Using accommodation price determinants to segment tourist areas

Juan M. Hernández, Jacques Bulchand-Gidumal, Rafael Suárez-Vega

University of Las Palmas de Gran Canaria (ULPGC)

To cite

Hernández, J. M., Bulchand-Gidumal, J., & Suárez-Vega, R. (2021). Using accommodation price determinants to segment tourist areas. Journal of Destination Marketing & Management, 21, 100622. https://doi.org/10.1016/j.jdmm.2021.100622

Abstract

Accommodation services oriented to different tourist segments usually have different price determinants. Thus, in multi-facet destinations such as large regions or cities, it should be possible to find and describe the underlying types of tourism in the destination by using a price determinant analysis. In this paper, a methodology based on stepwise geographically weighted regression (GWR) is developed, using a *k*-means clustering algorithm to determine the different types of tourism existing in a large geographical area. The method is applied to the island of Gran Canaria (Canary Islands, Spain), using a database of more than 2000 peer-to-peer accommodation units spread over the geography of the island. As a result, it was possible to identify and classify eight different clusters of types of tourism within this geographical area. This methodology can be used in other geographical areas to identify the different types of tourism developed in them.

Keywords: GWR, Airbnb, peer-to-peer accommodation, segmentation, hedonic pricing

1 Introduction

As with any other good or service, there are a series of determinants of the price of tourism products and accommodation has been one of the most extensively researched tourism services (e.g. Zhang, Ye, & Law, 2011). For example, in the case of hotels, the most common price determinants are, among many others, the stars of the hotel, the town in which the hotel is placed, the size of the property, the distance to certain resources (e.g. beach) and the availability of parking places (Espinet, Saez, Coenders, & Fluvia, 2003; Papatheodorou, 2002; Thrane, 2005). Identification of these price determinants is crucial for optimal hospitality management (Gibbs, Guttentag, Gretzel, Morton, & Goodwill, 2018).

Interestingly, the body of knowledge regarding price determinants of accommodation units has found varied and even, in some cases, contradictory results. The main reason for this is that the price determinants depend on the type of tourist area analyzed. For example, for a sun and beach hotel, some of the price determinants that have been found are all-inclusive service, distance to the beach and availability of a swimming pool (Soler, Gemar, Correia, & Serra, 2019; X. Wang, Sun, & Wen, 2019). In contrast, for a city hotel, the price determinants that were found include availability of mini-bar, hairdryer, free parking and location (Pawlicz & Napierala, 2017; Thrane, 2007; Valentin & O'Neill, 2019). The hypothesis that different market segments have different price determinants has also been extensively tested in the literature for different types of accommodation (e.g. Martín, Román, & Mendoza, 2018, for self-catering accommodation; ; Moreno-Izquierdo, Ramón-Rodríguez, Such-Devesa, & Perles-Ribes, 2019, for sharing accommodation units in urban and sun and beach destinations; Soler et al., 2019, for hotels in the Algarve).

Thus, segmentation of the tourism that is taking place in a certain area is important to properly identify the price determinants. In fact, most of the available studies focus on geographical areas in which the tourism segment is clear (e.g. a ski zone, a sun and beach destination, an urban area). However, with the rise of new types of tourism (e.g. urban short breaks) and new types of tourism services (e.g. sharing economy services), it is common to find geographical areas in which several tourism segments coexist and for which spatial identification of the existing segments and their characteristics is not an easy task.

Therefore, the objective of this research is to develop a methodology that will use a large pool of attributes of accommodation units in order to determine the latent tourism areas that exist in a region and the price determinants of the units in each of those areas. These price determinants will help in the process of identifying the types of tourism that are taking place in the region. More specifically, peer-to-peer (P2P) accommodation units are used, due to the large number of attributes available for each of the units.

For this purpose, a methodology based on geographically weighted regression (GWR) is applied to the case of P2P accommodation units on the island of Gran Canaria (Canary Islands, Spain). Gran Canaria offers several types of tourism, including sun and beach, rural and urban. The results show eight different tourism areas in the island and reveal a finer classification than the three that are commonly observed.

2 Literature review

2.1 Price determinants and tourism segmentation

Interest in tourism segmentation has grown in parallel to that in market segmentation (Dolničar, 2004). The possibility of segmenting tourists by dividing them into homogenous groups provides several benefits to the different stakeholders in the tourism sector (Bigné, Gnoth, & Andreu, 2007). Tourism segmentation has usually been carried out using tourist data, whether primary or secondary (Tkaczynski, Rundle-Thiele, & Beaumont, 2009). Different authors have used different criteria to segment tourists, including, amongst others, country of origin, socio-economic variables, demographic variables and psychographic segmentation. Generally speaking, tourist grouping can take place *a priori* (also called commonsense segmentation) and *a posteriori* (also called *post-hoc* or data-driven segmentation) (Dolnicar, Lazarevski, & Yanamandram, 2013). While in *a priori* segmentation the variables used are demographics, geography, seasons and intentions, among others, in *a posteriori* processes the segmentation takes place using, for example, travel motivations and activities carried out (Dolnicar et al., 2013; Tkaczynski, Rundle-Thiele, & Prebensen, 2015).

These segmentation processes have led to the appearance of certain segment tags that are widely used in the literature. One of the most frequently used is the one that segments tourists according to the main interest, attractions and motivation for the trip (Huertas & Marine-Roig, 2015): nature/rural, tangible heritage, intangible heritage, urban, gastronomy, leisure, sun and beach, business (including the events industry), sports, technology and services.

One of the main uses of tourism segmentation for researchers is to help find price determinants for different types of services. It seems natural that the price determinants for the services oriented to each segment will be different. For example, Thrane (2007) found the availability of mini-bar, hairdryer, parking and distance to city center to be among the price determinants of urban hotels. Soler et al. (2019) found that category and reputational variables were the most relevant in sun and beach hotels in the Algarve. Falk (2008) found that ski pass price was determined by length of ski runs, transport capacity, proportion of model transportation means and natural conditions, among others. Zhang et al. (2011) examined service attributes, and found that in an urban destination (in this case, New York) the characteristics that determined price depended on the category of the hotel (economy, midscale, luxury). Wang & Nicolau (2017) analyzed the case of sharing accommodation units in 33 cities in the US and found 24 variables that were predictors of price.

The aforementioned studies and, in general, studies that analyze price determinants tend to firstly categorize the type of tourism (e.g. urban, sun and beach, rural or ski) using the services that are being considered. However, this is not always a simple task. For example, many different types of tourism may be concurrently taking place in an urban area (e.g. sun and beach, shopping, cultural, gastronomic). In such cases, that is when several types of tourism coexist in a geographical area, it would be useful to carry out a segmentation of the tourism based on the attributes of the services that are available.

The spatial heterogeneity of some price determinants can help segmentation studies in destinations that include several types of tourism. The geographic weighted regression (GWR) (Brunsdon, Fotheringham, & Charlton, 1996), which is an extension of the ordinary least squares (OLS) procedure allowing spatial heterogeneity in the coefficient estimates, is a useful technique in this regard. Some price determinant studies have already applied this technique (Kim, Jang, Kang, & Kim, 2020; Latinopoulos, 2018; Lee, Jang, & Kim, 2020; Soler & Gemar, 2018; H. Zhang, Zhang, Lu, Cheng, & Zhang, 2011), and the estimation results commonly improve those obtained with methods which do not take into account spatial effects. More specifically, Kim et al. (2020) used GWR to describe the effect of size, category and location attributes in several hotels in Chicago, previously classified by chain and location. Lee et al. (2020) analyzed the influence of tourism clusters on the tourist performance, where clusters were previously defined according to the rate of specialization in a tourism sub-sector in a certain region of the Florida State. The results allowed the identification of clusters with higher influence on revenue per room.

The studies above start from pre-determined market segments (hotel types, location, rate of specialization) and estimate the influence of several attributes on the rental price of accommodation units in them. An exception is the study of Latinopoulos (2018), who used GWR estimation results in a coastal destination in Greece to determine areas or segments where tourist preferences are homogeneous (e.g. sea-view zone, integrated environmental zone). In this paper, a further step in this direction is proposed by combining the results of GWR with clustering analysis to perform a segmentation analysis.

2.2 Price determinants in peer-to-peer accommodation

In this study, Airbnb listings are used to conduct a market segmentation analysis. To help the selection of potential attributes, this section reviews the price determinants commonly described in the P2P accommodation industry. Some of these determinants have been previously identified in hotels (e.g. Espinet et al. 2003). However, the nature of P2P, where the client does not usually know in advance the service provider, means that new factors may influence the final price.

There has been growing interest in analyzing the price determinants of sharing accommodation units. Three reasons explain this growth. First, the recent expansion of P2P accommodation (Bakker & Twining-Ward, 2018), led most notably by Airbnb. Second, the fact that each unit offered in Airbnb is different to the rest, thus opening the door to the potential analysis of more than 7 million units worldwide (Airbnb, 2020) and their characteristics. Third, the fact that the information about these units (their characteristics and their prices) is available online, directly from the Airbnb page, and from providers that sell extensive datasets (e.g. AirDNA, Inside Airbnb).

Due to its recent implementation, most of the studies of price determinants in the P2P accommodation industry have been published in the last few years (2018-2020). Based on the literature review, the determinants can be classified in five groups:

1) Structural attributes (S): This group encompasses the property characteristics (e.g. property type, size, number of bedrooms), amenities (e.g. pool, hot tub, parking) and services (e.g. breakfast, kitchen, doorman). One of the differences between Airbnb accommodations and hotels is that the former is not standardized (in fact many service providers offer hosting in their own homes). Therefore, property attributes in Airbnb display a far greater degree of heterogeneity than hotels. As an illustrative example, up to 79 different property types were identified by Perez-Sanchez, Serrano-Estrada, Marti, & Mora-Garcia et al. (2018) in the region of Valencia (Spain). In general, most of the studies show that structural attributes determine the largest percentage of the rental price. Among them, the most frequent and most influential price determinants are the property type, the number of bedrooms and bathrooms of the property and the number of guests it can accommodate.

2) Host attributes (H): In the absence of direct contact, Airbnb buyers make use of the indications the platform offers to inform themselves on questions related to trust, quality and service 'promise' (Xie & Mao, 2017). One such indication is the information provided about the host's personal and relational attributes. For example, some studies have found that publishing personal photos and the gender of the host has an influence on trust and therefore on prices (Ert, Fleischer, & Magen, 2016). Additionally, Airbnb gives a specific label (superhost) to hosts that have met certain managerial criteria, which some studies have found has a positive effect on prices (Benítez-

Aurioles, 2018; Lorde, Jacob, & Weekes, 2019; Önder, Weismayer, & Gunter, 2019; Wang & Nicolau, 2017). The length of time a host has been registered in Airbnb also has a positive effect (e.g. Cai, Zhou, Ma, & Scott, 2019; Lorde, Jacob, & Weekes, 2019; Perez-Sanchez et al., 2018). Other host attributes that influence price include whether the host manages several properties or not, which is considered an indicator of professionalism (Chen & Xie, 2017; Gibbs et al., 2018; Magno, Cassia, & Ugolini, 2018). Previous studies have found that professional hosts tend to charge higher prices than amateur hosts (Bulchand-Gidumal, Melián-González, & López-Valcárcel, 2019).

3) Management attributes (M): This group includes the specific rental rules chosen by the host, such as the availability of instant booking, response time and flexible cancellation, among others. Some papers have found a significant effect of these factors on rental price (Benítez-Aurioles, 2018; Chen & Xie, 2017; Gibbs et al., 2018; Lorde et al., 2019; Wang & Nicolau, 2017). Other studies have found that rental price is positively influenced by the number of photographs included in the property advertisement (Gibbs et al., 2018; Lorde, Jacob, & Weekes, 2019b; Perez-Sanchez et al., 2018), which can be considered a management factor as well.

4) Reputation attributes (R): Reputation is an antecedent of trust and therefore plays an essential role in the purchase decision. As in the case of hotels, several reputation indicators are presented in Airbnb, such as reviews and rating (in quantitative and qualitative terms). Most of the studies that have been conducted report that these indicators exert a significant influence on price. More specifically, high rating scores tend to have a positive influence on rental price (Moreno-Izquierdo et al., 2019; Teubner, Hawlitschek, & Dann, 2017; Tong & Gunter, 2020). Importantly, a high number of reviews has a negative influence on price in most of the studies (Benítez-Aurioles, 2018; Gibbs et al., 2018; Magno et al., 2018; Wang & Nicolau, 2017). This phenomenon has been attributed to the strong demand for cheap properties, which in turn results in them having a large number of reviews (Gibbs et al., 2018).

