiveness of waterfronts has increased over the summer. Gede et al. [20] focused on the aggregation of photographs from the Flickr dataset along the Danube river. They formed a time series of data for each user and applied cluster analysis. Results show that concentrated points are considered interconnected destination systems, and the boundaries are defined by borders (i.e., Romania). Runge et al. [21] explained seasonal variation in arctic visitation with the Flickr dataset. They revealed that summer tourism is four times higher comparing 2006 to 2016, and winter tourism increased by 600% at the same time intervals. Also, they investigated the possibility of developing an early-warning system with Flickr; however, it was not possible due to the lack of data. Gong et al. [22] explained crowd characterization with social media dataset from Twitter and Instagram in city-scale events, namely Sail 2015 and King's Day 2016. They analyzed information related to the crowd, such as age, gender, temporal distribution, word use, and so on. Dane et al. [23]

focused on visitor flow with GPS at Dutch Design Week to explain the area of interest locations (AOI). Based on AOI, they analyzed visitors' spatial and temporal behaviors with network analysis. Shoval et al. [24] focused on the current states of tracking technologies in tourism research, and it was found that digital tracking data, such as GPS, mobile network, Bluetooth, geocoded social media data are important assets for these research because of spatial and temporal dimension. Duca et al. [25] focused on creating open data platform similar to Wikipedia extracted from Booking, Facebook, Foursquare and Google Places related to tourism places in European cities. It is found that tourists can benefit from this platform to arrange their trips considering the best accommodation places, tourist attractions, events and so on.

Previous studies show the potential of location-based big data and their value as sources for understanding people's movement patterns and behavior; therefore, to determine the point of interest and attractive locations because it contains two main key factors: space and time relations. In these studies, people's digital footprints were collected by different location-based social media platforms, such as Flickr, Twitter, and Trip Advisor; however, as suggested in the existing studies, Flickr has dominated the big data-based studies. The main reason can be that it was established in 2005, and the database offers millions of photographs. In these studies [13,16–23], it can be seen that researchers utilize various clustering algorithms, such as spatial autocorrelation with Getis-Ord Gi\*, restricted maximum-likelihood (REML), and DBSCAN to reveal hot spots and attractive areas.

The above-mentioned studies utilize newly available big datasets to understand people's movement patterns and behaviors and reveal the attractive places in the cities. However, these studies do not further explore the influence of the characteristics of the heritage and the characteristics of the built environment on the attractiveness of places. For instance, "facilities-services" are influential factors on the attractiveness of heritage locations [26]. Therefore, if visitor movements are traced around the facilities and services (e.g., shopping and eating locations), the relation between facilities and visitors' behavior can be described by statistical methods [27]. This relational behavior can also be visualized by mapping and simulation techniques. Another example is "overcrowding", which is associated with heritage attractiveness [26], and it can be analyzed whether overcrowding influences heritage locations or not. In that sense, large location-based social network data can also be utilized for such purposes [28].

To our knowledge, this is the first study where the influence of heritage attributes and built environment attributes are assessed on heritage's attractiveness by utilizing Flickr and city data on heritages and supporting services and facilities. In this research, we analyzed that photographed heritages are attractive and unphotographed heritages are unattractive. With the increase of social media usage, more people have a tendency to take photographs and leave their digital footprint voluntarily. In this sense, we designed research that was mainly based on online photo-sharing platforms (Flickr) and sub-datasets, such as monument data [29] and Amsterdam city data [30]. This research was done by first clustering Flickr data via the DBSCAN algorithm in order to identify the attractive areas. Later on, the city data on heritage buildings and sites were matched with attractive areas. In our assumption, different heritage types are not prone to photograph with the same frequency. In order to avoid location bias, we created a 100 m fixed buffer (see in Figure 3a) around hotspots where represent the most photographed places, then we counted heritages. As a result that, heritages were labeled as attractive and unattractive. Thereafter we analyze the relationship between these heritages and supporting products that can contribute to attractiveness. Moreover, the built environment characteristics around these heritages, such as the quantity of café, restaurants, museums, tram-metro stops, were also added to the data. Then this final dataset was analyzed by using ordinary least squares (OLS) and geographically weighted regression (GWR). The results of OLS allowed us to explain the parameters with a significant effect on the attractiveness and to classify these parameters based on their potential effects for further research. The results of GWR contributed to a

better understanding of what factors are influential by taking into consideration variables' locations on heritage attractiveness.

Our research has different viewpoints in several aspects. As seen in current studies that have utilized location-based social media data (i.e., Flickr, Panoramio, Twitter, TripAdvisor), focus on revealing the point of interest and hotspots in a city. Limited works have been done to explore the attractiveness of heritage with newly available big data, such as Flickr and city data on heritage and facilities and services. Therefore, our paper has two steps (i) finding the attractive areas in the city and the heritages that are within these areas (ii) understanding what makes these heritages attractive in terms of characteristics of heritages and the surrounding built environment. We selected Amsterdam as a case area for this study since it is a historical city that has faced overtourism issues in recent years.

This paper is organized as follows: first, the data collection procedure and variables are explained. Then, the methodologies that are used to analyze the data are introduced. After this, the findings of the study are shown. The paper concludes by discussing findings and the effectiveness of the methodologies and suggestions to policymakers.

# 2. Data and Methods

Our methodology was structured in four main parts, as shown in Figure 1. First, data collection from different sources is explained; thereafter, cluster analysis with the DBSCAN process is explained. After this, the identification of attractive heritages is presented. Then a new data matrix is designed to estimate heritage attractiveness with geographically weighted regression.



Figure 1. Methodological flow chart used in the data process.

#### 2.1. Data Collection and Preparation

Amsterdam was chosen as a case study, which is an important tourism destination, and its historical core has been listed within the World Heritage Site (WHS) by the Scientific and Cultural Organization (UNESCO) [31]. For this study, three different datasets were processed to understand the relation between heritages and supporting products in the urbanscape: Flickr dataset, national monument dataset, and Amsterdam city dataset of services and facilities.

