



Price indicators for Airbnb accommodations

Giulia Contu¹ · Luca Frigau¹ · Claudio Conversano¹

Accepted: 2 November 2022
© The Author(s) 2022

Abstract

New forms of hospitality grew increasingly more popular and successful during the last decades. Nowadays, they are chosen for different reasons, one of the most important certainly being price. Understanding the elements that can impact on price determination is crucial to increase profitability. We propose two price indicators for Airbnb accommodations, which are defined in three phases using proportional odds model as a reference model. The first phase focuses on the probability estimation of accommodations belonging to a specific class of price. The second phase aims to evaluate the ability of the model to make good predictions by computing three different indexes. Finally, the three indexes are combined to define the indicators q and r which evaluate, respectively, the impact that six different dimensions (transports, culture, crowd, property, management, and time) have with respect to price determination on Airbnb accommodations and their relative importance concerning neighborhoods. The analysis is focused on 61 neighborhoods of Rome. The findings show differences with respect to the impact of the dimensions on price for each neighborhood of Rome.

Keywords Proportional odds model · VGAM · Impact on price · Price classes · Rome

1 Introduction

New forms of hospitality have arisen and developed in the last two decades: couches, private and shared rooms, boats, and apartments are the new forms of accommodations that can be booked besides the more traditional hotel rooms. The choice among these new forms of accommodations is determined by different aspects, such as a wider range of listings, a favorable price-quality ratio, a lower price, the possibility of choosing between a private or a shared environment, the possibility of meeting new people, or living a more authentic experience (Skalska 2017). These new kinds of accommodations are booked online on different platforms. Among these platforms, probably the most famous and successful is Airbnb. Airbnb has been founded in 2008 in San Francisco and has grown significantly over the years. At the moment, Airbnb lists over 6 million accommodation located in

✉ Luca Frigau
frigau@unica.it

¹ Department of Economics and Business Sciences, University of Cagliari, Viale Sant' Ignazio 17, 09123 Cagliari, Italy

100,000 cities and 220 countries and regions (<https://news.airbnb.com/about-us/>) (accessed on October 31st, 2022).

One of the most important factors that influence booking choices on the Airbnb platform is price, as price is the key to increasing profitability after satisfying guests expectations (Chattopadhyay and Mitra 2019). It is a strategic element that should be defined considering different aspects such as the specific characteristics of the accommodation, the provided services, and the location.

Identifying the determinants of price on Airbnb is such an important topic that numerous articles have been written on the subject. For instance, Dudás et al. (2020) have discovered that Airbnb accommodation prices are influenced by property-related attributes, such as air conditioning, free internet, free parking, number of bedrooms, and the possibility of booking the entire house or only a room. Moreover, Lorde et al. (2018) have demonstrated that attributes such as location, reputation, convenience, personal and amenities attribute significantly affect prices in the Caribbean. Lawani et al. (2019) have evaluated the impact of three groups of variables on the booking price: structural variables (as number of bedrooms and bathrooms), neighborhood variables (as the distance in feet to the closest convention center and train stations), and the quality-signaling variables (as cleanliness, number of reviews and communication). They have discovered that price is influenced by review scores, room characteristics, and neighborhood features. Finally, Teubner et al. (2017) have highlighted how host reputation can also be a price determinant.

The studies investigating Airbnb price are carried out using different methods such as ordinary least squares (see for instance Voltes-Dorta and Sánchez-Medina (2020)), panel data analysis (see for instance Neumann and Gutt (2017); Falk et al. (2019)), quantile regression (see for instance Wang and Nicolau (2017); Perez-Sanchez et al. (2018); Lorde et al. (2018)), hedonic price models (see for instance Sainaghi et al. (2021); Voltes-Dorta and Sánchez-Medina (2020); Faye (2021)), price equations with spatial effects (Tang et al. 2019), pricing strategy model (Ye et al. 2018), and machine learning (see for instance Kalehbasti et al. (2019)).