5) Location attributes (L): This category includes spatial (e.g. distance to the city center and other attractions) and environmental factors (e.g. socioeconomic characteristics in the surroundings and competition among providers). Since Airbnb properties are located closer than hotels to the main attractions in some cities (Gutiérrez, García-

Palomares, Romanillos, & Salas-Olmedo, 2017), it is expected that spatial factors will be important in determining rental price in this market. In fact, several studies have found a significant influence on price of distance to attractions (Cai, Zhou, Ma, & Scott, 2019b; Gibbs et al., 2018; Önder et al., 2019; Perez-Sanchez et al., 2018; Tong & Gunter, 2020; Zhihua Zhang, Chen, Han, & Yang, 2017). Environmental factors, including population density, GDP/average income and number of pedestrian routes in the surroundings, have also been found to be price determinants (Chen & Xie, 2017; Chica-Olmo, González-Morales, & Zafra-Gómez, 2020; Önder et al., 2019; Tang, Kim, & Wang, 2019; Teubner et al., 2017), together with other indicators of competition among properties, such as the number of Airbnb properties and hotels in the neighborhood (Chen & Xie, 2017; Tang et al., 2019).

As a summary, Table 1 presents a review of the methodology, area, number, type of price determinant and the major findings of the empirical studies conducted to date on properties offered in Airbnb. It can be deduced from Table 1 that price determinants are commonly analyzed for Airbnb listings in urban destinations. Very few studies have focused on areas that include other types of destinations. Among them, Lorde et al. (2019) analyzed twelve Caribbean countries, finding specific price determinants such as the exchange rate. More related to the objectives of this paper are the studies by Falk, Larpin, & Scaglione (2019) and Moreno-Izquierdo et al. (2019) on properties in different tourist areas in Switzerland and the region of Valencia (Spain), respectively. Unlike the other studies, the first study (Falk et al., 2019) split the sample between urban and rural destinations and found that some structural factors, such as having a sauna and hot tub, exclusively influence the price of units in rural destinations, whereas property type, views and old town location influence the price of units in urban destinations. Similarly, Moreno-Izquierdo et al. (2019) split their study area into two segments, in this case urban and holiday destinations. They found that factors such as population density, income and tourist season had a greater influence in holiday than urban destinations.

[TABLE 1 ABOUT HERE]

As was the case of the contributions reviewed in section 2.1, the tourist market segments in the studies above were defined in advance and the analysis undertaken determined the effect of a price determinant in each segment. In contrast, in this paper tourist areas are classified from the observation and classification of the spatial heterogeneity of factors influencing price, without following a strict pre-classification of the market segments. In the case study, this technique is applied to a region (island) that includes urban, holiday and rural areas in order to identify the different types of tourism developed in it through the results of the price determinant analyses.

3 Methodology

3.1 Case study and data collection

Airbnb listings for Gran Canaria were used in this case study. Although this island has traditionally been a sun and beach tourist destination, tourism in urban and rural areas has significantly increased since the turn of the present century. The Canary Islands is the third highest region of Spain in terms of tourist flow. More specifically, Gran Canaria received 4.2 M visitors in 2018, 84.8% of them foreigners (FRONTUR, 2020), mostly British and German.

The three main types of tourism (sun and beach, urban, and rural) can be roughly spatially located (Fig. 1). First, the southern part of the island, where the larger beaches are, has been almost exclusively used for sun and beach tourism since the 1960s. It is therefore a mature coastal destination which includes different subtypes of sun and beach tourism (e.g. youngsters looking for nightlife, mature adults, retirees who stay for longer periods of time, gay) (Domínguez-Mújica, González-Pèrez, & Parreño-Castellano, 2011; Melián-González, Moreno-Gil, & Araña, 2011). Second, the main town (Las Palmas de Gran Canaria, c. 400,000 inhabitants), which is located in the northeast of the island (Fig. 1). While this city was a popular tourist destination in the 1960s and 1970s, it fell out of favor in the following two decades before seeing a resurgence which has continued unabated to the present day. Finally, tourism in rural areas of the center and north of the island has also gained increased importance, receiving mainly locals but also foreigners. However, this *a priori* classification of the three types of tourism does not take into account possible tourism-related specificities in different parts of each area due to geographical and/or socioeconomic characteristics and the interaction effect among the different destinations.

[FIGURE 1 ABOUT HERE]

In this study, all Airbnb listings in Gran Canaria and their attributes were considered. For the sake of homogeneity, the analysis is limited to entire properties and excluded private and shared rooms. Only properties that included at least one review were considered. In total, the information on 2238 units was downloaded in January 2018, with 140 possible price determinants included in the property information in the Airbnb platform. The determinants were mainly structural attributes (90), but also included host (13), management (27), reputational (8) and location (2) attributes. An additional 15 location attributes were calculated using geographic information systems (GIS). A variable showing the location of the property with respect to the main points of interest was generated from 206,897 pictures in Flickr in Gran Canaria from 2005 to March 2018 (Eugenio-Martin, Cazorla-Artiles, & González-Martel, 2019). The number of pictures in a certain location is assumed to be an indication of visitors' interest for that point. Other environmental attributes are also incorporated, such as distances to different cities. The list, description and descriptive statistics of these variables can be found in the Supplementary Material.

3.2 Geographically weighted regression (GWR) model

As Hong & Yoo (2020) pointed out, Airbnb pricing is spatially dependent because host strategies will depend on and at the same time affect other hosts nearby. Therefore, global regression models are not the most appropriate when prices are spatially dependent, and the use of models that take the geographical aspect into account is necessary. The existence of a spatial structure in the data in a regression model can be suggested by the presence of a spatial autocorrelation in the dependent variable or in the model residuals.

One of the statistics used to measure spatial autocorrelation is the Moran's I (Moran, 1950), which is given by the following expression:

$$I = \frac{n}{T}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2},$$

where z_i is the deviation of the variable from its mean, w_{ij} is the spatial weight between individuals *i* and *j*, n represents the total number of individuals, and *T* is the aggregated spatial weight:

$$T = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}.$$

Different functions can be used to obtain weights w_{ij}. In particular, in this study the adaptive Gaussian kernel was considered.

Moran's I index varies between -1 and +1, such that a positive value indicates clustering while a negative one means dispersion. These statistics can be tested where the null hypothesis is that the values are randomly distributed. The sign of Moran's I is only interpreted when the test is significant.

In this study, the price determinants were estimated by means of a spatial regression model. Each individual in the model corresponds to an entire geographically located property. The regression allows local estimation of the attributes that influence the price of each property in the sample. In other words, the price determinants can be heterogeneously distributed throughout the space. The GWR model, initially proposed by Brunsdon, Fotheringham, & Charlton (1996), assumes that the effect of explanatory variables can be spatially dependent, implicitly including the geographic coordinates in the model definition.

In the GWR model, each property *i* is associated to a pair (u_i, v_i) representing the geographic coordinates of the property. Given *n* properties and a set of determinants $\{x_1, x_2, \ldots, x_K\}$ (covariates or independent variables), the model assumes that the rental price *p* (dependent variable) can be expressed as:

$$p_i(u_i, v_i) = \beta_0(u_i, v_i) + \sum_{k=1}^{K} \beta_k(u_i, v_i) x_{ik} + \varepsilon_i, \; i = 1, \ldots, n, \quad (1)$$

where $\beta_k(u_i, v_i)$ is the estimated coefficient for variable x_k associated to the i^{th} property, and ε_i is the error term in regression at the coordinates (u_i, v_i), which is independently normally distributed with mean zero and constant variance.

The conceptual basis of the GWR model is that closer observations in the data sample are more influential on the estimation than more distant ones. The spatial relationship between observations is defined by means of a diagonal weight matrix, $W^i_{nxn} = diag(w_1(u_i, v_i), w_2(u_i, v_i), \dots, w_n(u_i, v_i))$, where the components depend on the coordinates of each property. The coefficient estimation procedure is similar to that followed for weighted least squares models, but in GWR each individual in the sample has associated its own weight matrix and the coefficients are estimated for each one of them (Fotheringham, Brunsdon, & Charlton, 2002). The diagonal elements in this matrix are calculated using a distance-decreasing kernel function. In this paper the Gaussian kernel is used:

$$w_j(u_i, v_i) = e^{-0.5\left(\frac{d_{ji}}{h}\right)^2},$$

where d_{ji} is the Euclidean distance between locations (u_i, v_i) and (u_j, v_j), and h is called the bandwidth (measured in the same units as the distance). The bandwidth mission is to determine the influence area as well as the effect of the distance on the estimation. Although the kernel choice does not usually affect significantly the estimates, the bandwidth selection does. The bandwidth may be fixed or adaptive. In the fixed option, the same bandwidth is considered for all observations in the sample. In the adaptive form, the bandwidth varies to obtain similar subsamples for all observations. Adaptive bandwidth is commonly expressed as the k nearest neighbours and h represents the distance to the k^{th} nearest neighbour. The fixed bandwidth is recommended when the observations are uniformly distributed in the space, otherwise the adaptive form is preferred. See (Páez, Uchida, & Miyamoto, 2002a, 2002b) for a more detailed discussion about the bandwidth selection in GWR models.

In this paper, the corrected Akaike Information Criterion proposed by Fotheringham, Brunsdon, & Charlton (2002) to test the goodness of fit for the GWR models is followed:

$$AICc = 2n\log_e(\hat{\sigma}) + n\log_e(2\pi) + n\left\{\frac{n + tr(S)}{n - 2 - tr(S)}\right\},$$

where $\hat{\sigma}$ is the estimated standard deviation of the error term. The lower the AICc, the better the model fit. In addition, a minimum difference of three units in AICc is chosen to confirm that one model is a better fit of the data than another.

Therefore, the GWR model involves *n* local coefficient estimations. These local regressions usually share sample elements, leading to an artificial increase of the t-statistical values obtained to test parameter significance. To prevent this problem, da Silva & Fotheringham (2016) proposed the following corrected significance level (α) for the estimates:

$$\alpha = \frac{\xi_m}{\frac{p_e}{K}},$$

where ξ_m is the desired significance level for the estimations, p_e is the effective number of parameters $p_e = 2tr(S) - tr(S'S)$, with *S* the hat matrix such that $\hat{p} = Sp$, with $\hat{p}$ being the estimated values for *p*, and *K* is the number of parameters in each model.

When considering a large set of possible covariates, as GWR involves multiple coefficient estimations, it is necessary to make use of heuristic procedures to find the combination of them that best fits the data. To solving the variable selection problem in the GWR context, the recently proposed SW-GWR procedure by Suárez-Vega & Hernández (2020) was followed. Essentially, this procedure proposes the use of a stepwise (SW) algorithm over the total set of variables, in such a way that the variable with the best goodness of fit is added to the model in a sequential manner. The procedure stops when a certain condition is fulfilled. More specifically, the AICc is used as goodness of fit measure and the algorithm stops execution when a minimal improvement in the AICc is not obtained after a given number of steps.

In order to contrast the improvement in the adjustment produced by the GWR with respect to the OLS, Leung, Mei, & Zhang (2000) propose two tests (F1 and F2) in which the null hypothesis is that the GWR does not improve the adjustments of the OLS. Those authors also propose another test (F3) in which they contrast the variability of each of the parameters involved in the model. If it is significant, it means that the model behaves better if the parameter is considered variable instead of fixed.

Multicollinearity may also be a problem when linear regression is considered because it can produce problems for computing the estimation (existence of singular or near singular matrices) and for checking the significance of the parameter. To evaluate multicollinearity, the variance inflation factor (VIF) (for global models), or the local version (for GWR models), can be used. This value is calculated for each variable in the model. In the local version, it is calculated for each variable for each of the local models and therefore an nxK matrix of VIFs is obtained. Values greater than 10 may indicate a multicollinearity problem for that variable in the corresponding model.

3.3 The clustering method

The estimates of the GWR model in this paper show how the effect on price of each determinant varies throughout the space. On the basis of these results, a clustering method is applied to identify groups of significant factors, which are associated to different areas of the studied region. The GWR method has been used in several works to perform clustering. Among others, Cahill, Mulligan, Cahill, & Mulligan (2007) and Windle, Rose, Devillers, & Fortin (2010) can be mentioned. Cahill et al. (2007) used hierarchical clustering to detect clusters in the parameters that describe criminality ratios in Portland, Oregon. For their part, Windle et al. (2010) used a *k*-means cluster analysis based on GWR *t*-values in a context of fisheries in the Northwest Atlantic. In each resulting group they calculated the values of the parameters by averaging the values of the parameters of the individuals that belong to each of the clusters.

In this paper, the *k*-means clustering procedure (MacQueen, 1967) is also applied, which aims to divide the set of multivariate points into a predetermined number of *k* clusters in such a way that the distance between the points belonging to each cluster is minimized. The solution procedure proposed by Hartigan & Wong (1979) is followed, where the *k* clusters are found by minimizing the total within-cluster sum of squared distances. Since the solution depends on the initial points assigned to cluster centers, the algorithm is solved several times starting from random seeds in order to obtain a robust solution.