Flickr API allows downloading large datasets along with the metadata; it was employed to determine the attractive areas for visitors in time and space. Note that an attractive area refers to the clusters of a high number of photos taken and tagged at a location. The metadata of 285,130 photographs were downloaded within bordered area the boundary of "minx": 4.867080, "miny": 52.357924, "maxx": 4.933176, "maxy": 52.390259. These coordinates represent the border of the study area. Code snippets were run by using GO language, and the dataset was saved as a comma separated value (CSV) file. Flickr database contains photo's taken time, uploaded time, location (latitude and longitude), and description (tags) of photos. The location of photos can be uploaded automatically from the camera or manually assigned on the map. In this research, latitude, longitude, owner, date taken, and URL information were retained. Due to the data privacy, user-related information, such as first name, last name, website, occupation and hometown, were not analyzed, and the results did not contain any personal information. The attributes of Flickr data per photo can be seen in Appendix A, Table A1.

The downloaded Flickr dataset was cleaned by removing the duplicate and invalid records to minimize the dataset's errors. Flickr downloading algorithm had to be run many times to retrieve all the Flickr data from Amsterdam. However, it resulted in duplicated data in some locations. The repeated data were removed by cleaning the same URLs. The timestamps of uploaded photos were used to divide the users into tourists and locals by considering taken time range of the photographs. As suggested in the literature [15,19], the classification between tourists and locals was processed based on the taken date intervals of photos. Timestamps were divided into 30 day periods. If a user uploaded more than one photograph within 30 days, that user was assumed as a tourist since shorter periods are preferred by tourists. [13]. If the period is greater than 30 days, they were considered locals. This enabled us to understand the clusters per user group so that a variety of clusters could be found.

The unprocessed dataset contained the data records of 28,130 photographs. In the dataset, timestamps of photographs varied from 1927 and 2019. The photographs were taken before 2007 were checked manually. These old photos were kept in the dataset because they still represent spatiotemporal information. As a result, the photographs that were taken before 2007 were merged both for locals and tourists. The distribution of photographs per year can be seen in Appendix A, Table A2. Data records of 93,752 photographs belong to tourists, and 191,378 photographs belong to locals. Figure 2a represents a valid photograph that was used in the research. The upper graph shows the number of photographs taken by tourists and locals before and after the cleaning process. The bottom graphs with percentages show the number of users. Although the majority of photographs were taken by locals, the number of users was less than the tourists. For the final Flickr dataset, 12,766 photographs remained from 1808 tourists, and 25,445 photographs remained from 654 locals. 73% of photos were uploaded by tourists, and 27% were uploaded by locals. Valid photographs and user distribution (Figure 2a) and spatial distribution of Flickr photographs (Figure 2b) are given below. The final data file consisted of latitude, longitude, date taken and URL columns per photo record.

It can be seen in Figure 2b that the photos of tourists were taken mainly in the urban core, while the photos of locals covered a wider range of the city. The cleaned Flickr data were utilized to cluster analysis to identify the most attractive locations in Amsterdam.

National monuments data were used to analyze the relationship between the Flickr photo clusters and urban heritage areas. Seven thousand five hundred heritages are registered in the national monuments dataset of Amsterdam, which was downloaded from the Cultural Heritage Agency website of the Ministry of Education, Culture and Science [29]. All data were provided in shape format; therefore, it was possible to use them directly within a GIS software environment. Each heritage object has its own attributes, including coordinates, function, purpose (i.e., building, church, monument, object), postcode, street name, municipality and so on.

For the analysis, 25 predefined heritage types were assigned to 14 categories regarding their functions. This was done because some of the heritage types had less representation in the dataset, which could influence the results of regression analysis. Thus, heritages were



assigned to more concentrated groups considering their functions. Heritage classification can be seen in Table 1.

(a)

(b)

Figure 2. (a) Valid photographs and user distribution; (b) Flickr photographs.

Table 1. Heritage classification.

Heritage Type	Group
Catering	Catering
Church	Church
Education	Education
House	House
Industry	Industry
Shopping	Shopping
Storage	Storage
Transportation	Transportation
Uncategorized	Uncategorized
Office building	Office
Garden and park Zoo	Garden and zoo
Administration building Court Governmental building Military Service home Social care	Governmental building
Fort/fortress Housing part Memorial Remains Street furniture	Remains
Art and culture Sport and recreation	Culture and sport

Finally, Amsterdam city data were gathered for this study. These data were used to understand the influences of urban facilities on the attractiveness of heritages. The city data consists of a wide range of themes, including public spaces, tourism, culture, infrastructure, energy, population. The dataset is presented with different formats, for instance, CSV, microsoft word file (i.e., docx, doc), javascript object notation (JSON), portable document format (PDF), and established in collaboration with Amsterdam Municipality and their partners. All data are open and suitable for researchers and kept up to date [30]. Shopping, public toilet, local product, tram-metro stops, eating, museum, open market and attraction datasets were chosen for this study to explain the degree of their influences on heritages. These datasets contain the location information of each facility and service in Amsterdam.

#### 2.2. Cluster Analysis

In this study, the attractive locations were identified by applying a clustering algorithm on the cleaned Flicker dataset. Cluster analysis is a statistical method to make a group and classify objects according to some features. It is an unsupervised analysis method; therefore, it can be applied without data training. Spatial clustering can be defined as transforming an object into clusters that have similar specifications within groups that can be defined as high homogeneity; however, dissimilar qualification among the other groups, which can be accepted as high heterogeneity. In that sense, for our study, each cluster represents the points of interest (POIs), and concentrated locations can be defined as the most photographed areas [15].

Current literature has plentiful algorithms for clustering, such as K-means, fuzzy C-means, DBSCAN [32–36]. K-means is one of the popular clustering methods that was introduced in 1967 [32]. It sets the mean value of objects in a cluster as a cluster center; also, it is a simple method that computes complexity [33]. K-means demands the number of clusters in advance as an input that can influence clusters' aspects. Another method is fuzzy C-means, which was developed in 1973 [34], and is mainly used for pattern recognition. Both methods, k-means and fuzzy c-means, demonstrate nearly the same strategy. They are based on the Euclidean distance in order to determine the similarities between the considered objects and cluster centroids [35]. Fuzzy c-means are sensitive to the initial cluster centers selection, slowness of convergence, and it has a tendency to become stuck in the optimum local value [36]. Another method is DBSCAN, which is widely utilized in urban planning studies with big data [15,16,33,37]. It is an algorithm that searches for areas of high-density. DBSCAN is run with two parameters; areas of the neighborhood (eps) and minimum points (MinPts) within these areas. When comparing the methods, k-means appeal to researchers who focus on a location optimization problem, while the DBSCAN is better to find geospatial aggregation [33]; in addition, noise points, which are calculated in the DBSCAN, can be assumed as less interesting areas for further analysis.