The variables included in the models are usually the price as the response variable, together with the independent variables that are normally grouped in categories such as site characteristics, reputation, convenience, personal and amenities attributes Faye (2021); Barnes and Kirshner (2021); Chattopadhyay and Mitra (2019). Generally speaking, previous studies investigated the impact of different groups of variables taking into account a whole city (see for instance Aznar et al. (2018); Moreno-Izquierdo et al. (2018); Voltes-Dorta and Sánchez-Medina (2020)) or more cities, comparing the results among them (see for instance Perez-Sanchez et al. (2018); Rodríguez-Pérez et al. (2019); Chattopadhyay and Mitra (2019)). Only some researchers as Faye (2021), Sainaghi et al. (2021) and Falk et al. (2019) have included in their models variables related to neighborhood's features in order to evaluate the impact that these aspects might have on price determination.

However, to our knowledge, nobody has studied the price determinants concerning the neighborhoods of a city to investigate the existence of differences in terms of variables' impact within a specific city. For this reason, we propose two indicators, capable of evaluating the impact of different groups of variables, hence called also dimensions, on Airbnb prices, that could therefore support hosts in the difficult task of price determination.

The two proposed indicators, which we will call respectively *q indicator* and *r indicator*, aim to estimate the importance of the dimensions in the price determination for each neighborhood, and the relative importance of a single dimension concerning the other neighborhoods. In particular, the indicators allow identifying the existence of possible differences related to the location of the accommodation.

The variables taken into account in the definition of the indicators have been previously used by other researchers in the investigation of the Airbnb price (see for instance Faye (2021); Önder et al. (2019); Cai et al. (2019)). As evidenced before, in literature they are grouped in a specific dimension. We decide to group the variables into six different dimensions. Specifically, we have grouped the variables related to the proximity to the center and the presence of subway and bus stops nearby in a dimension called *transports*; those related to the proximity of the accommodation to monuments and museums in a dimension called *culture*; the presence of other Airbnb accommodations and hotels in the neighborhood in the dimension *crowd*; the aspects related to the management of the accommodations such as the provided services, the badge of Superhost, the number of photos in the Airbnb platform in the dimension *management*; the characteristics of the accommodation such as the number of bedrooms and bathrooms in the dimension *property*; and, finally, the seasonality of the reservation in the dimension *time*. The choice of name dimensions is related to specific aspects that we have supposed to have a direct impact on the price determination. This is the first time that all these dimensions and variables are combined together in the estimation of a price indicator.

In this study, we illustrate the application of the two indicators on the Airbnb accommodations located in 61 neighborhoods of the capital of Italy, the city of Rome, which is one of the most visited cities in Italy and Europe. The findings of the analysis have shown, generally speaking, a different impact of the six dimensions on the entire city, but also specific differences within the single neighborhoods.

Five sections, besides the introduction, complete this study. The second and third sections are related to the research design: data and methodology are here described. The results and implications are explained in the fourth and fifth sections. Finally, the sixth and last section focuses on theoretical and practical implications, limitations, and future developments.

2 Data

The study has been realized using a dataset composed by different groups of variables (Table 1).

The choice of these dimensions and the relative variables has been affected by the literature reviews. Specifically, we have investigated the covariates used by other researchers with the aim to include in the model only the elements that previous studies have demonstrated to have really influenced Airbnb price determination. More in details, in literature the covariates used for price determination can be grouped in specific categories such as site characteristics, reputation, personal and amenities attributes. Analyzing the literature, we have discovered that researchers have included different variables inside each category. For instance, in the group *site characteristics*, Faye (2021) has included aspects such as the distance to the city center and the proximity to shopping areas or tourist attraction areas; Önder et al. (2019) the number of points of interest in the surrounding area; Cai et al. (2019) the number of Airbnb listings in the same district. In the group *reputation*, Faye (2021) has included aspects such as superhost badge, number of reviews, star ratings, host listing counts, host profile picture, response time, and cancellation rate. Instead, Cai et al. (2019) have considered only the variables number of reviews and the overall rating as part of this group. They have inserted the variables response time, cancellation rate, and Superhost into the last group of variables, called *personal and amenities attributes*.