The cluster center can be interpreted as the representative point belonging to that cluster. In the context of this study, the center's components represent the effect of the determinants on the price of the properties belonging to that group.

Thus, it is used the approach of determining clusters of properties where their attributes had similar effects on price, obtaining a more clearly defined differentiation of the geographical areas in which a particular type of tourism is developed. The number k of clusters must be chosen *a priori*. The procedure to determine k is essential since it determines the number of market segments in the destination. The popular procedure of observing the inertia plot was chosen to find the best k. Given k clusters, the inertia is the sum of squared distances of sample points to the nearest cluster center. The inertia tends to zero as k increases, and the best k is selected when the inertia reduction is significantly lower than the previous ones (Thorndike, 1953).

The methodology described above was applied to find the determinants of the rental price of Airbnb listings in Gran Canaria. The logarithm of price, as usual in this type of studies (Cai et al., 2019; Chica-Olmo et al., 2020; Soler & Gemar, 2018), was considered as the dependent variable.

4 Results

4.1 GWR model estimates

Moran's I index revealed the existence of a significant positive spatial dependence for the logarithm of prices (I = 0. 088, p-value= 2.2e-16), meaning that prices for houses in the vicinity tend to be close. This spatial dependence suggested the need to use a spatial regression model as the GWR. Therefore, the SW-GWR algorithm was run over the set of 156 possible determinants. A new determinant was added to the model in each step. Figure 2 shows the AICc increments along the execution of the algorithm. The stopping rule was met when the AICc was reduced by less than 3 points in three consecutive steps (Suárez-Vega & Hernández, 2020). In Figure 2, steps for which the AICc reduction is less than this limit are those over the horizontal line. Once the algorithm had stopped and the increments analyzed, it was decided to choose the model with the first 30 variables because from that point onwards the successive increments did not provide significant continuous improvements.

[FIGURE 2 ABOUT HERE]

Table 2 shows these 30 variables. The order of the variables in the Table coincides with the order of their entry in the model following the SW-GWR procedure. The earlier the variables enter the model the more they contribute to price explanation. The last two columns in Table 2 represent the AICc obtained and the adjusted R^2 with a model including the first *r* variables, with *r* being the row in the table.

[TABLE 2 ABOUT HERE]

The local VIFs analysis for the model revealed four variables with severe spatial collinearity problems (some local VIFs higher than 10). In order to avoid these problems, the variables corresponding to the number of properties within 500m, beach distance and beach distance at 200m-500m were removed from the model. The new 27-variable model does not present severe collinearity problems, since only the variable corresponding to the number of bedrooms includes VIF values above 5. More specifically, it includes 106 cases in which the local VIF has values between 5 and 5.71. These 106 cases represent 0.17% of all VIFs calculated, which amounted to a total of 60,426 (2238 regressions with 27 variables each).

The descriptive statistics of the significant coefficients of the final model are also shown in Table 2. The value of the estimated coefficients of each determinant varies across the sample space. Columns 3 to 7 in Table 2 show the mean and dispersion measures of the significant coefficients, as well as the percentage of properties for which this factor is significant. This table also shows the p-values for the F1, F2 and F3 tests. The p-values for the F1-test and F2-test show that the GWR is a better fit than the OLS model. This is also confirmed by the adjusted R^2 of the GWR model (0.651) and the OLS model (0.558). The Table 2 also reveals that 20 out of 28 parameters (including the intercept) are significant at the 5% level for the F3-test, confirming that a local regression model is more adequate than a global one. Moreover, if the residual spatial autocorrelation is analyzed, it is still significant in the OLS model (I=0.016, p-value= 2.2E-16) whereas in the GWR model it is no longer significant (I=-5.83E-04, p-value=0.550). This indicates that the use of the GWR solves the problem of spatial dependence of the model errors.

The model includes 15 structural variables. Four of them (number of bathrooms, number of bedrooms, pool availability and dryer availability) are significant for more than half of the sample. Specifically, the number of bathrooms and the number of bedrooms are proxies for the size of the unit, thus indicating that the bigger the size, the higher the price. Certain host attributes were also found to have an influence on price, such as the number of properties the host manages (indicator of professionalism). These five factors explain around 60% of the data variance, as shown by the adjusted R^2. The rest of the 22 significant determinants add only 4% more to explain the data variance. Among them are management, reputation and location attributes.

As column 7 in Table 2 shows, the number of properties for which each determinant is significant varies. For instance, the number of bathrooms is significant for 88.78% of properties while the French-speaking host is only significant for 2.41%. Interestingly, suitability for events enters in the first steps of the procedure, but is only significant for 3.93% of the sample, suggesting that this factor influences a specific and small market segment.

In order to gain further insight about the spatial heterogeneity of the estimates, Figure 3 shows the coefficient values for a representative sample of the significant factors. As can be observed, the number of bathrooms and the existence of a pool affect the rental prices of properties throughout the island, but to a different degree depending on the area. The accommodation market in certain coastal areas (south) values the number of bathrooms more than others (northwest). In contrast, swimming pools are more appreciated in the rural areas in the northeast than on the coast. Suitability for events only influences some properties in the northwest, but is not significant for the rest. The local regression model for each of the properties has a different R^2 depending on where the property is located. The Local R^2 distribution presented in Figure 3 shows that properties in the south of the island are better explained by the explanatory factors than those in the center and northwest.

[FIGURE 3 ABOUT HERE]

4.2 Clusters in the GWR estimates

The findings described above revealed a diverse valuation of the P2P accommodation attributes throughout space. Following the methodology described in section 3.3, Figure 4 was used to determine the number of clusters. Figure 4 shows the inertia graph for the data sample of coefficients. The elbow is the point from which the reductions start to be non-significant (the inertia tends to flatten). In this specific case, the reductions in inertia for k= 7, 8 and 9 are 11.9, 12.7 and 6.6, respectively. Thus, it was considered that *k*=8 is where the trend breaks (the improvements for *k*=7 and *k*=8 are similar, but much higher than for *k*=9). This shows that 8 is the best number of clusters selected in the algorithm.

[FIGURE 4 ABOUT HERE]

Figure 5 shows the spatial distribution of the eight groups of coefficients. As can be seen, the data groups coincide with geographical areas, representing different tourist zones in terms of orography, weather, socioeconomic and tourism supply conditions. These results agree with the previous classification of three types of tourism in the island (urban, rural and beach tourism), but interestingly disaggregate them and find different clusters in each type, two for each of urban and beach tourism and four for rural tourism. This finer classification reveals new sub-types of tourism not previously identified.

[FIGURE 5 ABOUT HERE]

In order to interpret the different groups, Figure 6 shows the center values of the coefficients of attributes in each cluster, ordered by type of determinant. The center values are selected as the representative effect of the determinants in the corresponding group. Figure 6 therefore shows the relative effect of the different determinants on the Airbnb properties belonging to each group. As can be observed, there are marked differences between clusters in some of the factors.

[FIGURE 6 ABOUT HERE]

In order to better visualize the effect of the variables in each cluster, the percentages of Figure 6 were discretized into three categories (low, mid and high), according to the relative position of each coefficient with respect to the maximum value (minimum value in case of negative sign) in all clusters. Thus, a coefficient value is marked as low if it is below one third of the maximum value of the coefficient across all coefficients, mid if it is between one and two thirds of the maximum value and high in the rest of the cases. The results are shown in Table 3. Where no category is assigned the variable does not exert any effect on the cluster. Table 3 also shows the number of cases in each of the clusters.

[TABLE 3 ABOUT HERE]

Some general deductions can be made from Figure 6 and Table 3. For example, the most relevant structural variables (number of bathrooms and bedrooms, availability of swimming pool) influence the rental price in almost all the clusters. However, suitable for events is exclusively appreciated in northern rural tourism clusters (R-NW and R-NE). The different weather conditions in the north (humid), south (dry), along the coast (warm) and in the center (cold) determine different valuations of some facilities, such as dryer machine, hot tub and heating system. Some other structural variables (essentials and elevator availability) show negative effects mainly in some rural clusters, suggesting that units which mention these items are likely to be low-quality units that need to specify this fact. This reason may also apply to the highly negative influence of TV availability in one of the beach tourism clusters (B-SW).

The number of properties the host manages influences more in B-SW, but also in both urban clusters to a lower extent. The language capabilities have a positive effect on rental price mostly in R-SW (French) and R-NW (German). Among the reputation attributes influencing rental price, cleanliness rating has a positive effect mainly in U-NB and location rating in R-NE and R-NW.

As for the location variables, the number of points of interest near the property influences almost exclusively urban-beach tourism and the rural tourism cluster in the surroundings of the main city (U-B and R-NE), being low or negligible for the rest. The negative influence of this variable in the rental prices in cluster U-B is explained by a competition effect, since the points of interest are located in the surroundings of the properties. The distance to airport shows positive (U-NB) and negative (R-NW) effects as well. Since the airport is located in the southeast, it is clear that the distance to the airport is positively valued in properties close to the airport but negatively valued by properties far away. The effect of having a beach closer than 500 m is positive in almost all clusters.

The analysis above shows that each cluster is influenced by a specific group of determinants. In order to summarize the combined effect of them, below each of the clusters and the variables that influence them were reviewed and catalogued, taking into account the geographical location (Fig. 5) and the characteristics of the determinants in each cluster (Fig. 6 and Table 3).

Cluster U-B: Urban tourism, beach area. Land cost is high in this area, so the size of the property is highly appreciated. The market is a mature one with a highly demanding customer base, as shown by the importance of the competition effect.

Cluster U-NB: Urban tourism, non-beach area. Property size is also relevant in this cluster, where the market is professionalized with high values for cleanliness rating. Unlike the U-B cluster, swimming pool availability is highly appreciated.

Cluster B-SW: Beach tourism in low density area. Property size and professionalism are positively valued. Distance to the beach is appreciated, since properties are widespread distributed in this low-density area.

Cluster B-S: Beach tourism in high density area. This cluster has the highest positive effect of the number of bathrooms and bedrooms, revealing the high cost of land. Unlike to cluster B-SW, distance to the beach is not an issue in this cluster since this is quite small and high density urban area.

Cluster R-SW: Mature coastal rural tourism. Some luxury amenities, such as a hot tub, are appreciated in this cluster. Moreover, the market is professional and competitive. Host language skills are also appreciated, showing that these properties are sought after by foreigners.

Cluster R-NW: Non-mature remote rural tourism. Properties that are suitable for events are highly appreciated in this cluster, showing that, unlike the other clusters, properties here are usually large houses that can offer these characteristics. Distance to the airport is negatively valued. The effect of professionalism is low, revealing that this cluster is still in the first stages of development.

Cluster R-SE: Non-mature rural tourism, near to the airport. The effect of professionalism is non-existent in this cluster and, unlike the R-NW cluster, distance to the airport is positively valued. Heating system is also highly appreciated in this area.

Cluster R-NE: Mature non-coastal rural tourism. Unlike the other rural clusters, professionalism is appreciated together with certain management attributes, such as the requirement of a security deposit. The number of points of interests in the surrounding area and the location rating are also valued.

5 Discussion

Most tourism studies that analyze price determinants in a particular geographical zone start with the assumption that the tourism in the zone is homogeneous or delimited by prefixed geographical areas. This *ex-ante* classification is possible when dealing with high level rough classifications (e.g. beach, urban and rural tourism). In contrast, this paper proposes to classify tourist areas through an *ex-post* analysis of price determinants. This approach is specifically useful when the tourism zone to be analyzed includes several subtypes of tourism which are not clearly delimited. By means of this methodology, the hidden sub-types can be revealed without the need to previously associate them to a specific geographical area.

This paper extends previous attempts to segment tourism market using results from price determinant analysis (Latinopoulos, 2018) by proposing a methodology combining GWR with clustering analysis to identify market segments associated to geographical areas. Nevertheless, the process is not fully automated. It is not proposed an algorithm that can be run to find the available clusters and their characteristics. Instead, it is a process that fits with the current body of research in the areas of big data and data analytics (Li, Xu, Tang, Wang, & Li, 2018). It includes a series of tools that can help tourism experts in the process of understanding the tourism segments that are available in a particular geographical area. The participation of local tourism experts will be necessary in the process of choosing the best available model, in the determination of the optimal final number of clusters, and in the process of interpreting the resulting clusters based on the significant variables in each of them.

Additionally, while most tourism segmentation methods currently available use primary and secondary tourist data (Tkaczynski et al., 2009), or in other words demand-side data, in this research data on the available offer was used to segment the tourist areas. In this respect, this research has been able to segment tourism types on the basis of the properties on offer and their characteristics. In other words, using spatial information and data on P2P accommodation units, it has been possible to develop a new methodology and approach to the tourism segmentation process.