For this study, DBSCAN was selected as the clustering algorithm [38]. Clusters can be explained as the common shared properties, and noise points can be described as lowdensity regions [39]. The algorithm starts with picking a core point, and it continues to enlarge it until for all density reachable points from the core point. The points that are not listed within any clusters (not reach reachable density point) are assigned as noise points and continue to search until no points remain. Clusters depend on different criteria; core, border, noise, directly density reachable, and density reachable. The core point is in the center of density-based clusters, and it is an array within *eps* and *MinPts*. The border point lies within the neighborhood of the core point. Noise is the point, which is neither the core point nor the border point. Directly density reachable (DDR) is a point *r* is directly density reachable from *s. eps* and *MinPts* belonged to *NEps*(*s*) and  $|NEps(s)| \ge MinPts$ . Density reachable (DR) is a point *r* is reachable from point *s. eps* and *MinPts* if there is a sequence of points  $r1 \dots rn$ , r1 = s, rn = s such that ri + 1 is directly reachable from ri [39]. The algorithm finds dense areas and creates arbitrarily shaped clusters [40].

#### 2.3. Estimation of Heritage's Attractiveness

As suggested in the literature, heritage attractiveness is shaped by several factors, such as heritage's atmosphere, special events, conservation status [41], heritage's historical value [26], and cultural background [7]. In this study, considering the heritage's surroundings, supporting products, such as shopping facilities, public toilets, local products, tram-metro stops, eating points, museums, open markets, and attractions were analyzed to understand these factor's degree of influence on heritage attractiveness. Therefore, first, the clusters were identified by the DBSCAN algorithm on the Flickr photo locations. Later, these clusters were extended with 100 m buffers in order to avoid location biases. Following that, the heritages that fall under the buffered clusters were identified as attractive heritages, and the rest were labeled as unattractive heritages. Then the labeled heritage data set with their locations were brought together with the secondary datasets of facilities and services from the Amsterdam city data. This final dataset was used to apply OLS and GWR analysis.

On the final dataset, OLS and GWR analyses were applied. Regression analysis is a statistical model to reveal a correlation between one dependent and several independent variables. Linear regression is the most used mode in geographical analysis; however, non-stationary variation can be missed in simple global fitting methods, such as OLS. On the other hand, GWR can provide an alternative method for the analysis to consider spatial variations; GWR enables local variations in the estimation of coefficient; therefore, the regression coefficient is calculated by different values for each location [42]. The benefit of GWR is that it allows for the identification of selected attributes from a large group of possible criteria with a significant impact on attributes and ranks them in line with their weights.

Before starting the OLS, multicollinearity among the variables was checked by variance inflation factor (VIF) scores by processing the exploratory regression tool in ArcGIS 10.8. According to theory, multicollinearity is a problem for estimation when the VIF score is greater than 10 [43]. In order to avoid misinterpretations, local products were subtracted from the matrix. In addition, spatial autocorrelation was checked by Moran's Index to observe how variables were distributed. As a final step, GWR was conducted to observe the spatial distribution of attractive heritages. A set of models was analyzed with the assistance of ArcGIS 10.8.

#### 3. Results

In this section, we first explain the results of cluster analysis; then, we describe the identifying heritage attractiveness. After this, OLS and GWR results that were used for understanding heritage attractiveness are explained. Finally, OLS and GWR model diagnostics are presented.

#### 3.1. Cluster Analysis

In this study, for the selection of parameters (*MinPts* and *eps*) of the DBSCAN clustering algorithm, different values were tested, and *minPts* = 125 for tourists, *minPts* = 175 for locals datasets were determined after several trials. Then, *eps* was calculated using the *kNN* function in the R programming language, and the best *eps* output was 70, both for local and tourist datasets. Smaller *minPts* results in more clusters and this could lead to deceptive analysis because clusters were spread within the core. The experiment resulted in 9 clusters for photographs by tourists and 12 clusters for photographs by locals. Results are shown in Figure 3a with a map and Figure 3b with numbers. The clusters for tourists and locals were combined for the next steps.

#### 3.2. Identifying Attractive Heritages

This step aimed to identify the attractive heritages in the city of Amsterdam, based on the clusters of most photographed locations. Existing clusters that represent the most photographed locations in Amsterdam were processed to create buffers. The buffering can allow avoiding location bias that photographs' geotags might have. In this sense, 100 meter fixed buffers around the clusters were created in ArcGIS 10.8. They represent a polygon within a specified proximity of each photographed point. Afterward, heritages were intersected with the buffered areas. The heritages, which fall into the buffered areas (Figure 3a), were assumed as attractive because photographs were taken around them; otherwise, non-intersected heritages were assumed as unattractive heritages. Figure 4 represents the number of heritages per heritage category that were found to be attractive and unattractive.



Figure 3. Distribution of clusters (points of interest (POIs)) for tourist and local: (a) map; (b) counts of photos per POI.

According to Figure 4a, the most attractive heritage types were house buildings, including narrow canal houses that are registered as a national monuments and attract visitor's attention. Following that, culture-sport heritages were found the second most attractive type, including art-culture heritages and sport-recreation heritages.

#### 3.3. Estimation of Heritage Attractiveness

In order to analyze the attractiveness of heritages, the photographs by locals and tourists were merged together. For the final sample, 118 heritages were found attractive, and 704 heritages were found unattractive. In order to avoid spatial bias, heritages were aggregated into three groups. The main reason is that some heritage types were underrepresented, such as shopping (n = 4), catering (n = 18), transportation (n = 35), on the other hand, some of them were overrepresented storage (n = 327), governmental building (n = 154), a church (n = 58). As seen in Table 2, heritage types were formed into three groups: culturescience, commercial-governmental and recreation, considering the heritages' functions. For instance, enjoyment-related heritages, including caterings (i.e., eating-drink points) and recreational places (i.e., garden and park), were assigned to the same group. The houses remain, and uncategorized heritages were removed from the heritage types in order to ensure data robustness. As indicated in Figure 4, houses were found the most attractive heritages because the majority of heritages were registered as houses. They were overrepresented in data, and experiment results can be biased. However, the influences of houses cannot be neglected; therefore, houses were renamed as residential and included in the regression as supporting facility/services. Narrow canal houses can be considered an attractive factor due to the UNESCO heritage status. It can be evaluated as an independent explanatory variable to describe the influences of canal houses on attractiveness. Finally, heritages' surrounding facilities and services were added to the data set as independent variables: shopping facilities, public toilets, local products, tram-metro stations, eating points, museums, open markets, and attractions. The number of these variables was counted within each neighborhood of Amsterdam using the boundaries from the Centraal Bureau voor de Statistiek (CBS) dataset [44]. The number of points within the neighborhood polygons was counted added to the data matrix. The final dataset was uploaded to ArcGIS 10.8 for analysis.