Table 1 Variables used in the analysis

Label	Group	Description
DistCenter	<i>Transports</i>	Distance from city center (i.e. Pantheon)
DistSubway		Minimum distance from the closest subway stop
Bus200		Number of bus stops within 200 ms
Airbnb50	<i>Crowd</i>	Number of airbnb in a range of 50 ms
Airbnb500		Number of airbnb in a range of 500 ms
Hotels50		Number of hotels in a range of 50 ms
Hotels500	<i>Culture</i>	Number of hotels in a range of 500 ms
Monuments500		Number of monuments in a range of 500 ms
Monuments2000		Number of monuments in a range of 2000 ms
Bedrooms	<i>Property</i>	Number of bedrooms in a vacation rental listing
Bathrooms		Number of bathrooms in a vacation rental listing
Entirehome		Dummy variable: Entire home
FlexCancellation	<i>Management</i>	Dummy variable: Flexible cancellation policy
ResponseTime		Average time in minutes a host responds to
MinimumStay		The minimum night stay to book required by host
BusinessReady	<i>Time</i>	Dummy variable: Host who provides business facilities
Superhost		Dummy variable: High quality experienced host
NumberPhotos		Number of photos in a vacation rental listing
DayHigh	<i>Time</i>	Dummy variable: Monday, Friday or Saturday
MonthHigh		Dummy variable: March, April, May, September or October
Bridge		Dummy variable: Bridge Holiday
PNR	<i>Instrumental</i>	Published nightly rate for a vacation rental listing
Price		Categorization of PNR into five classes
Neighborhood		Neighborhood of Rome

The first column provides their label, the second their dimension, the third their description

Chattopadhyay and Mitra (2019) have taken into account the variables bathrooms, beds, number of reviews, review scores rating, and reviews per month as elements of the last group.

Analyzing the similarities and the differences in the variables classification, we have decided to define six specific groups of variables. They are called: *transports*, *culture*, *crowd*, *property*, *management*, and *time*.

More in detail, the variables in *transports* include the distance from the accommodations to the closest subway stop and to the city center (i.e. Pantheon), as well as the number of bus stops within 200 ms. Some researchers have discovered that Airbnb accommodations are often close to the main tourist attractions (see for instance Gutiérrez et al. (2017); Romano (2021); Celata and Romano (2022)); concentrated in specific areas of tourist cities (see for instance Dudás et al. (2017)); and located *in areas that are attractive and accessible by public transport* (Quattrone et al. 2016, p. 1391). Moreover, the same researchers have investigated the impact of distance to the nearest bus stop, to the nearest bus/train station, or to the city center (see for instance Voltes-Dorta and Sánchez-Medina (2020)). All these aspects have been summarized in the categories *transports* with the aim to investigate if the proximity to the city center or to public transportation facilities can influence the Airbnb price.

The group of variables called *crowd* measures the number of other accommodations or hotels close to each accommodation. Specifically, we consider the number of Airbnb accommodations and hotels within 50 meters for each accommodation, as well as those within 500 meters. Other researchers have inserted these variables in their model and they have labelled them with names such as *market conditions* (Faye 2021), *listing location* (Cai et al. 2019), and *supply and demand dynamics* (Ye et al. 2018). In this study, we have chosen the label *crowd* focusing on the aim to comprehend if the major presence of Airbnb listings and hotels, which can be assimilated to a crowd, can impact on price.

The dimension of *culture* concerns the number of monuments located within 500 and 2000 meters from the accommodation. Different researchers have analyzed the impact of tourism attractions and their distance from Airbnb accommodations (see for instance Voltes-Dorta and Sánchez-Medina (2020)). In this study, we take into consideration tourist attractions such as cultural monuments, museums, and churches. We believe that people choose Rome as a tourist destination also to visit its famous monuments, and for this reason we add these variables in the model and called their group *culture*, as the goal is in fact to investigate the impact of cultural attractions on price determination.

The group of variables related to the characteristics of the *property* includes the number of bedrooms and of bathrooms in the accommodation, as well as a dummy variable concerning the type of accommodation (1 if entire home, 0 if room). Analyzing the literature, we have found out that several researchers have included these variables in their price determination models (see for instance Chen and Xie (2017); Gibbs et al. (2018)). They have used different labels to define this category, such as *characteristics of listing* (Chen and Xie 2017), *listing attributes* (Cai et al. 2019), *rental listing features* (Kakar et al. 2016). We have also discovered that the characteristics of the accommodation are usually joined with other variables such as the number of reviews, the provided services, and the Superhost Badge. In this paper, we have decided to distinguish between the characteristics of the accommodation and services with the aim to comprehend if specific accommodation features are able to impact on the price determination independently of the provided services and the positive reputation of the host. This decision has also influenced the choice of the labels. In fact, we have chosen the label *property* for the specific characteristics of the accommodation; and the label *management* for those aspects related with both the activities directly handled by the hosts and their reputation.