The most relevant price determinants found in the specific case study are consistent with previous findings. This is the case of the number of bathrooms and bedrooms, which were commonly significant price determinants in the aforementioned studies shown in Table 1. Another two relevant structural variables were detected, namely the availability of a pool and dryer machine. The latter variable is a consequence of the humid weather conditions in most of the territory analyzed. However, in contrast with most price determinant studies of the P2P accommodation industry, the availability of a swimming pool is appreciated in most of the tourism clusters (urban, beach and rural). This result is expected for the beach clusters, as was the case in price determinant analyses for sun and beach hotels (Soler et al., 2019; X. Wang et al., 2019). The findings reveal that this attribute is also appreciated in other coexisting types of tourism in the island, pointing to the existence of some specific attributes that are shared by all clusters in the geographical area.

Additionally, previous studies (Chen & Xie, 2017; Chica-Olmo et al., 2020; Gibbs et al., 2018; Magno et al., 2018; Moreno-Izquierdo et al., 2019) have reported that professional hosts charge more than amateur hosts. The findings in this paper also reveal that the effect of professionalism is higher in areas where tourism is more mature, such as the coastal areas in the island. This observation cannot be deduced from applying OLS techniques to the sample, as was the case in the aforementioned studies.

The positive influence of the strict cancellation policy is also consistent with previous results on price determinants of P2P accommodation units (Benítez-Aurioles, 2018; Cai et al., 2019; Moreno-Izquierdo et al., 2019; Wang & Nicolau, 2017). Other findings, such as the fact that the number of reviews negatively influences rental price whereas ratings have a positive influence, have also been observed in previous studies (Benítez-Aurioles, 2018; Cai et al., 2019; Gibbs et al., 2018; Lorde et al., 2019; Magno et al., 2018; Moreno-Izquierdo et al., 2019; Wang & Nicolau, 2017; Zhang et al, 2017, among others).

The distance to main attractions (e.g. beach) has been recurrently identified as a price determinant in hotels (Espinet et al., 2003; Papatheodorou, 2002; Thrane, 2005) and P2P accommodation units (Cai et al., 2019a; Chica-Olmo et al., 2020; Gibbs et al., 2018; Perez-Sanchez et al., 2018; Tong & Gunter, 2020). The methodology applied in this paper allows determination of the specific effect of the distance to several attractions, such as beach and airport, in the different clusters.

As usual in studies using GWR to analyze price determinants (Lee et al., 2020; Soler & Gemar, 2018; H. Zhang et al., 2011; Zhihua Zhang et al., 2017), the adjusted R^2 is heterogeneously distributed in the area, being higher in the south than in the northwest of the island. In order to interpret this result, it is necessary to bear in mind the general type of tourism developed in the areas where differences in the adjusted R^2 are observed: the south is a traditional beach tourism area and the north is a rural area where tourist activity is a more recent phenomenon. Therefore, the differences in the adjusted R^2 are interpreted in terms of market clearing: the P2P accommodation market in the south of the island is mature and therefore prices are well described by common P2P accommodation characteristics in mature coastal destinations, such as in Caribbean countries (Lorde et al., 2019a) and the region of Valencia, Spain (Moreno-Izquierdo et al., 2019; Perez-Sanchez et al., 2018). In contrast, the market in the northwest has not been cleared yet and therefore the price determinants have not been clearly established.

The findings above were possible through the use of econometric techniques which take into account the spatial distribution of the sample, such as GWR. The importance of using this methodology has been highlighted in other studies (Soler & Gemar, 2018; H. Zhang et al., 2011). In this regard, this paper stresses the suitability of GWR in conjunction with clustering techniques to conduct a segmentation analysis.

6 Conclusions

In this paper, it is proposed a method for the determination of the existing latent tourism clusters in an area using the characteristics of the accommodation units in it. The methodology is developed in five steps. First, geospatial variables that are added to each unit are calculated. Second, the price determinant variables using GWR are calculated. Third, the optimal number of clusters is obtained. Fourth, the clustering method is applied to find the existing clusters. Finally, it is analyzed whether the resulting clusters fitted appropriately the geographical areas in which the different types of tourism take place and the sub-type of tourism that was taking place in each of the clusters is interpreted.

This methodology can be used in any geographical area or destination in which several types of tourism are taking place. The case of P2P accommodation units was considered, but if sample size is sufficiently large this method could also be applied to hotel characteristics and average daily hotel rates. Of course, the advantage of this method is that P2P accommodation units are currently available almost everywhere, while this is not the case with hotels.

Specifically, this research has shown the application of GWR to a large dataset of P2P accommodation units (more than 2200) on the island of Gran Canaria (Canary Islands, Spain). Although the island is traditionally considered a sun and beach destination, its more than 1500 km^2 also accommodates urban and rural tourism, among other types. For each of the units, more than 150 available variables were available. In this particular case study, an *ex-ante* classification would have probably determined the existence of the three main types of tourism, but would have not been able to go much deeper. While it may have been possible to determine the existence of different types of tourism in the sun and beach areas, the method proposed in this paper allowed to identify eight tourism clusters.

Although this method can be used in any tourism zone, it is especially relevant in tourist destinations that include many types and subtypes of tourism. Properly understanding the different subtypes of tourism that are taking place in a destination is important for all stakeholders. This finer segmentation has different advantages for all the tourism operators in the area in processes that can lead to competitive advantages (Dolničar, 2004) such as promotion message segmentation, adaptation of public infrastructures, worker training, and the development of new tourism products, among many others. Additionally, this segmentation can be compared with a detailed analysis of the demand available in order to better understand the tourism product-market fit in a particular geographic area.

For example, destination management organizations (DMOs) can better develop their promotion strategies if they know the specific type of tourists that are interested in a certain area and their requirements. DMOs can offer language courses for hosts in certain areas and can prepare a list of the relevant price determinants for each of the areas, so that hosts can take them into account. While some of the determinants may be difficult to fulfil (e.g. changing the size of the unit or including a swimming pool), others can be more feasible (e.g. speaking a certain language, hot tub, dryer, dishwasher, breakfast).

7 Limitations

This study is not exempt of certain limitations. It should be noted that this methodology is useful for the identification of market segments which can be geographically delimited. When different types of tourism co-exist in the same area and cannot be spatially separated (e.g. culture, shopping and gastronomy in the same neighborhood in urban destinations), the methodology is not suitable to conduct a market segmentation analysis. Other limitations related to the case study can be described. First, only one geographical region was analyzed. While the method can be applied to different territories, this needs to be verified by using it in other environments that may be even more complex than the one chosen for this case study. Second, the location of the units is not exact but approximate. This limitation is imposed by the data offered by Airbnb. Third, most hosts are not professional ones, and may have not correctly reported some of the information of their units. Finally, while the characteristics of units that had actually been rented were used, not all the units have a similar demand level.

Acknowledgments

The authors wish to thank two anonymous reviewers for their helpful comments to the previous version of the manuscript. This work was supported by the Department of Economy, Industry and Trade of the Government of Canary Islands, Spain [grant number PROID2017010040]; the University of Las Palmas de Gran Canaria [grant number COVID 19–04] and the Ministry of Science, Innovation and Universities [grant number GOB-ESP2019-07].

References

Airbnb. 2020. Fast Facts. Available at https://news.airbnb.com/fast-facts/.

Bakker, M. H. E., & Twining-Ward, L. D. (2018). *Tourism and the sharing economy: Policy & potential of sustainable* (No. 130054).

Benítez-Aurioles, B. (2018). Why are flexible booking policies priced negatively? *Tourism Management*, *67*, 312–325. https://doi.org/10.1016/j.tourman.2018.02.008

Bigné, E., Gnoth, J., & Andreu, L. (2007). Advanced topics in tourism market segmentation. In *Tourism Management: Analysis, Behaviour and Strategy* (pp. 151–173). https://doi.org/10.1079/9781845933234.0151

Brunsdon, C., Fotheringham, A. S., & Charlton, M. E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical Analysis*, *28*(4), 281–298. https://doi.org/10.1111/j.1538-4632.1996.tb00936.x

Bulchand-Gidumal, J., Melián-González, S., & López-Valcárcel, B. G. (2019). Is the sharing economy for all? An answer based on neighbourhoods, types of hosts, and user complaints. In *Information and Communication Technologies in Tourism 2019* (pp. 55–66). Springer International Publishing. https://doi.org/10.1007/978-3-030-05940-8_5

Cahill, M., Mulligan, G., Cahill, M., & Mulligan, G. (2007). Local crime patterns. *Social*

Science Computer Review, *25*(2), 174–193.

Cai, Y., Zhou, Y., Ma, J., & Scott, N. (2019a). Price determinants of Airbnb listings: Evidence from Hong Kong. *Tourism Analysis*, *24*(2), 227–242. https://doi.org/10.3727/108354219X15525055915554

Cai, Y., Zhou, Y., Ma, J., & Scott, N. (2019b). Price determinants of Airbnb listings: Evidence from Hong Kong. *Tourism Analysis*, *24*(2), 227–242. https://doi.org/10.3727/108354219X15525055915554

Chen, Y., & Xie, K. L. (2017). Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management*, *29*(9), 2405–2424. https://doi.org/10.1108/IJCHM-10-2016-0606

Chica-Olmo, J., González-Morales, J. G., & Zafra-Gómez, J. L. (2020). Effects of location on Airbnb apartment pricing in Málaga. *Tourism Management*, *77*. https://doi.org/10.1016/j.tourman.2019.103981

da Silva, A. R., & Fotheringham, A. S. (2016). The multiple testing issue in geographically weighted regression. *Geographical Analysis*, *48*(3), 233–247. https://doi.org/10.1111/gean.12084

Dolničar, S. (2004). Beyond "Commonsense segmentation": A systematics of segmentation approaches in tourism. *Journal of Travel Research*, *42*(3), 244–250. https://doi.org/10.1177/0047287503258830

Dolnicar, S., Lazarevski, K., & Yanamandram, V. (2013). Quality of life and tourism: A conceptual framework and novel segmentation base. *Journal of Business Research*, *66*(6), 724–729. https://doi.org/10.1016/j.jbusres.2011.09.010

Domínguez-Mújica, J., González-Pèrez, J., & Parreño-Castellano, J. (2011). Tourism and human mobility in spanish archipelagos. *Annals of Tourism Research*, *32*(2), 586–606.

Ert, E., Fleischer, A., & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism Management*, *55*, 62–73. https://doi.org/10.1016/j.tourman.2016.01.013

Espinet, J. M., Saez, M., Coenders, G., & Fluvia, M. (2003). Effect on prices of the attributes of holiday hotels: A hedonic prices approach. *Tourism Economics*, *9*(2), 165–177. https://doi.org/10.5367/000000003101298330

Eugenio-Martin, J. L., Cazorla-Artiles, J. M., & González-Martel, C. (2019). On the determinants of Airbnb location and its spatial distribution. *Tourism Economics*. https://doi.org/10.1177/1354816618825415

Falk, M. (2008). A hedonic price model for ski lift tickets. *Tourism Management*, *29*(6), 1172–1184.

Falk, Martin, Larpin, B., & Scaglione, M. (2019). The role of specific attributes in determining prices of Airbnb listings in rural and urban locations. *International Journal of Hospitality Management*, *83*, 132–140. https://doi.org/10.1016/j.ijhm.2019.04.023

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically Weighted Regression: the analysis of spatially varying relationships*. (Wiley, Ed.). Chichester.

FRONTUR. (2020). Turistas en Gran Canaria y Canarias, por meses. Retrieved from https://www.grancanaria.com/turismo/fileadmin/PDF/informes/PDF/FRONTUR.pdf

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J., & Goodwill, A. (2018). Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of Travel and Tourism Marketing*, *35*(1), 46–56. https://doi.org/10.1080/10548408.2017.1308292

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

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means clustering algorithm author (s): J . A . Hartigan and M . A . Wong Published by : Blackwell Publishing for the Royal Statistical Society Stable URL : http://www.jstor.org/stable/2346830. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/https://doi.org/10.2307/2346830

Hong, I., & Yoo, C. (2020). Analyzing spatial variance of airbnb pricing determinants using multiscale GWR approach. *Sustainability (Switzerland)*, *12*(11). https://doi.org/10.3390/su12114710

Huertas, A., & Marine-Roig, E. (2015). Destination brand communication through the social media: What contents trigger most reactions of users? In *Information and Communication Technologies in Tourism 2015* (pp. 295–308). Springer International Publishing. https://doi.org/10.1007/978-3-319-14343-9_22

Kim, J., Jang, S., Kang, S., & Kim, S. H. (James). (2020). Why are hotel room prices different? Exploring spatially varying relationships between room price and hotel attributes. *Journal of Business Research*, *107*(September 2018), 118–129. https://doi.org/10.1016/j.jbusres.2018.09.006

Latinopoulos, D. (2018). Using a spatial hedonic analysis to evaluate the effect of sea view on hotel prices. *Tourism Management*, *65*, 87–99. https://doi.org/10.1016/j.tourman.2017.09.019

Lee, Y., Jang, S., & Kim, J. (2020). Tourism clusters and peer-to-peer accommodation. *Annals of Tourism Research*, *83*(102960).