Figure 4. Distribution of heritages per type: (a) attractive; (b) unattractive.

Table 2. Group of heritage types.

Culture and Science	Commercial and Governmental	Recreation
Church	Governmental building	Catering
Culture and sport	Industry	Garden, park and zoo
Education	Shopping	-
	Storage	
	Transportation	
	Office building	

#### 3.3.1. The Results of Global Model (OLS)

Regression analysis is a commonly used statistical method in spatial research to evaluate the relationship between explanatory and response variables. OLS provides a global model of the variable to understand relations among the estimators. Thus, OLS presents the probabilities of estimators and the strength of their coefficients.

Before starting with OLS, we investigated multicollinearity. It was indicated in the summary of multicollinearity statistics from the ArcGIS 10.8 exploratory regression tool that at least two or more variables are strongly correlated. When VIF values were checked, local products and tram-metro stops showed perfect multicollinearity. Therefore, local products were omitted from the regression analysis instead of tram metro stops because other variables, such as open market and shopping, can be analyzed to estimate commercial-related factors. Table 3 shows descriptive statistics of each estimator with mean, maximum, minimum values, standard deviations, and VIF scores. As a rule of thumb, chosen parameters (estimators) should not have a strong correlation; therefore, the VIF value should be below 10.

Variables	Mean	Standard Deviation	Minimum	Maximum	VIF
Attractive heritage	0.25	1.447	0	18	-
Commercial	1.16	4.645	0	54	1.495
Recreation	0.17	0.971	0	16	1.217
Residential	10.53	36.857	0	371	3.076
Attraction	0.68	1.527	0	12	4.216
Eating	1.31	2.885	0	31	2.231
Museum	0.25	0.802	0	6	3.251
Open market	0.09	0.373	0	3	1.186
Public toilet	0.38	1.001	0	10	1.313
Shopping	1.21	4.047	0	57	2.713
Tram-metro stops	0.18	0.507	0	4	1.338

Table 3. Descriptive statistics of the data.

We employed the model using OLS with attractive heritages as the dependent variable. Independent variables were selected as two groups. Heritage types (Table 2.) as commercial and recreational heritages were added. Then supporting products surrounding heritages (i.e., residential heritages, attractions, eating points, museums, open markets, public toilets, shopping locations, and tram-metro stops) were added. Table 4 highlights the results of OLS regression with an adjusted  $R^2$  of 0.384; nearly 39% of variances (see Table 5 for the model's performance) are represented by the model.

#### Table 4. OLS regression results.

Coefficients	Estimate	Std. Error	<i>t</i> -Statistics	Probability <sup>1</sup>
Intercept	-0.011	0.059	-1.851	0.064
Commercial	0.115	0.013	8.455	0.000 ***
Recreation	0.417	0.058	7.086	0.000 ***
Residential	-0.002	0.002	-0.911	0.362
Attraction	0.273	0.069	3.927	0.000 ***
Eating	0.048	0.026	1.800	0.072 *
Museum	-0.093	0.116	-0.800	0.423
Open market	-0.410	0.151	-2.715	0.006 ***
Public toilet	0.149	0.059	2.513	0.012 **
Shopping	-0.085	0.021	-4.055	0.000 ***
Tram metro stops	0.201	0.120	1.673	0.094 *

<sup>1</sup> Significance codes: significant *p* < 0.01 \*\*\*, *p* < 0.05 \*\*, *p* < 0.10 \*.

Model	R <sup>2</sup>	Adjusted R <sup>2</sup>	AICc	RSS
OLS	0.397	0.384	1500.634	605.788
GWR	0.734	0.657	1297.702	266.631

**Table 5.** Ordinary least square (OLS) and Geographically weighted regression (GWR) model performance.

According to the results, the increasing number of commercial heritages, recreational heritages, attractions, eating points, public toilets and tram-metro stops in a neighborhood significantly increased the attractiveness of heritages. Both heritage types (i.e., commercial and recreational) were found significant, meaning that these types of heritages were attractive to people. Considering the supporting products, the increased number of attractions and public toilets in a neighborhood significantly increases the heritage attractiveness; this result was expected because a combination of heritages and services contributes to attractiveness in historical places [5,6]. It is also possible that supporting services and facilities were built because of the existence of cultural heritage in these areas. On the other hand, the increasing number of open markets and shopping facilities that can be supposed as sales location significantly decreased the attractiveness of heritages. Therefore, it can be deduced that commercial heritages as a heritage type are found to be attractive by people. Moreover, commercial supporting products significantly decreased the attractiveness of any heritage type.

In this study, global Moran's I was used as a measure of spatial autocorrelation to calculate the degree of dispersion, randomness and clustering of the data. The Moran's I value is represented by a value of between "1" and "-1". Thus, "1" means perfect positive spatial autocorrelation, "-1" shows perfect negative spatial autocorrelation and "0" means perfect spatial randomness [45]. A high positive local Moran's I value shows the selected target value is similar to within neighborhood, and then the locations are spatial clusters, while a high negative local Moran's I value implies a potential spatial outlier, which is different from the values of within neighborhood locations [46].

For the analysis, residuals from OLS regression with Moran's I were tested by considering neighboring locations. The Moran's I test results showed that the OLS residuals were indicated spatial randomness with Moran's I = 0.003, which was close to "0". In addition to the Moran's I index, z-score and *p*-value, which represent the statistical significance, were calculated. Results from spatial autocorrelation indicated that the z-score was not statistically significant, and residuals showed they were randomly distributed. The distribution of spatial autocorrelation can be seen in Appendix B, Figure A1.