As we just noticed, the group of variables called *management* includes those aspects that can be directly handled by hosts when managing their accommodations, i.e. the flexibility of cancellation policies, the amount of time that guests must wait on average to get an answer to their questions, the minimum number of nights that must be booked, the number of photos published on the Airbnb online platform, the Business Ready status, and the Superhost status. Before moving on to explain the choice of these variables, a few words are needed here to explain the last two variables. The Business Ready status is granted to accommodations equipped with services to support business travelers, such as Wi-Fi, laptop-friendly workspace, iron, hangers, shampoo, hairdryer, and all the other essentials such as toilet paper and clean towels. On the other hand, the Superhost status is attributed by Airbnb to accommodations that have fulfilled in the previous 12 months the following criteria (<https://www.airbnb.com/d/superhost>, accessed on October 31st, 2022): host at least 10 stays per year; honor at least 99% of the reservations unless there is an extenuating circumstance; answer guests' requests within 24 h at least 90% of the time; and achieve a 4.8+ overall rating.

Now let us return to explain the choice of variables in the *management* group. The variables *Cancellation policy* and *Response Time* have been chosen for two reasons: firstly,

Airbnb has declared to use them in the attribution of the badge of Superhost; and secondly, Gunter (2018) has discovered that they are actually relevant in the Superhost badge attribution. Since in literature researchers have discovered the existence of a positive relationship between the Superhost status and the price (see for instance Ma et al. (2017)), we believe that the two aspects, being involved in the attribution of the Superhost badge, can have an impact on the price determination, too. Similarly we believe that the variable *Business Ready*, collecting together several services of the accommodation, can be relevant in the price determination. Finally, the variable *Superhost* has been chosen as it is intended as a measure of the hosts' reputation, whereas the variable *Number of Photos* measures the host's commitment in the promotion of their accommodation.

In literature, different researchers have used the same variables, but grouped them differently. In some cases, the researchers have combined the five variables in two different groups. For instance, Ye et al. (2018) have included the first three variables in a category called *listing features* and Lorde et al. (2018) in the group *reputation of the product*. Meanwhile, the last two variables are grouped together in categories called *host attributes* by Cai et al. (2019) or *reputation of the host* by Lorde et al. (2018). In other cases, the five variables are joined together in the same group called *reputational variables* as in Faye (2021). We have decided to join all this variables in one group called *management* because this way we can put together aspects that are directly influenced by the host, as *Cancellation policy*, *Response Time* and the *Business Ready*, with others that can be interpreted as the result of the good management of an accommodation, as the badge of Superhost and the number of photos.

The group of variables called *time* consists of three dummy variables created to represent how low season and high season influence prices. More specifically, these variables estimate the impact that the high season might have on prices, in order to comprehend if and how prices are influenced during those periods of time with high tourism flows. We have created the variables attributing the value 1 to each day or month in the high season (like national holidays, the summer and so on). Other researchers have created similar variables, like for instance, Ye et al. (2018) who have used specific days of the year and the different days of the week to create a group called *temporal features*.

The final group of variables used in the analysis is called *instrumental* and includes the response variable "published nightly rate of the accommodation", as well as the neighborhood where the accommodation is located. The response variable has been transformed into an ordinal variable of five classes according to its quantiles (i.e. $1 = [0, 57]$, $2 = (57, 82]$, $3 = (82, 110]$, $4 = (110, 155]$, $5 = (155, \infty)$). We chose five classes as this number is a good trade-off to detect the variability of the price (from very low to very high), avoiding at the same time its excessive spreading out.

The city of Rome has been divided into 116 non-administrative units: 22 districts, called in Italian *rioni*, located in the historic center of the city, 35 *quarters* surrounding the historic center of Rome, 6 suburbs, and 53 zones in the countryside of the city, as shown in Fig. 1. Not all 116 non-administrative units have Airbnb accommodations, and for this reason the analysis has been held only for those areas of Rome where the listings are located. Specifically, we have focused on 61 different geographical areas of Rome, which are represented in different colors with respect to the typology of non-administrative units in Fig. 1, where more than 20 Airbnb accommodations are located. The aim was to obtain a significant sample that could be actually useful for the analysis.