Leung, Y., Mei, C. L., & Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, *32*(1), 9–32. https://doi.org/10.1068/a3162

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301–323. https://doi.org/10.1016/j.tourman.2018.03.009

Lorde, T., Jacob, J., & Weekes, Q. (2019a). Price-setting behavior in a tourism sharing economy accommodation market: A hedonic price analysis of AirBnB hosts in the caribbean. *Tourism Management Perspectives*, *30*, 251–261. https://doi.org/10.1016/j.tmp.2019.03.006

Lorde, T., Jacob, J., & Weekes, Q. (2019b). Price-setting behavior in a tourism sharing economy accommodation market: A hedonic price analysis of AirBnB hosts in the caribbean. *Tourism Management Perspectives*, *30*(February), 251–261. https://doi.org/10.1016/j.tmp.2019.03.006

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–296). Berkeley, CA: University of California Press.

Magno, F., Cassia, F., & Ugolini, M. M. (2018). Accommodation prices on Airbnb: effects of host experience and market demand. *TQM Journal*, *30*(5), 608–620. https://doi.org/10.1108/TQM-12-2017-0164

Martín, J. C., Román, C., & Mendoza, C. (2018). Determinants for sun-and-beach self-catering accommodation selection: A stated preference approach. *Tourism Economics*, *24*(3), 319–336. https://doi.org/10.1177/1354816618758731

Melián-González, A., Moreno-Gil, S., & Araña, J. E. (2011). Gay tourism in a sun and beach destination. *Tourism Management*, *32*(5), 1027–1037. https://doi.org/10.1016/j.tourman.2010.08.015

Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1–2), 17–23. https://doi.org/10.1093/biomet/37.1-2.17

Moreno-Izquierdo, L., Ramón-Rodríguez, A. B., Such-Devesa, M. J., & Perles-Ribes, J. F. (2019). Tourist environment and online reputation as a generator of added value in the sharing economy: The case of Airbnb in urban and sun- and-beach holiday destinations. *Journal of Destination Marketing and Management*, *11*(December 2018), 53–66. https://doi.org/10.1016/j.jdmm.2018.11.004

Önder, I., Weismayer, C., & Gunter, U. (2019). Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tourism Economics*, *25*(8), 1150–1166. https://doi.org/10.1177/1354816618805860

Páez, A., Uchida, T., & Miyamoto, K. (2002a). A general framework for estimation and inference of geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for locational heterogeneity. *Environment and Planning A*, *34*(4), 733–754. https://doi.org/10.1068/a34110

Páez, A., Uchida, T., & Miyamoto, K. (2002b). A general framework for estimation and inference of geographically weighted regression models: 2. Spatial association and model specification tests. *Environment and Planning A*, *34*(5), 883–904.

https://doi.org/10.1068/a34133

Papatheodorou, A. (2002). Exploring competitiveness in Mediterranean resorts. *Tourism Economics*, *8*(2), 133–150. https://doi.org/10.5367/000000002101298034

Pawlicz, A., & Napierala, T. (2017). The determinants of hotel room rates: an analysis of the hotel industry in Warsaw, Poland. *International Journal of Contemporary Hospitality Management*, *29*(1), 571–588. https://doi.org/10.1108/IJCHM-12-2015-0694

Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R. T. (2018). The what, where, and why of airbnb price determinants. *Sustainability (Switzerland)*, *10*(12). https://doi.org/10.3390/su10124596

Soler, I. P., & Gemar, G. (2018). Hedonic price models with geographically weighted regression: An application to hospitality. *Journal of Destination Marketing and Management*, *9*(May 2016), 126–137. https://doi.org/10.1016/j.jdmm.2017.12.001

Soler, I. P., Gemar, G., Correia, M. B., & Serra, F. (2019). Algarve hotel price determinants: A hedonic pricing model. *Tourism Management*, *70*(August 2018), 311–321. https://doi.org/10.1016/j.tourman.2018.08.028

Suárez-Vega, R., & Hernández, J. M. (2020). Selecting prices determinants and including spatial effects in peer-to-peer accommodation. *International Journal of Geo-Information*, *9*, 259. https://doi.org/10.3390/ijgi9040259

Tang, L. (Rebecca), Kim, J., & Wang, X. (2019). Estimating spatial effects on peer-to-peer accommodation prices: Towards an innovative hedonic model approach. *International Journal of Hospitality Management*, *81*, 43–53. https://doi.org/10.1016/j.ijhm.2019.03.012

Teubner, T., Hawlitschek, F., & Dann, D. (2017). Price determinants on Airbnb: How reputation pays off in the sharing economy. *Journal of Self-Governance and Management Economics*, *5*(4), 53–80. https://doi.org/10.22381/jsme5420173

Thorndike, R. L. (1953). Who belongs in a family? *Psychometrica*, *18*(4), 267–276.

https://doi.org/10.1117/12.620399

Thrane, C. (2005). Hedonic price models and sun-and-beach package tours: The Norwegian case. *Journal of Travel Research*, *43*(3), 302–308. https://doi.org/10.1177/0047287504272034

Thrane, C. (2007). Examining the determinants of room rates for hotels in capital cities: The Oslo experience. *Article in Journal of Revenue & Pricing Management*, *5*(4), 315–323. https://doi.org/10.1057/palgrave.rpm.5160055

Tkaczynski, A., Rundle-Thiele, S. R., & Beaumont, N. (2009). Segmentation: A tourism stakeholder view. *Tourism Management*, *30*(2), 169–175. https://doi.org/10.1016/j.tourman.2008.05.010

Tkaczynski, A., Rundle-Thiele, S. R., & Prebensen, N. K. (2015). Segmenting potential nature-based tourists based on temporal factors: The case of Norway. *Journal of Travel Research*, *54*(2), 251–265. https://doi.org/10.1177/0047287513514296

Tong, B., & Gunter, U. (2020). Hedonic pricing and the sharing economy: how profile characteristics affect Airbnb accommodation prices in Barcelona, Madrid, and Seville. *Current Issues in Tourism*, *0*(0), 1–20. https://doi.org/10.1080/13683500.2020.1718619

Valentin, M., & O'Neill, J. (2019). The value of location for urban hotels. *Cornell Hospitality Quarterly*, *60*(1), 5–24. https://doi.org/10.1177/1938965518777725

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, *62*, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

Wang, X., Sun, J., & Wen, H. (2019). Tourism seasonality, online user rating and hotel price: A quantitative approach based on the hedonic price model. *International Journal of Hospitality Management*, *79*, 140–147. https://doi.org/10.1016/j.ijhm.2019.01.007

Windle, M. J. S., Rose, G. A., Devillers, R., & Fortin, M. J. (2010). Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression

(GWR): An example from the Northwest Atlantic. *ICES Journal of Marine Science*, *67*(1), 145–154. https://doi.org/10.1093/icesjms/fsp224

Xie, K., & Mao, Z. (2017). The impacts of quality and quantity attributes of Airbnb hosts on listing performance. *International Journal of Contemporary Hospitality Management*, *29*(9), 2240–2260. https://doi.org/10.1108/IJCHM-07-2016-0345

Zhang, H., Zhang, J., Lu, S., Cheng, S., & Zhang, J. (2011). Modeling hotel room price with geographically weighted regression. *International Journal of Hospitality Management*, *30*(4), 1036–1043. https://doi.org/10.1016/j.ijhm.2011.03.010

Zhang, Zhihua, Chen, R. J. C., Han, L. D., & Yang, L. (2017). Key factors affecting the price of Airbnb listings: A geographically weighted approach. *Sustainability (Switzerland)*, *9*(9), 1–13. https://doi.org/10.3390/su9091635

Zhang, Ziqiong, Ye, Q., & Law, R. (2011). Determinants of hotel room price: An exploration of travelers' hierarchy of accommodation needs. *International Journal of Contemporary Hospitality Management*. https://doi.org/10.1108/09596111111167551

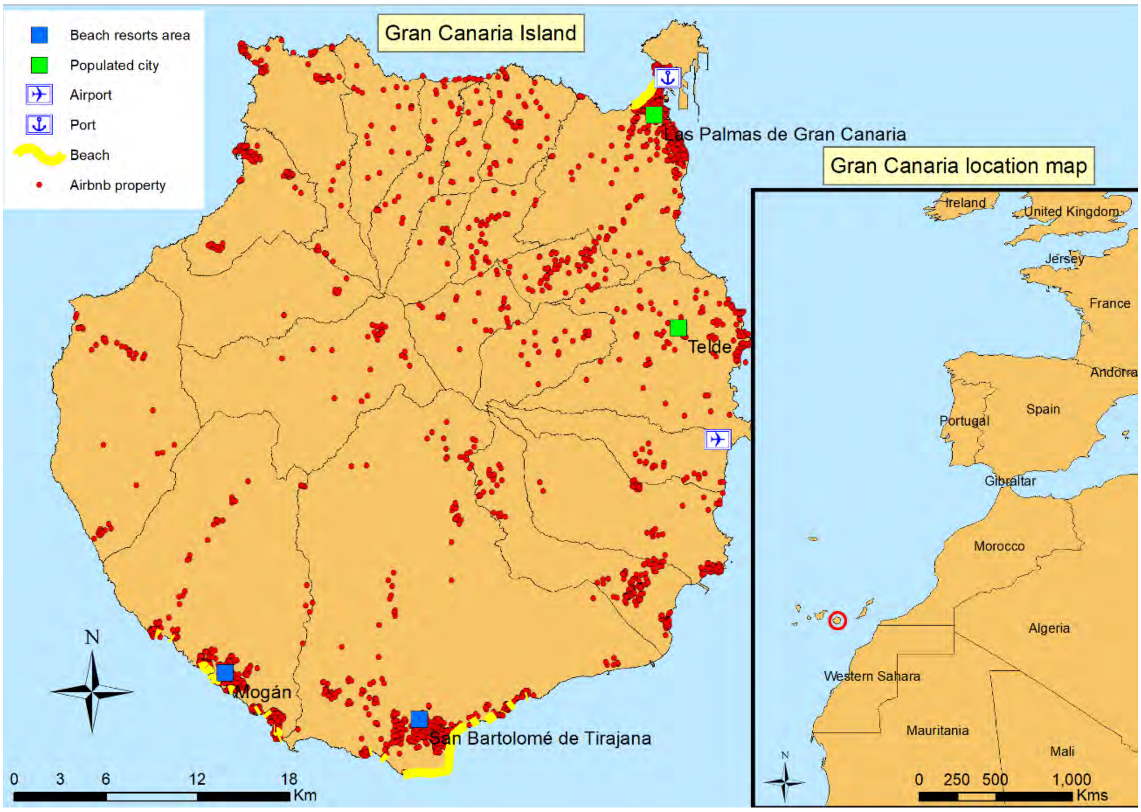

Figure 1. Gran Canaria island scenario.

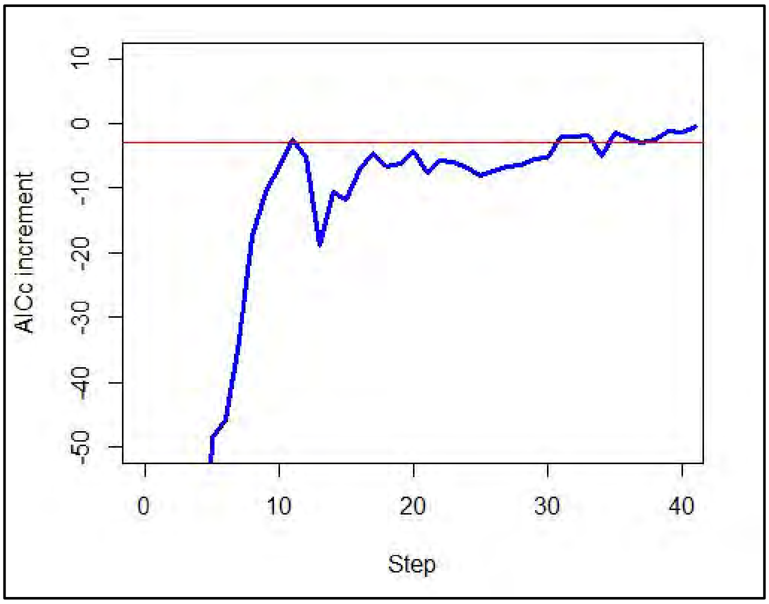

Figure 2. AICc increments along the SW-GWR procedure.