Together with the OLS regression, we also measured the Koenker (BP) statistic, which assesses stationarity. The Koenker (BP) statistic (Koenker's studentized Breusch–Pagan statistic) is a test to determine whether the independent variables in the model have a relationship to the dependent variable in geographic space [47]. We tested the stationarity of heritage attractiveness with Koenker (BP) statistics. The result was significant (p < 0.01); thus, a geographically weighted (non-stationarity) regression model is employed to estimate heritage attractiveness. Diagnostic information of OLS can be seen in Appendix B, Table A3.

#### 3.3.2. The Results of Local Model (GWR)

The global statistics (OLS) that we used previously to explain heritage attractiveness above, generalized throughout the studied area. However, GWR is a statistical modal that covers the localization of regression modeling, and it extends the global analysis to local by involving the spatial component. For the GWR, the following parameters need to be specified; dependent variable, explanatory variables, kernel type, and bandwidth method. In this study, we specified the same key explanatory variables, which were used with OLS. GWR equation is constructed for every feature in the data matrix (dependent and exploratory variables) by specified bandwidth for each target feature. Bandwidth can be selected as either fixed or adaptive. We specified an adaptive kernel, which allows GWR to consider the optimum bandwidth by repeating the number of nearest neighbors for the local regression. The GWR's output map (GWR residuals) can be seen in Figure 5. The results showed that a given z-score -0.264 indicating the model residuals were random (see Appendix B, Figure A2), which displayed the GWR model is able to classify variables.



Figure 5. Spatial distribution of geographically weighted regression (GWR) residuals.

The GWR coefficients with standard deviations intervals are shown with the map in Figure 6, including commercial heritages, recreational heritages, attractions, eating, open market, public toilet, shopping, and tram-metro stops. Each map represents the explanatory variable, and these coefficients indicate how the relationship between each explanatory variable changes across the area. Standard deviation class breaks were used for mapping. When the relation is positive, the dark green areas show where the coefficients were large; on the other hand, negative predictors were shown with dark brown in which the coefficients were small, considering the influence on heritage attractiveness.

The strong positive influence of attractions was observed in the historical core around Amsterdam Centraal Station, De Oude Kerk, Church of Saint Nicholas, Heineken Experience and Zoo Artis; on the other hand, weak predictors were located in more peripheral areas, such as Het Schip, NSDM Werf and Sloterdijk. The positive influence of eating locations was observed strongly in the northern part, including Het Stenen Hoofd, Eye Museum and slightly strong in the southwestern part, such as Occii and Westindische Buurt; however, they showed negative relation at the rest of the Amsterdam excluding southern west parts, such as Occii and Westindische Buurt. A negative correlation was observed between attractive heritages and open markets in the historical core; on the other hand, residuals of public toilets showed positive influence in the core, exclusively in the northern part. The residuals of shopping locations were observed negative to the northern part; however, they showed a slight positive influence around the Zoo Artis, Het Schip and NDSM Werf. The last estimator is that the locations of tram-metro stops were strongly positive to the northern and southeastern parts. The reason could be that Amsterdam Centraal station is located in the northern part, and it can be found attractive heritage spot for visitors.









(c)

(**d**)



(e)



Figure 6. Cont.



(i)

**Figure 6.** GWR residuals for each explanatory variable: (**a**) commercial heritages; (**b**) recreational heritages; (**c**) attractions; (**d**) eating locations; (**e**) open markets; (**f**) public toilets; (**g**) shopping locations; (**h**) tram-metro stops; (**i**) legend.

The result of the GWR revealed that attractions close to Damrak, Bloemenmarkt, Rembrandt House Museum, Museumplein, De Oude Kerk, Church of Saint Nicholas and iconic canals; public toilets around the Amsterdam Centraal Station and Zoo Artis, and tram-metro stops that similar to attractions were contributed to heritage attractiveness in the historical core. On the other hand, open markets close to Zoo Artis, Melkweg, Vondelpark; museums around the Amsterdam Centraal Station, De Oude Kerk, Church of Saint Nicholas; and shopping locations within the northern part of Amsterdam made heritages less attractive in the historical core.

#### 3.3.3. Model Performance

Model diagnostics can be done with R<sup>2</sup>, adjusted R<sup>2</sup>, Akaike's information criterion (AICc), and residual sum squares (RSS) values. The R<sup>2</sup> shows the variation in the dependent variable with the possible number of 0 and 1. The values closer to 1 demonstrate that model has better performance. The adjusted R<sup>2</sup> was found to be 0.657 and showed that GWR could explain almost 66% of the model variance. On the other hand, the OLS performed an adjusted R<sup>2</sup> of 0.384. AICs is another method of model selection that explains the goodness of fit measure. According to theory, lower AICc values are preferable then higher values [48]. While the OLS performed AICc of 1500.634, GWR's AICc was 1292.702. RSS highlights unexplained variations, which were represented by 605.788 from OLS and 266.631 from GWR. Table 5 shows the model performances with R<sup>2</sup>, adjusted R<sup>2</sup>, AICc and RSS. Detailed diagnostic information of OLS and GWR can be seen in Appendix B, Tables A3 and A4 respectively.

## 4. Conclusions

The successful development of touristic heritage areas is associated with identifying the destinations' attractiveness and presence. Heritages and their supporting products should be determined appropriately. The results of a study are presented that encompasses the distribution of the tourists who visit the heritage sites. The variables that influence visitation are explained by combining the newly available dataset and other sources.

Our paper aim was two-fold (i) finding the attractive areas in the city and the heritages that are within these areas (ii) understanding what makes these heritages attractive in terms of characteristics of heritages and the surrounding built environment. The novelties of this study can be explained by these two main aspects. First, DBSCAN was applied to find attractive areas (i.e., overrepresented places) considering the division of tourists and locals by the Flickr data set. An intersection algorithm was employed to extract overlapping heritages obtained national monument database within these areas. Cluster analysis was convenient for the classification of hotspots. When we checked the locations of these clusters, tourists were commonly concentrated around the museums (i.e., Museumplein, Eye Film Museum, Heineken Experience and Het Schip Museum), churches (i.e., De Oude Kerk and Church of Saint Nicholas) and Amsterdam Centraal Station; locals' were aggregated at recreational areas (i.e., Vondelpark, Zoo Artis, NDSM Werf, and Het Stenen Hoofd) and residential places (i.e., Westindische Buurt and Sloterdijk). It can be said that mainstream touristic locations were found attractive by tourists, and less-known hidden places were discovered by the locals. Our analysis showed that houses, culture-sport heritages, and industrial buildings were found attractive (Figure 4). One of the aims of this study was to reveal attractive heritage spots that can be identified as overly touristed locations. Therefore, the Flickr data set is useful to explain overrepresented and underrepresented places. Obtaining visitor data from traditional sources, such as surveys, questionnaires, and the number of visitors from official agencies, could be expensive and demanding in a certain period. This study proves that social media data are capable of capturing visitors' distribution spatially.