The data that have been used in the analysis have been either bought or downloaded for free. Specifically, the data about *transports*, *crowd*, and *culture* have been downloaded for free from the website *Open Data Roma Capitale* (<https://dati.comune.roma.it/>), which is

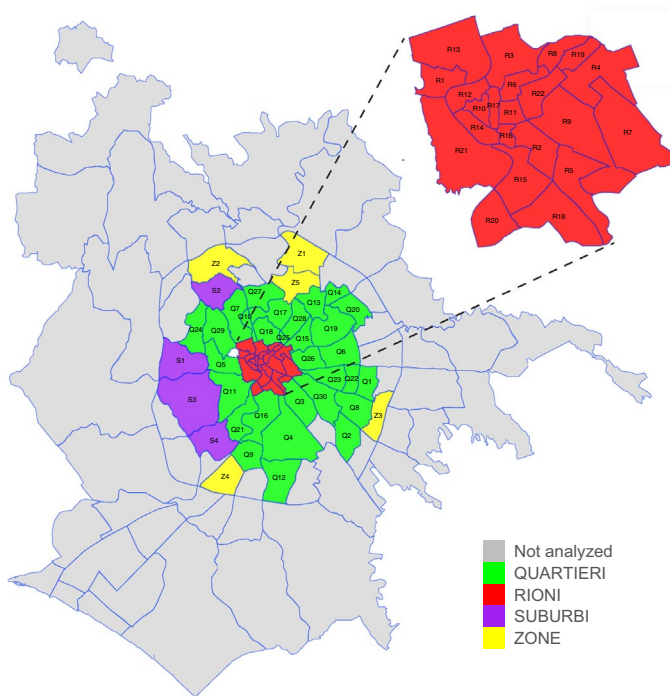


Fig. 1 Neighborhoods of Rome: in green (*Quartieri*), red (*Rioni*), purple (*Suburbi*) and yellow (*Zone*) the neighborhoods considered in the work, in grey those not analyzed. In the centroid of each analyzed neighborhood a corresponding label is printed. (Color figure online)

an official website of the Municipality of Rome. The *Open Data Roma Capitale* includes information on different topics, such as public transportation, libraries, local police, and nurseries. On the other side, the data related to *management*, *property*, *time* and *instrumental* have been provided by AirDNA LLC, a company that manages Airbnb data. Our analysis takes into consideration exclusively the data for one year, since the aim is to carry out a cross-section analysis capable of supporting hosts in the definition of a price strategy in line with the most relevant dimensions and the characteristic of the neighborhood where the accommodation is located. For data coherence reasons, we used the data of 2016.

3 Methodology

In order to define the two price indicators q and r for a specific neighborhood, we operate in three different phases. In the first phase the *Proportional Odds Model* framed in Vector Generalized Additive Model introduced by Yee (2015) is used to model the price by the variables of one dimension each time, and to estimate the probability for accommodations to belong to a specific class of price. In the second phase, the estimated probability is used to evaluate the ability of the model to correctly predict the price class of accommodations by three indexes defined in $[0, 1]$. Finally, in the third phase, these three indexes are combined and the two price indicators are obtained.

3.1 Price modeling

Notationally, let us consider $\mathbf{D} = [\mathbf{y}, \mathbf{x}_1, \dots, \mathbf{x}_p]$ the dataset made up of an ordinal outcome variable \mathbf{y} that can assume values in the set $\{1, 2, \dots, K\}$, where 1 corresponds to the lowest class and K to the highest one, and p independent variables \mathbf{x}_i each of them can be mapped to a set $M \in \mathcal{M}$ gathering variables of a specific dimension. We define a function $\nu(\cdot)$ that corresponds to the fitted probabilities of a Proportional Odds Model trained on bootstrap samples \mathbf{T} and tested on corresponding out-of-bag samples \mathbf{V} using \mathbf{y} as the dependent variable and those belonging to a set M as independent ones. The bootstrap strategy is carried out to reduce the uncertainty and avoid potential overfitting issues. Specifically, Proportional Odds Model is defined through logit

$$P(y \leq k | \mathbf{x}) = \eta_k(\mathbf{x}) \quad (1)$$

subject to the constraint

$$\eta_k(\mathbf{x}) = \beta_{(k)1}^* + \mathbf{x}'_{[-1]} \beta_{[-1:k]}^* \quad (2)$$

with $k = 1, \dots, K$, where k identifies the level of y and moves from 1 to K , $\mathbf{x}_{[-1]}$ is the \mathbf{x} with the first element deleted (the intercept), $*$ denotes the regression coefficients that have to be estimated (Yee 2015, p. 11).