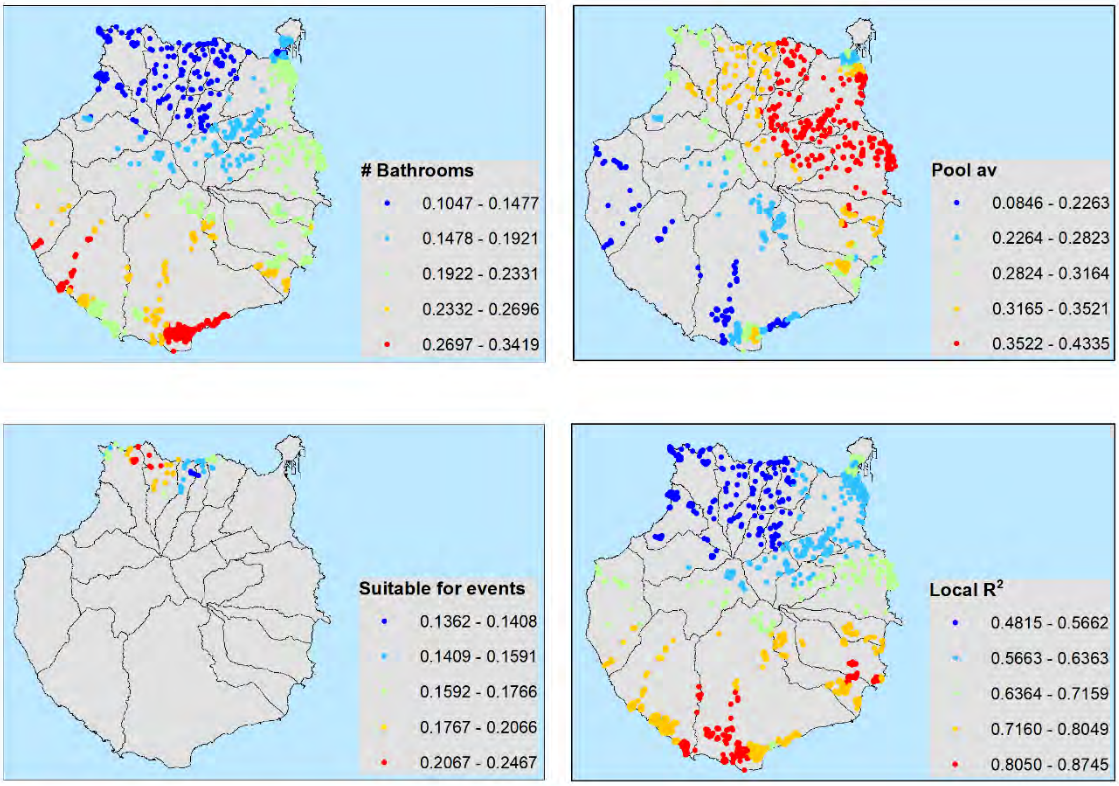

Figure 3. Distribution of the estimates of some of the significant price determinants throughout space.

= Number of; av = available.

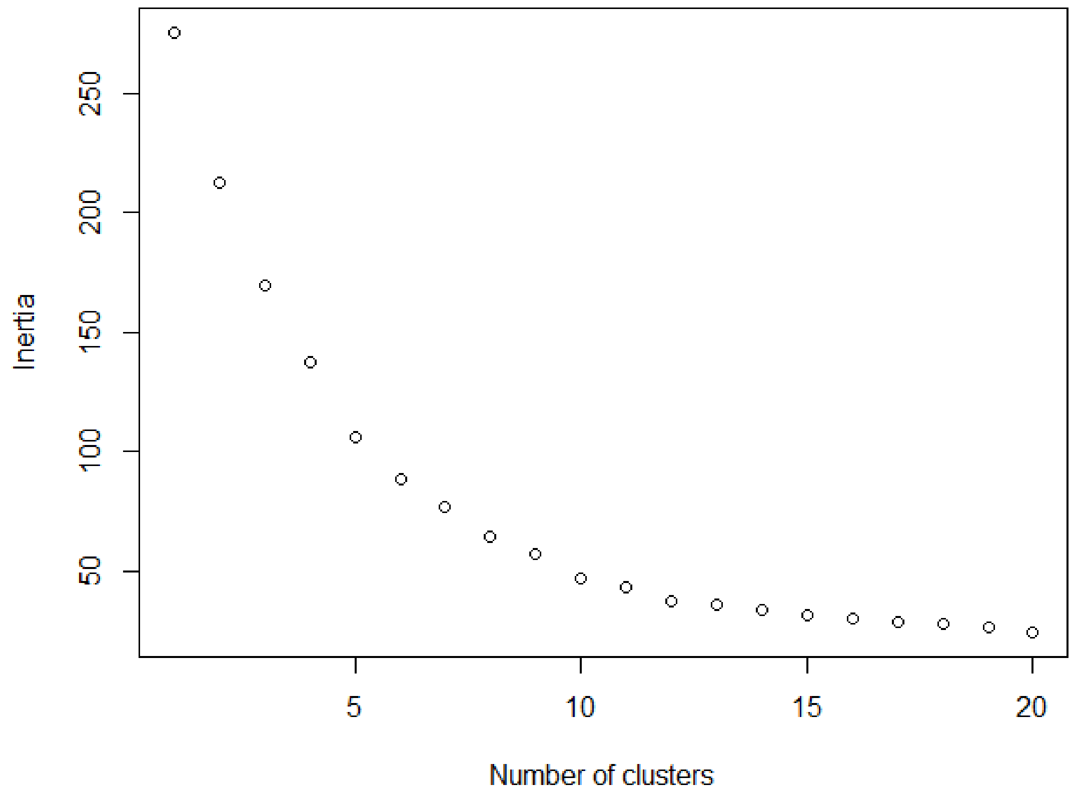

Figure 4. Inertia function for different numbers of clusters.

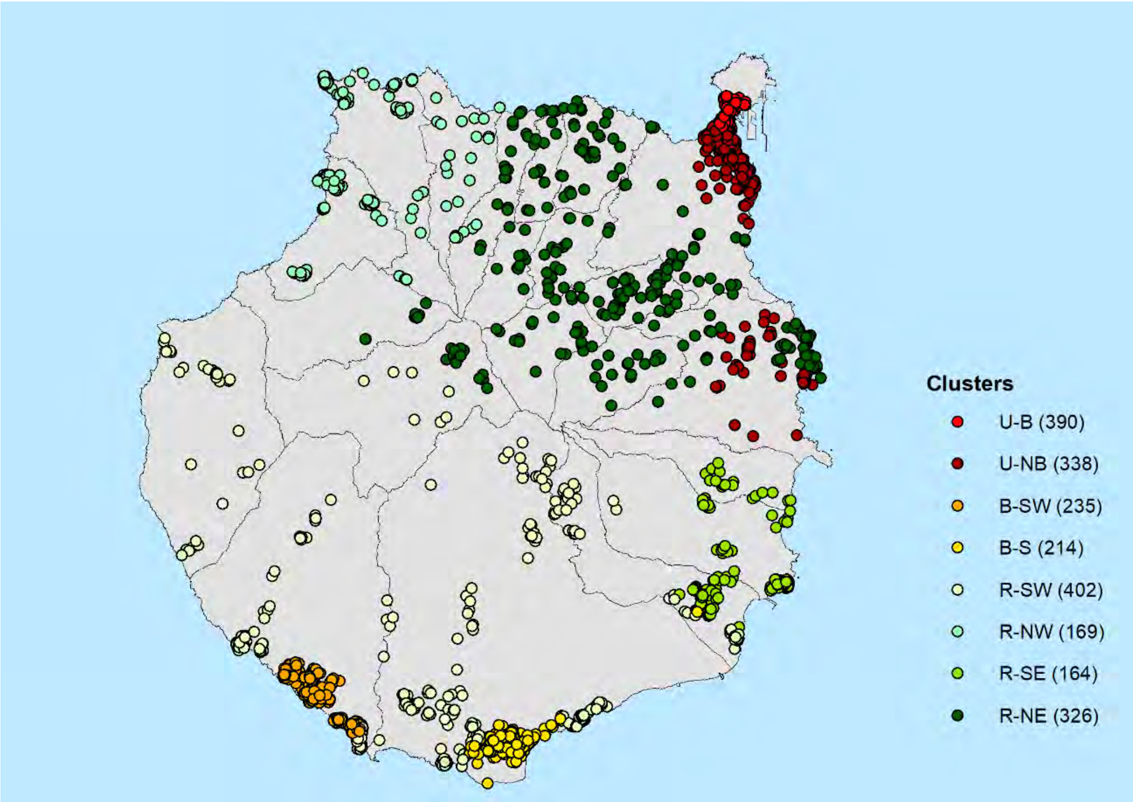

Figure 5. Spatial result for the k-means algorithm.

Note: The clusters are: U-B (Urban tourism, Beach); U-NB (Urban tourism, Not beach); B-SW (Beach Southwest area); B-S (Beach South area); R-SW (Rural Southwest area); R-NW (Rural Northwest area); R-SE (Rural Southeast area); R-NE (Rural Northeast area). Number of properties belonging to the cluster is shown in brackets. Cluster U-NB includes some properties in the second largest town (Telde-east).

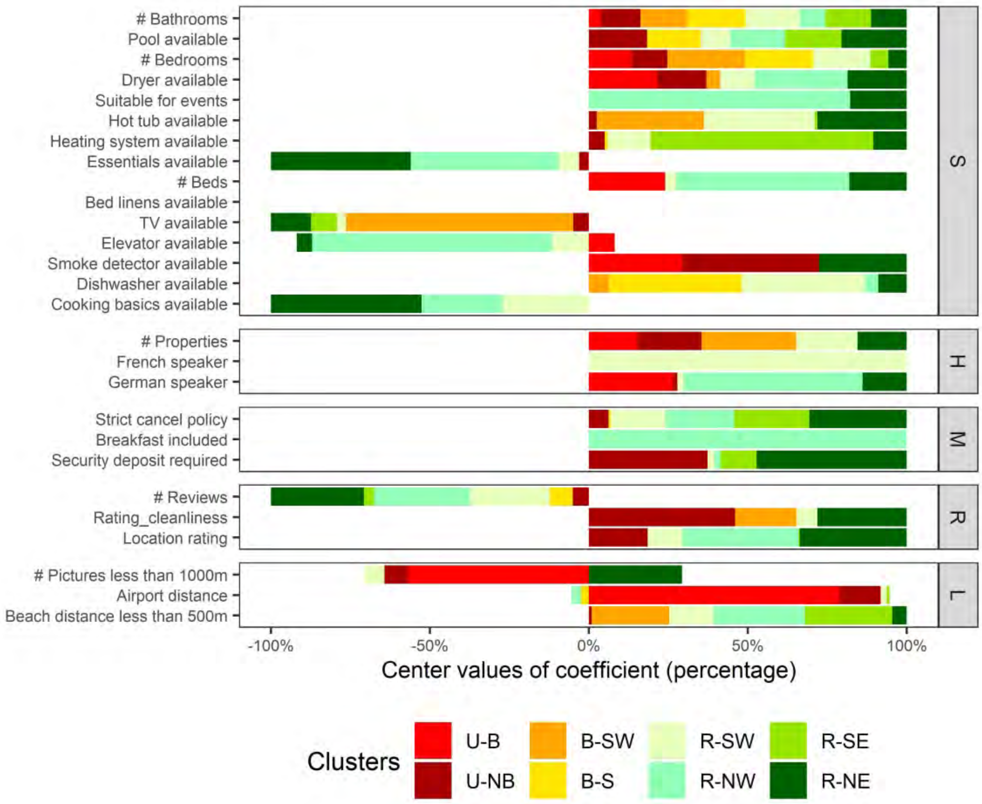

Figure 6. Stacked bar plot of the center values of coefficients in each cluster in percentage terms, ordered by type of determinant.

Note: The clusters are: U-B (Urban tourism, Beach); U-NB (Urban tourism, Not beach); B-SW (Beach Southwest area); B-S (Beach South area); R-SW (Rural Southwest area); R-NW (Rural Northwest area); R-SE (Rural Southeast area); R-NE (Rural Northeast area).