According to findings, tourists' clusters were aggregated within Amsterdam's historical core, while locals were distributed over other parts of the city (Figure 3). Because it is important to understand what makes heritages attractive; therefore, the second part of the study focused on a more in-depth understanding of heritage attractiveness to explain the attractiveness with heritage's characteristics and its surrounding built environment. In order to achieve this aim, two regression models, OLS and GWR, were applied to predict the attractiveness of heritages. Our models were successfully classified by OLS with 39.7% variance and GWR with 73.4% variance. In these two models, we used the same explanatory key variables to estimate heritage attractiveness. OLS results showed that heritage characteristics (i.e., commercial heritages and recreational heritages) and supporting products (i.e., attractions, eating points, public toilets, and tram-metro stops) have an influence on attractiveness (Table 4). OLS is a global statistic, and it was applied over the whole study area; however, GWR allowed us to analyze local statistics for each feature. According to GWR results, attractions and tram-metro stops around the Damrak, Bloemenmarkt, Rembrandt House Museum, Museumplein, De Oude Kerk contribute to heritage attractiveness positively as supporting products within the historical core. As it can be seen in Figure 6, other supporting products, such as eating locations in the northern part and the southwestern part, open markets and shopping locations in the east and west part of Amsterdam, can be promoted to reduce tourist pressure in the historical core. Although they did not show a strong influence on the attractiveness in the core, their presence can make the rest of Amsterdam more attractive. Policymakers, destination management organizations (DMOs), and Amsterdam Visitor Centers can draw tourists' attention to these locations. They can offer new routes for visitors because tram-metro stops have already been found attractive. The combination of less attractive heritage with strong influential supporting products, such as tram-metro stops, could pave the way for sustainable tourism in Amsterdam.

# 5. Discussion

The methods (i.e., DBSCAN and GWR) that were employed in this research can also be applied to predict the heritage attractiveness of other cities that will be exposed to overtourism in the future. Combining newly available big data sets from social media, open data sources, such as Amsterdam city data and monument data, allows for deeper analysis. In addition, new models can be developed by considering additional exploratory variables, such as accommodation spots can be analyzed to the relevance of overnight stay, and other travel modes (i.e., walkability index, bike-sharing, bus stops) can be considered to evaluate the impact of the accessibility to the heritages depending on the city. Moreover, temporal characteristics of newly available big data can be used for the future because it can provide more in-depth insight into tourism dynamics. For instance, hourly, daily, monthly, and yearly intervals can contribute to additional understanding about the distribution of visitors in historical cities. With these data, better suggestions can be given to promote locations.

This study has some limitations. Currently, it is mainly based on the Flickr dataset, which can be downloaded at any time with multiple parameters. By nature of these newly available big datasets, they are growing each time of the day automatically. Therefore, results can change with different datasets. Due to the downloading algorithm, it had to be run many times to have the best data, and it resulted in duplicated data in some locations. After the data processing, the amount of data used for the analysis was reduced and cannot be called big data. In addition, not everyone uses social media such as Flickr; therefore, the results may not be representative of the population. Another limitation is that tourists were assigned by considering their time intervals of photographs. As the threshold of time intervals is adjusted, the number of tourists and locals is subject to change. Finally, some heritages may be wrongly identified as unattractive because they were remote to social media interaction. Further research can be employed for cross-validation by making a comparison with different datasets, such as surveys, questionnaires and the number of tickets, to confirm the results. Moreover, to validate the research outcomes, the discussion can be conducted with focus groups and experts. It can contribute to facilitating new development around the unattractive heritages to make them attractive; therefore, the pressure of tourism and the crowd can be relieved in the historical core.

Overall, our study reveals that analyzing the attractiveness of heritages with their types and supporting products in the surrounding built environment provides a valuable perspective for diverting visitors to less crowded areas in terms of overly touristed places. It highlights that understanding the distribution of heritage visitors can show what makes attractive or unattractive in historical cities. In order to relieve the burden of tourism, municipalities and historical city organizations can work on making unattractive heritages more attractive considering the recommendations of this study. For future research, limitations can be diminished by taking into account cross-validation with other datasets.

Author Contributions: Conceptualization, Sevim Sezi Karayazi; data curation, Sevim Sezi Karayazi; formal analysis, Sevim Sezi Karayazi; supervision, Gamze Dane and Bauke de Vries; writing original draft, Sevim Sezi Karayazi and Gamze Dane; writing—review and editing, Gamze Dane and Bauke de Vries. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by The Republic of Turkey Ministry of National Education, Directorate-General for Higher and Overseas Education.

Acknowledgments: We want to thank Theo A. Arentze and Tao Feng for their suggestions on the study design.

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

# Appendix A

 Table A1. Flickr attributes per photo.

Attribute	Explanation	Attribute	Explanation
latitude *	Photo's latitude	url *	Photo's URL
longitude *	Photo's longitude	join date	User's join time
owner *	Photo's identifier	occupation	User's occupation of
date_taken *	Photo's taken time	hometown	User's hometown of
date_unknown *	Unknown taken time	first_name	User's name
date_uploaded *	Photo's upload time	last_name	User's last name
title	Photo's title	website	User's personal website
description	Photo's description	city	User's city
tags	Photo's tags	country	User's country
4 1011			

\* required fields.

**Table A2.** Distribution of photographs per year.

Year	Tourist	Local	Total
2007 and earlier	197	830	1027
2008	165	336	501
2009	298	499	797
2010	330	644	974
2011	274	800	1074
2012	140	906	1046
2013	674	1634	2308
2014	964	1576	2540
2015	710	4686	5396
2016	732	2970	3702
2017	2107	2252	4359
2018	5092	5289	10,381
2019	1083	3023	4106
Grand total	12,766	25,445	38,211

# Appendix B



Figure A1. Spatial autocorrelation with OLS residuals.



Figure A2. Spatial autocorrelation with GWR residuals.

Table A3.	Diagnostic	information	of OLS
-----------	------------	-------------	--------

Diagnose Name	Value	Definition
AIC	1499.96782597000	Akaike's information criterion: a relative measure of performance used to compare models; the smaller AIC indicates the superior model.
AICc	1500.63449263000	Corrected Akaike's information criterion: second-order correction for small sample sizes.
R2	0.39725686606	R-Squared, coefficient of determination: the proportion of variation in the dependent variable that is explained by the model.
AdjR2	0.38443254406	Adjusted R-squared: R-squared adjusted for model complexity (number of variables) as it relates to the data.
F-Stat	30.97683184330	Joint F-statistic value: used to assess overall model significance.
F-Prob	0.00000000000	Joint F-statistic probability (p-value): the probability that none of the explanatory variables have an effect on the dependent variable.
Wald	29.45291722620	Wald statistic: used to assess overall robust model significance.
Wald-Prob	0.00105205020	Wald statistic probability (p-value): the computed probability, using robust standard errors, that none of the explanatory variables have an effect on the dependent variable.
K(BP)	172.61536356700	Koenker's studentized Breusch–Pagan statistic: used to test the reliability of standard error values when heteroskedasticity (nonconstant variance) is present.
K(BP)-Prob	0.00000000000	Koenker (BP) statistic probability (p-value): the probability that heteroskedasticity (nonconstant variance) has not made standard errors unreliable.
JB	73438.08015430000	Jarque–Bera statistic: used to determine whether the residuals deviate from a normal distribution.
JB-Prob	0.00000000000	Jarque–Bera probability (p-value): the probability that the residuals are normally distributed.
Sigma2	1.28891101535	Sigma-squared: OLS estimate of the variance of the error term (residuals).

Diagnostic Name	Value
Neighbors	103.000000000
Residual squares	266.63162555400
Effective number	109.62242880200
Sigma	0.84732104742
AICc	1292.70285212000
R2	0.73470861988
R2Adjusted	0.65711482778

Table A4. Diagnostic information of GWR.

#### References

- 1. van der Borg, J.; Costa, P.; Gotti, G. Tourism in European Heritage Cities. Ann. Tour. Res. 1996, 23, 306–321. [CrossRef]
- García-Hernández, M.; de la Calle-Vaquero, M.; Yubero, C. Cultural Heritage and Urban Tourism: Historic City Centres under Pressure. Sustainability 2017, 9, 1346. [CrossRef]
- 3. van der Zee, E.; Bertocchi, D.; Vanneste, D. Distribution of Tourists within Urban Heritage Destinations: A Hot Spot/Cold Spot Analysis of TripAdvisor Data as Support for Destination Management. *Curr. Issues Tour.* **2018**, 1–22. [CrossRef]
- 4. Ashworth, G.; Page, S.J. Urban Tourism Research: Recent Progress and Current Paradoxes. Tour. Manag. 2011, 32, 1–15. [CrossRef]
- Tu, H.-M. Sustainable Heritage Management: Exploring Dimensions of Pull and Push Factors. Sustainability 2020, 12, 8219. [CrossRef]
- Li, Y.; Lo, R.L.B. Applicability of the Market Appeal—Robusticity Matrix: A Case Study of Heritage Tourism. *Tour. Manag.* 2004, 25, 789–800. [CrossRef]
- Trinh, T.T.; Ryan, C. Visitors to Heritage Sites: Motives and Involvement—A Model and Textual Analysis. J. Travel Res. 2017, 56, 67–80. [CrossRef]
- 8. Song, H.; Liu, H. Predicting Tourist Demand Using Big Data. In *Analytics in Smart Tourism Design*; Xiang, Z., Fesenmaier, D.R., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 13–29. ISBN 978-3-319-44262-4.
- 9. Laney, D. 3D Data Management Controlling Data Volume Velocity and Variety. META Group Res. Note 2001, 6, 1.
- 10. Beyer, M.; Laney, D. *The Importance of "Big Data": A Definition;* Gartner: Stamford, CT, USA, 2012; Available online: https://www.gartner.com/en/documents/2057415/the-importance-of-big-data-a-definition (accessed on 31 December 2020).
- 11. Li, G. Big data related technologies, challenges and future prospects. Inf. Technol. Tourism 2015, 15, 283–285. [CrossRef]
- Karayazi, S.S.; Dane, G.; de Vries, B. Leveraging Newly Available Big Data for Urban Architectural Heritage: Designing a Recommendation System for Heritage Sites through the Lens of Social Media. In *Real Corp 2020: Shaping Urban Change Livable City Regions for the 21st Century, Proceedings of 25th International Conference on Urban Planning, Regional Development and Information Society,* 1st ed.; Schereng, M., Popovich, V.V., Zeile, P., Elisei, P., Beyer, C., Ryser, J., Reicher, C., Eds.; RWTH-Aachen University: Aachen, Germany, 2020; pp. 553–563. ISBN 978-3-9504173-9-5.
- 13. Kádár, B. Measuring Tourist Activities in Cities Using Geotagged Photography. Tour. Geogr. 2014, 16, 88–104. [CrossRef]
- 14. Ginzarly, M.; Pereira Roders, A.; Teller, J. Mapping Historic Urban Landscape Values through Social Media. *J. Cult. Herit.* 2018. [CrossRef]
- Koutras, A.; Nikas, I.A.; Panagopoulos, A. Towards Developing Smart Cities: Evidence from GIS Analysis on Tourists' Behavior Using Social Network Data in the City of Athens. In *Smart Tourism as a Driver for Culture and Sustainability*; Katsoni, V., Segarra-Oña, M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 407–418. ISBN 978-3-030-03909-7.
- Devkota, B.; Miyazaki, H.; Pahari, N. Utilizing User Generated Contents to Describe Tourism Areas of Interest. In Proceedings of the 2019 First International Conference on Smart Technology & Urban Development (STUD), Chiang Mai, Thailand, 13–14 December 2019; pp. 1–6.
- 17. 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]
- 18. Ganzaroli, A.; De Noni, I.; van Baalen, P. Vicious Advice: Analyzing the Impact of TripAdvisor on the Quality of Restaurants as Part of the Cultural Heritage of Venice. *Tour. Manag.* **2017**, *61*, 501–510. [CrossRef]
- 19. Girardin, F.; Vaccari, A.; Gerber, A.; Biderman, A. Quantifying Urban Attractiveness from the Distribution and Density of Digital Footprints. *Int. J. Spat. Data Infrastruct. Res.* **2009**, *4*, 26.
- 20. Gede, M.; Kádár, B. Analysing Tourism Movements along the Danube River Based on Geotagged Flickr Photography. *Proc. Int. Cartogr. Assoc.* 2019, 2, 1–5. [CrossRef]
- 21. Runge, C.A.; Daigle, R.M.; Hausner, V.H. Quantifying Tourism Booms and the Increasing Footprint in the Arctic with Social Media Data. *PLoS ONE* **2020**, *15*, e0227189. [CrossRef]
- 22. Gong, V.X.; Daamen, W.; Bozzon, A.; Hoogendoorn, S.P. Crowd Characterization for Crowd Management Using Social Media Data in City Events. *Travel Behav. Soc.* 2020, 20, 192–212. [CrossRef]
- Dane, G.; Borgers, A.; Kaya, D.I.; Feng, T. Visitor Flows at a Large-Scale Cultural Event: GPS Tracking at Dutch Design Week. *IJGI* 2020, 9, 661. [CrossRef]