S = Structural attributes; H = Host attributes; M = Management attributes; R = Reputation attributes; L = Location attributes; # = Number of

Table 1. Review of price determinants estimations in Airbnb listings

Reference	Method	Area	Sample[1]	Determinants[2]						Main findings
				(S)	(H)	(M)	(R)	(L)	Total	(structural attributes not included)[3]
Ert et al. (2016)	OLS	Stockholm	395	3	5	0	2	0	10	+ trustworthy host photos, # reviews[4]
Chen & Xie (2017)	OLS	Austin, Texas	5779	5	2	3	7	3	20	+ response time, Host identity verified, professionalism - # hotels in surroundings
Teubner et al. (2017)	OLS	86 German cities	13,889	4	7	7	3	3	24	+ rating score, duration of membership, city GDP, population - # photographs, # ratings, professionalism
Wang & Nicolau (2017)	OLS and QR	33 cities around the World	180,533	13	4	5	2	1	25	+ being superhost, strict cancellation policy - Host photographs, # reviews, instant booking, smoking allowed
Zhang et al. (2017)	GWR	Nashville, Tennessee	974	0	0	0	3	2	5	- #reviews, rating, distance to city center
Benítez-Aurioles (2018)	OLS	44 cities in the world	497,507	5	1	2	2	0	10	+ being superhost - # reviews, flexible cancellation, instant booking
Cai et al. (2018)	OLS and QR	Hong-Kong	3351	9	4	3	2	4	22	+ being superhost, duration of membership, review score, strict cancellation - # reviews, instant booking, distance to point of interest
Gibbs et al. (2018)	OLS	5 Canada cities	11,239	9	2	2	2	1	16	+ # photographs, professionalism, being superhost - # reviews, distance to City Hall
Magno et al. (2018)	OLS	Verona	1056	2	2	0	1	1	6	+ professionalism, market demand - # reviews
Perez-Sánchez et al. (2018)	OLS and QR	Four cities in the region of Valencia (Spain)	19,578	5	1	5	1	10	22	+ # photographs, duration of membership, rating - distance to coast
Chattopadhyay & Mitra (2019)	OLS, random forest and decision tree	11 US cities	151,955	143				0	143	+ suitable for events

										- # reviews, # rating, # beds
Falk et al. (2019)	Panel data analysis and QR	Switzerland (urban and rural areas)	50,858	34	1	0	0	8	43	Use the title as explanatory variable (objective and subjective variables) objective factors influence more than subjective + sauna and jacuzzi (rural), suite, penthouse, view, old town location (urban)
Lorde et al. (2019)	OLS and QR	12 Caribbean countries	3046	13	3	11	2	5	34	+ # photographs, duration of membership, rating, being superhost, GDP, exchange rate - # rating, population
Moreno-Izquierdo et al (2019)	OLS and clustered errors	Region of Valencia (Spain) (urban and holiday destinations)	11,257	5	2	5	3	5	20	+ professionalism, strict cancellation, rating, # photographs, population, income (holiday destinations) - # reviews, occupation rate, tourist season (holiday destination)
Önder et al. (2019)	OLS with spatial adjusted data	Tallinn	1164	2	1	0	1	4	8	+ price of Airbnb in surroundings - distance to center
Tang et al. (2019)	Spatial econometrics	10 US cities	51,125	3	0	0	1	9	12	+ # Airbnb and median income in surroundings - # unemployed individuals in surroundings
Chica-Olmo et al. (2020)	Spatial econometrics	Malaga (Spain)	2967	9	2	3	2	7	22	+ professionalism, pedestrian routes density - # reviews, instant booking, noise, distance to points of interest
Tong & Gunter (2020)	Weighted Least Squares and QR	Barcelona, Madrid, Seville	29,675	4	2	5	2	1	14	+ rating, response rate, superhost, # photographs - # reviews, instant booking, distance to city center

[1] Number of properties in the sample.

[2] Number of determinants: (S) Number of structural attributes; (H) Number of host attributes; (M) Number of management attributes; ® Number of reputation attributes; (L) Number of location attributes.

[3] The structural attributes positively influence on price in most of the studies. The most influent attributes are property type, number of bedrooms, number of bathrooms, number of accommodates.

[4] Symbol # means 'number of'.

Works are sorted by year and alphabetically for each year.

Table 2. GWR model. Descriptive statistics of the coefficient estimations.

Coefficient	Type	Mean	SD	Min	Max	P. Sig.[1]	AICc	Adj.R^2
Intercept*		2.9750	0.4861	1.7546	4.1534	79.09		
Number of bathrooms*	S	0.2120	0.0540	0.1047	0.3419	88.78	1930.06	0.493
Pool available*	S	0.3058	0.0565	0.0846	0.4335	67.11	1751.43	0.538
Number of bedrooms*	S	0.1289	0.0505	0.0431	0.2205	84.41	1607.44	0.572
Number of properties*	H	0.0172	0.0061	0.0065	0.0276	51.88	1504.51	0.597
Dryer available*	S	0.1357	0.0296	0.0756	0.2157	63.94	1455.92	0.612
Number of pictures less than 1000m*	L	0.0000	0.0000	0.0000	0.0000	12.87	1410.19	0.623
Suitable for events	S	0.1787	0.0267	0.1362	0.2467	3.93	1376.56	0.632
Beach distance[2]	L	-	-	-	-	-	1359.04	0.640
Strict cancel policy	M	0.0545	0.0083	0.0385	0.0744	30.7	1348.45	0.638
Number of reviews	R	-0.0029	0.0007	-0.0049	-0.0017	38.16	1341.80	0.655
Hot tub available	S	0.2042	0.0307	0.1374	0.2906	42.14	1339.17	0.648
Heating system available*	S	0.1662	0.0734	0.0701	0.3068	22.12	1333.96	0.635
Essentials available*	S	-0.1942	0.0337	-0.2583	-0.1026	26.72	1315.17	0.641
Distance to airport*	L	0.0001	0.0001	0.0000	0.0003	32.57	1304.68	0.644
Cleanliness rating *	R	0.0622	0.0130	0.0366	0.0897	18.86	1292.94	0.650
Number of beds*	S	0.0482	0.0133	0.0222	0.0657	25.51	1285.71	0.654
Host speaks French	H	0.1224	0.0175	0.0689	0.1499	2.41	1281.18	0.654
Beach distance under 500m*	L	0.1663	0.0551	0.0786	0.4844	38.83	1274.40	0.655
Bed linen available*	S	0.0000	0.0000	0.0000	0.0000	0	1268.17	0.659
TV available*	S	-0.2266	0.1149	-0.4654	-0.0839	22.12	1263.82	0.662
Elevator available*	S	-0.1018	0.1158	-0.2937	0.1376	18.1	1256.16	0.658
Smoke detector available*	S	0.1411	0.0356	0.0729	0.2129	35.75	1250.34	0.651
Location rating	R	0.0649	0.0076	0.0449	0.0803	36.24	1244.28	0.649
Host speaks German*	H	0.1321	0.0251	0.0690	0.1690	18.68	1237.30	0.650
Breakfast included	M	0.2737	0.0527	0.1277	0.3603	7.55	1229.25	0.647
Security deposit required*	M	0.0004	0.0001	0.0001	0.0005	19.93	1221.80	0.650
Dishwasher available*	S	0.2568	0.0557	0.1083	0.3669	36.46	1215.18	0.651
Beach dist. at 200m-500m[2]	L	-	-	-	-	-	1208.77	0.653
Cooking basics available	S	-0.1205	0.0263	-0.1827	-0.0826	18.54	1203.34	0.655
Number of properties within 500m[2]	L	-	-	-	-	-	1198.06	0.656

Adjusted-α= 0.0049 OLS Adj.R^2: 0.558 GWR Adj.R^2: 0.651	F1 test p-value: 3.54e-08 F2 test p-value: 2.2e-16

[1] Percentage of properties in the sample where the variable is significant

[2] These variables were removed from the model

Type (of attribute): S – Structural; H – Host; M – Management; R: Reputation attributes; L – Location

* means that the F3-test is significant at 5%

Table 3. Discretization of values in Figure 6.

Variables	Cluster[1]							
	U-B	U-NB	B-SW	B-S	R-SW	R-NW	R-SE	R-NE
Number of properties	390	338	235	214	402	169	164	326
Structural								
No. of bathrooms	low	high	high	high	high	mid	high	mid
Pool available	-	high	-	high	mid	high	high	high
No. of bedrooms	mid	mid	high	high	high	-	low	low
Dryer available	high	mid	low	-	mid	high	-	mid
Suitable for events	-	-	-	-	-	high	-	low
Hot tub available	-	low	high	-	high	-	low	high
Heating system av.	-	low	-	low	low	-	high	low
Essentials available	low (-)	low (-)	-	-	low (-)	high (-)	-	high (-)
Number of beds	mid	-	-	-	low	high	-	mid
Bed linen available	-	-	-	-	-	-	-	-
TV available	-	low (-)	high (-)	-	low (-)	-	low (-)	low (-)
Elevator available	low (-)	-	-	-	low	high	-	low
Smoke detector av.	high	high	-	-	-	-	-	mid
Dishwasher av.	-	-	low	high	high	low	low	low
Cooking basics av.	-	-	-	-	mid (-)	mid (-)	-	high (-)
Host								
No. of properties	mid	high	high	-	mid	low	-	mid
French speaker	-	-	-	-	high	-	-	-
German speaker	mid	low	-	-	low	high	-	low
Management								
Strict cancel policy	-	low	-	low	mid	high	high	high
Breakfast included	-	-	-	-	-	high	-	-
Security deposit req.	-	high	-	-	low	low	low	high
Reputation								
Number of reviews	-	low (-)	-	low (-)	high (-)	high (-)	low (-)	high (-)
Rating cleanliness	-	high	mid	-	low	-	-	mid
Location rating	-	mid	-	-	low	high	low	high
Location								
Pictures less 1000m	high (-)	low (-)	-	-	low	-	-	mid
Airport distance	high	low	-	low (-)	low	low (-)	low	-
Beach less 500m	-	low	high	-	mid	high	high	low

[1]The clusters are: U-B (Urban tourism, Beach), U-NB (Urban tourism, Not beach), B-SW (Beach Southwest area), B-S (Beach South area), R-SW (Rural Southwest area), R-NW (Rural Northwest area), R-SE (Rural Southeast area), R-NE (Rural Northeast area).
Symbol (-) means the attribute exerts a negative effect on price in the cluster.

Supplementary Material

Table SM1 includes the list of explanatory variables used in the GWR model. Most of them were extracted from the Airbnb website (140), whereas some others were generated using GIS and Flickr photos repository.

Table SM1. Description of variables used in the GWR model**.**

Type[1]	Source	Description	Mean	sd	Min.	Max.
S	Airbnb website	Number of bathrooms	1.289	0.644	0	6
S	Airbnb website	Pool existence (1: Yes; 0: No)	0.357	0.479	0	1
S	Airbnb website	Max. no. people who can be accommodated	4.187	1.93	1	16
S	Airbnb website	Air conditioning availability (1: Yes; 0: No)	0.274	0.446	0	1
S	Airbnb website	Cable TV available (1: Yes; 0: No)	0.278	0.448	0	1
S	Airbnb website	Dryer available (1: Yes; 0: No)	0.228	0.42	0	1
S	Airbnb website	Number of bedrooms	1.765	1.059	0	10
S	Airbnb website	Hot tub (1: Yes; 0: No)	0.057	0.231	0	1
S	Airbnb website	Essentials available (1: Yes; 0: No)	0.942	0.233	0	1
S	Airbnb website	Dishwasher available (1: Yes; 0: No)	0.086	0.28	0	1
S	Airbnb website	Cooking basics available (1: Yes; 0: No)	0.255	0.436	0	1
S	Airbnb website	BBQ grill available (1: Yes; 0: No)	0.058	0.234	0	1
S	Airbnb website	TV available (1: Yes; 0: No)	0.928	0.259	0	1
S	Airbnb website	Hair dryer available (1: Yes; 0: No)	0.741	0.438	0	1
S	Airbnb website	Security deposit required	79.224	173.493	0	5000
S	Airbnb website	Wireless Internet available (1: Yes; 0: No)	0.855	0.352	0	1
S	Airbnb website	Elevator existence (1: Yes; 0: No)	0.357	0.479	0	1
S	Airbnb website	Extra pillows and blankets (1: Yes; 0: No)	0.181	0.385	0	1
S	Airbnb website	Refrigerator available (1: Yes; 0: No)	0.27	0.444	0	1
S	Airbnb website	Free parking on premises (1: Yes; 0: No)	0.444	0.497	0	1
S	Airbnb website	Bathtub with shower chair (1: Yes; 0: No)	0.003	0.056	0	1
S	Airbnb website	Availability of kitchen (1: Yes; 0: No)	0.982	0.133	0	1
S	Airbnb website	Bathtub existence (1: Yes; 0: No)	0.051	0.22	0	1
S	Airbnb website	Oven available (1: Yes; 0: No)	0.162	0.368	0	1
S	Airbnb website	First aids kit existence (1: Yes; 0: No)	0.403	0.491	0	1
S	Airbnb website	Safety card existence (1: Yes; 0: No)	0.181	0.385	0	1
S	Airbnb website	Hangers availability (1: Yes; 0: No)	0.853	0.354	0	1
S	Airbnb website	Table corner guards (1: Yes; 0: No)	0.008	0.091	0	1
S	Airbnb website	Wide entryway (1: Yes; 0: No)	0.043	0.204	0	1
S	Airbnb website	Crib available (1: Yes; 0: No)	0.185	0.389	0	1
S	Airbnb website	Number of beds	2.907	1.741	0	15
S	Airbnb website	Wheelchair accessible (1: Yes; 0: No)	0.121	0.326	0	1