- Shoval, N.; Ahas, R. The Use of Tracking Technologies in Tourism Research: The First Decade. *Tour. Geogr.* 2016, 18, 587–606. [CrossRef]
- 25. Duca, L.; Marchetti, A. Open data for tourism: The case of Tourpedia. J. Hosp. Tour. Technol. 2018, 10, 351–368.
- 26. Vong, L.T.-N.; Ung, A. Exploring Critical Factors of Macau's Heritage Tourism: What Heritage Tourists Are Looking for When Visiting the City's Iconic Heritage Sites. *Asia Pac. J. Tour. Res.* **2012**, *17*, 231–245. [CrossRef]
- 27. Dane, G.; Borgers, A.; Feng, T. Subjective Immediate Experiences during Large-Scale Cultural Events in Cities: A Geotagging Experiment. *Sustainability* **2019**, *11*, 5698. [CrossRef]
- 28. 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*, 5. [CrossRef]
- 29. Ministerie van Onderwijs, C. en W. Rijksmonumenten-kaartlagen-Rijksmonumenten register-Rijksdienst voor het Cultureel Erfgoed. Available online: https://www.cultureelerfgoed.nl/onderwerpen/rijksmonumentenregister/rijksmonumenten-kaartlagen (accessed on 31 August 2020).
- 30. Data En Informatie-Amsterdam. Available online: https://data.amsterdam.nl/ (accessed on 31 August 2020).
- 31. Centre, U.W.H. Seventeenth-Century Canal Ring Area of Amsterdam inside the Singelgracht. Available online: https://whc. unesco.org/en/list/1349/ (accessed on 26 December 2020).
- 32. Macqueen, J. Some Methods for Classification and Analysis of Multivariate Observations. Multivar. Obs. 1967, 1, 17.
- 33. Lee, I.; Cai, G.; Lee, K. Exploration of Geo-Tagged Photos through Data Mining Approaches. *Expert Syst. Appl.* **2014**, *41*, 397–405. [CrossRef]
- 34. Dunn, J.C. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters. *Null* **1973**, *3*, 32–57. [CrossRef]
- Mai, D.S.; Ngo, L.T. Semi-Supervised Fuzzy C-Means Clustering for Change Detection from Multispectral Satellite Image. In Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Istanbul, Turkey, 2–5 August 2015; pp. 1–8.
- Winkler, R.; Klawonn, F.; Kruse, R. Problems of Fuzzy c-Means Clustering and Similar Algorithms with High Dimensional Data Sets. In *Challenges at the Interface of Data Analysis, Computer Science, and Optimization*; Gaul, W.A., Geyer-Schulz, A., Schmidt-Thieme, L., Kunze, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 79–87. ISBN 978-3-642-24465-0.
- Kisilevich, S.; Krstajic, M.; Keim, D.; Andrienko, N.; Andrienko, G. Event-Based Analysis of People's Activities and Behavior Using Flickr and Panoramio Geotagged Photo Collections. In Proceedings of the 2010 14th International Conference Information Visualisation, London, UK, 26–29 July 2010; pp. 289–296.
- Ester, M.; Kriegel, H.-P.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
- 39. Batra Nagpal, P.; Ahlawat Mann, P. Comparative Study of Density Based Clustering Algorithms. *Int. J. Comput. Appl.* **2011**, 27, 44–47. [CrossRef]
- 40. Götz, M.; Bodenstein, C.; Riedel, M. HPDBSCAN: Highly Parallel DBSCAN. In Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments-MLHPC '15, Austin, TX, USA, 15–20 November 2015; 2015; pp. 1–10.
- Kempiak, J.; Hollywood, L.; Bolan, P.; McMahon-Beattie, U. The Heritage Tourist: An Understanding of the Visitor Experience at Heritage Attractions. *Int. J. Herit. Stud.* 2017, 23, 375–392. [CrossRef]
- 42. Brunsdon, C.; Fotheringham, S.; Charlton, M. Geographically Weighted Regression-Modelling Spatial Non-Stationarity. J. R. Stat. Soc. Ser. D 1998, 47, 431–443. [CrossRef]
- 43. Wooldridge, J.M. Introductory Econometrics: A Modern Approach. *Nelson Educ.* **2012**, *910*, 94–99.
- Statistiek, C.B. voor de Wijk-En Buurtkaart. 2020. Available online: https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/ geografische-data/wijk-en-buurtkaart-2020 (accessed on 3 November 2020).
- 45. Moran, P.A.P. Notes on Continuous Stochastic Phenomena. Biometrika 1950, 37, 7. [CrossRef]
- 46. Fu, W.J.; Jiang, P.K.; Zhou, G.M.; Zhao, K.L. Using Moran's I and GIS to Study the Spatial Pattern of Forest Litter Carbon Density in a Subtropical Region of Southeastern China. *Biogeosciences* **2014**, *11*, 2401–2409. [CrossRef]
- 47. How OLS Regression Works—ArcGIS Pro | Documentation. Available online: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/how-ols-regression-works.htm (accessed on 19 November 2020).
- 48. Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. In *Selected Papers of Hirotugu Akaike;* Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer: New York, NY, USA, 1998; pp. 199–213. ISBN 978-1-4612-1694-0.