S	Airbnb website	Stair gates existence (1: Yes; 0: No)	0.016	0.127	0	1
S	Airbnb website	Wide doorway (1: Yes; 0: No)	0.071	0.257	0	1
S	Airbnb website	Beach essentials available (1: Yes; 0: No)	0.067	0.251	0	1
S	Airbnb website	High chair available (1: Yes; 0: No)	0.153	0.36	0	1
S	Airbnb website	Stove available (1: Yes; 0: No)	0.165	0.371	0	1
S	Airbnb website	It is a flat (1: Yes; 0: No)	0.045	0.207	0	1
S	Airbnb website	Existence of a smooth pathway to front door (1: Yes; 0: No)	0.045	0.207	0	1
S	Airbnb website	Well lit path to entrance (1: Yes; 0: No)	0.083	0.276	0	1
S	Airbnb website	Heating available (1: Yes; 0: No)	0.243	0.429	0	1
S	Airbnb website	Garden or backyard existence (1: Yes; 0: No)	0.082	0.274	0	1
S	Airbnb website	Private entrance (1: Yes; 0: No)	0.171	0.376	0	1
S	Airbnb website	Ground floor access (1: Yes; 0: No)	0.002	0.047	0	1
S	Airbnb website	Shampoo available (1: Yes; 0: No)	0.551	0.497	0	1
S	Airbnb website	Access by mart lock (1: Yes; 0: No)	0.003	0.056	0	1
S	Airbnb website	Washer available (1: Yes; 0: No)	0.879	0.326	0	1
S	Airbnb website	Iron available (1: Yes; 0: No)	0.713	0.453	0	1
S	Airbnb website	Indoor fireplace (1: Yes; 0: No)	0.045	0.207	0	1
S	Airbnb website	Luggage drop-off allowed (1: Yes; 0: No)	0.086	0.28	0	1
S	Airbnb website	Fixed grab bars for shower toilet (1: Yes; 0: No)	0.007	0.084	0	1
S	Airbnb website	Internet available (1: Yes; 0: No)	0.366	0.482	0	1
S	Airbnb website	Electric vehicle charger (1: Yes; 0: No)	0.001	0.03	0	1
S	Airbnb website	Wide clearance to shower toilet (1: Yes; 0: No)	0.024	0.154	0	1
S	Airbnb website	Wide clearance to bed (1: Yes; 0: No)	0.046	0.21	0	1
S	Airbnb website	Step free access (1: Yes; 0: No)	0.106	0.308	0	1
S	Airbnb website	Coffee maker available (1: Yes; 0: No)	0.258	0.438	0	1
S	Airbnb website	Access by keypad (1: Yes; 0: No)	0.006	0.078	0	1
S	Airbnb website	Pocket Wi-Fi available (1: Yes; 0: No)	0.049	0.216	0	1
S	Airbnb website	Existence of fire extinguisher (1: Yes; 0: No)	0.319	0.466	0	1
S	Airbnb website	Baby bath available (1: Yes; 0: No)	0.053	0.223	0	1
S	Airbnb website	Buzzer wireless intercom av. (1: Yes; 0: No)	0.257	0.437	0	1
S	Airbnb website	Wide hallway clearance (1: Yes; 0: No)	0.056	0.23	0	1
S	Airbnb website	Lockbox existence (1: Yes; 0: No)	0.035	0.184	0	1
S	Airbnb website	Children's books and toys av. (1: Yes; 0: No)	0.07	0.256	0	1
S	Airbnb website	Accessible height toilet (1: Yes; 0: No)	0.037	0.189	0	1
S	Airbnb website	Accessible height bed (1: Yes; 0: No)	0.051	0.221	0	1
S	Airbnb website	Existence of smoke detector (1: Yes; 0: No)	0.201	0.401	0	1
S	Airbnb website	Hot water available (1: Yes; 0: No)	0.239	0.426	0	1
S	Airbnb website	Air purifier available (1: Yes; 0: No)	0	0.021	0	1

S	Airbnb website	Bed linens available (1: Yes; 0: No)	0.247	0.431	0	1
S	Airbnb website	Pack&Play travel crib available (1: Yes; 0: No)	0.126	0.332	0	1
S	Airbnb website	Handheld shower head existence (1: Yes; 0: No)	0.033	0.179	0	1
S	Airbnb website	Pets allowed (1: Yes; 0: No)	0.186	0.389	0	1
S	Airbnb website	Gym (1: Yes; 0: No)	0.031	0.173	0	1
S	Airbnb website	Microwave available (1: Yes; 0: No)	0.249	0.433	0	1
S	Airbnb website	Roll in shower with chair (1: Yes; 0: No)	0.007	0.084	0	1
S	Airbnb website	Outlet covers existence (1: Yes; 0: No)	0.026	0.159	0	1
S	Airbnb website	Single level home (1: Yes; 0: No)	0.054	0.226	0	1
S	Airbnb website	Suitable for events (1: Yes; 0: No)	0.039	0.195	0	1
S	Airbnb website	Patio or balcony existence (1: Yes; 0: No)	0.131	0.338	0	1
S	Airbnb website	Ethernet connection available (1: Yes; 0: No)	0.043	0.203	0	1
S	Airbnb website	Room darkening shades av. (1: Yes; 0: No)	0.152	0.359	0	1
S	Airbnb website	Dishes and silverware available (1: Yes; 0: No)	0.265	0.441	0	1
S	Airbnb website	Window guards existence (1: Yes; 0: No)	0.032	0.176	0	1
S	Airbnb website	Game console available (1: Yes; 0: No)	0.009	0.094	0	1
S	Airbnb website	Lock on bedroom door (1: Yes; 0: No)	0.128	0.334	0	1
S	Airbnb website	Fireplace guards (1: Yes; 0: No)	0.005	0.073	0	1
S	Airbnb website	Disabled parking spot (1: Yes; 0: No)	0.015	0.124	0	1
S	Airbnb website	Changing table available (1: Yes; 0: No)	0.013	0.113	0	1
H	Airbnb website	Number of properties managed by the host	3.207	4.566	1	28
H	Airbnb website	The host speaks French (1: Yes; 0: No)	0.131	0.337	0	1
H	Airbnb website	The host speaks German (1: Yes; 0: No)	0.165	0.371	0	1
H	Airbnb website	The host speaks Polish (1: Yes; 0: No)	0.012	0.111	0	1
H	Airbnb website	The host is a superhost (1: Yes; 0: No)	0.231	0.422	0	1
H	Airbnb website	The host speaks Portuguese (1: Yes; 0: No)	0.03	0.17	0	1
H	Airbnb website	The host speaks Spanish (1: Yes; 0: No)	0.525	0.499	0	1
H	Airbnb website	The host speaks English (1: Yes; 0: No)	0.554	0.497	0	1
H	Airbnb website	The host does not declare knowledge of any language (1: Yes; 0: No)	0.404	0.491	0	1
H	Airbnb website	The host speaks Danish (1: Yes; 0: No)	0.016	0.127	0	1
H	Airbnb website	The host speaks Norwegian (1: Yes; 0: No)	0.021	0.144	0	1
H	Airbnb website	The host has profile (1: Yes; 0: No)	0.997	0.056	0	1
H	Airbnb website	The host speaks Italian (1: Yes; 0: No)	0.11	0.313	0	1
M	Airbnb website	Cancel policy: 3: Flexible; 4: Moderate; 5: Strict; 9: Super Strict	4.317	0.79	3	9
M	Airbnb website	Family kids friendly (1: Yes; 0: No)	0.792	0.406	0	1
M	Airbnb website	Instant bookable allowed for experienced guests with government id (1: Yes; 0: No)	0.067	0.25	0	1
M	Airbnb website	Smoking allowed (1: Yes; 0: No)	0.294	0.456	0	1

M	Airbnb website	Cleaning before checkout av. (1: Yes; 0: No)	0.008	0.091	0	1
M	Airbnb website	Laptop friendly workspace av. (1: Yes; 0: No)	0.511	0.5	0	1
M	Airbnb website	Minimum number of nights to be rented	3.698	1.901	1	21
M	Airbnb website	Long term stays allowed (1: Yes; 0: No)	0.154	0.361	0	1
M	Airbnb website	The host's identity is verified (1: Yes; 0: No)	0.501	0.5	0	1
M	Airbnb website	Existence of carbon monoxide detector (1: Yes; 0: No)	0.105	0.306	0	1
M	Airbnb website	Babysitter recommendations av.(1: Yes; 0: No)	0.033	0.179	0	1
M	Airbnb website	Baby monitor available (1: Yes; 0: No)	0.001	0.036	0	1
M	Airbnb website	Children's dinnerware available (1: Yes; 0: No)	0.056	0.23	0	1
M	Airbnb website	Breakfast included (1: Yes; 0: No)	0.04	0.196	0	1
M	Airbnb website	Self-check In allowed (1: Yes; 0: No)	0.073	0.259	0	1
M	Airbnb website	Instant bookable allowed for guests with government id (1: Yes; 0: No)	0.098	0.297	0	1
M	Airbnb website	Host greets you (1: Yes; 0: No)	0.124	0.329	0	1
M	Airbnb website	24 hour check in available (1: Yes; 0: No)	0.23	0.421	0	1
M	Airbnb website	Instant bookable allowed for experienced guests (1: Yes; 0: No)	0.032	0.176	0	1
M	Airbnb website	Discount factor for weekly rentals	0.702	0.457	0	1
M	Airbnb website	Maximum number of nights to be rented	945.575	2604.701	3	112,030
M	Airbnb website	Doorman existence (1: Yes; 0: No)	0.1	0.3	0	1
M	Airbnb website	Has dismissed the Instant Booking for salmon flow (1: Yes; 0: No)	0.194	0.396	0	1
M	Airbnb website	Flexible check in is allowed (1: Yes; 0: No)	0.32	0.466	0	1
M	Airbnb website	Pets live on the property (1: Yes; 0: No)	0.042	0.2	0	1
M	Airbnb website	Instant booking allowed (1: Yes; 0: No)	0.529	0.499	0	1
M	Airbnb website	Instant bookable allowed for everyone (1: Yes; 0: No)	0.432	0.495	0	1
R	Airbnb website	Number of comments on the property	15.059	18.476	0	181
R	Airbnb website	Host's review count	60.75	102.947	0	1085
R	Airbnb website	Accuracy rating for the expected experience	9.689	0.580	2	10
R	Airbnb website	Check-in rating	9.800	0.508	2	10
R	Airbnb website	Cleanliness rating	9.625	0.646	2	10
R	Airbnb website	Communication rating	9.789	0.536	2	10
R	Airbnb website	Location rating	9.385	0.734	6	10
R	Airbnb website	Overall rating (between 0 and 10)	9.485	0.662	2	10
L	GIS	Distance, in meters, to the nearest beach	6582.82	7840.81	0	26,083
L	GIS	The beach is less than 200 meters from the property (1: Yes; 0: No)	0.169	0.375	0	1
L	GIS	The beach is less than 500 meters from the property (1: Yes; 0: No)	0.336	0.472	0	1
L	GIS	The beach is less than 1000 meters from the property (1: Yes; 0: No)	0.434	0.496	0	1

L	GIS	The beach is between 201 and 500 meters from the property (1: Yes; 0: No)	0.166	0.372	0	1
L	GIS	The beach is between 501 and 1000 meters from the property (1: Yes; 0: No)	0.098	0.298	0	1
L	GIS	Distance to the airport	23,873	9051	1809	42,751
L	GIS	Distance to the ship port	24,554	18,443	187.8	50,277
L	GIS	Distance to Las Palmas de Gran Canaria	21,709	10,466	0	40,639
L	GIS	Distance to Telde	26,684	15,520	241.4	45,740
L	GIS	Distance to San Bartolomé de Tirajana	22,967	17,514	147.2	47,735
L	GIS	Number of Airbnb properties at 100 meters	5.136	6.6	1	37
L	GIS	Number of Airbnb properties at 300 meters	27.024	36.989	1	170
L	GIS	Number of Airbnb properties at 500 meters	52.664	69.773	1	268
L	Flickr and GIS	Number of Flickr pictures taken within 1000 meters of the property	4606	5228	2	16,841
L	Airbnb website	Beachfront located (1: Yes; 0: No)	0.064	0.244	0	1
L	Airbnb website	Waterfront located (1: Yes; 0: No)	0.097	0.296	0	1

[1] S: Structural attribute; H: Host attribute; M: Management attribute; R: Reputation attribute; L: Location attribute.