The dataset \mathbf{T} , is a bootstrap sample of \mathbf{D} , that is it is created by sampling with replacement from \mathbf{D} the same number of observations. Instead, \mathbf{V} is made up of the out-of-bag observations, that are the observations of \mathbf{D} which are not included in \mathbf{T} .

The output of the function $\nu(\cdot)$ is a $N \times K$ matrix $\mathbf{P} = \nu(\mathbf{y} \sim M, \mathbf{T}, \mathbf{V})$, where N is the number of the observations that composed the dataset \mathbf{V} . Specifically, each row \mathbf{P}_i is a vector $(p_{i1}, p_{i2}, \dots, p_{iK})$ of probabilities, with $\sum_{j=1}^K p_{ij} = 1$, that express the ability of the variables belonging to the dimension M to explain the target variable \mathbf{y} (i.e. price). In fact, if the estimated probabilities are similar to each other, the prediction uncertainty is high and the dimension is not able to explain the price classes. On the contrary, if one probability is close to one, and the others to zero, the prediction uncertainty is low and the dimension explains very well the price classes.

3.2 Assessment of price class prediction

Referring to the notation above, the first phase consists in estimating a matrix \mathbf{P} for each set of variables M belonging to \mathcal{M} . In the second phase, for each dimension, the corresponding estimated matrix \mathbf{P} is used to compute three different indexes expressing the ability of the model to predict correctly the class of price through the variables of the dimension. More in detail, the first index, called θ , measures the concentration of the probabilities of \mathbf{P}_i and corresponds to the complementary of the Gini's Index normalized. Namely,

$$\theta_i = 1 - \left[\frac{(1 - \mathbf{P}_i \mathbf{P}'_i) \cdot K}{K - 1} \right]. \quad (3)$$

$\theta_i = 1$ means that the model assigns a probability of 100% to one class and 0% to the other $K - 1$ classes to be the price of the accommodation i . On the contrary, $\theta_i = 0$ means that the model gives the same probability to all classes to be the price of the accommodation i .

For assessing the goodness of fit of the model, another aspect to take into account is prediction correctness. Being the dependent variable ordinal, so the predicted class \hat{y} is not that characterized by the highest probability, but the median one. The absolute difference between \hat{y} and y provides the information of how wrong the prediction is: in fact, if the prediction is correct $|\hat{y} - y| = 0$, otherwise the higher value of the difference the more wrong the prediction is. Consequently, to assess the prediction correctness the following normalized index γ is considered

$$\gamma_i = 1 - \frac{|\hat{y}_i - y_i|}{K - 1}. \tag{4}$$

If γ_i assumes a value equal to 1 then the model has predicted correctly the class of the accommodation i . The lower value the higher difference between the correct class and the predicted one. The worst situation occurs when the lowest class (i.e. 1) is predicted while the highest one (i.e. K) was the true or vice versa.

The third index, called δ , measures the gap between the probability of the true class and that of the highest one, that is the probability of the predicted class. It computes the complementary absolute difference value between the probability fitted for the true class ($\mathbf{P}_{\hat{y}}$) and that for the predicted one (\mathbf{P}_{iy}). Specifically,

$$\delta_i = 1 - |\mathbf{P}_{\hat{y}} - \mathbf{P}_{iy}|. \tag{5}$$

If $\delta_i = 1$ then either the predicted class is the true one or they differ but have an identical probability. In both situations, the model provides the highest probability to the true class for the accommodation i . The lower δ_i the less likely the true class can be predicted.

3.3 Price indicators definition

Finally, the third phase aims to estimate the two price indicators able to measure the impact and the relative importance of the dimensions on the price determination. Specifically, for each dimension the values of the accommodations of three indexes introduced above are combined by multiplying them into a unique index that summarizes the capacity of the corresponding independent variables to explain the price. Then for all the dimensions, the values of the accommodations of this unique index are both averaged obtaining the first price indicators q organized in a vector \mathbf{q} , and aggregated according to the true classes of each accommodation by averaging, obtaining a descriptive matrix \mathbf{A} . The matrix \mathbf{A} is made up of K rows and a number of columns equal to the number of dimensions and provides information useful to analyze deeply the neighborhoods. It gives information about the capacity of each dimension to explain a specific class of price. Nonetheless, that is inconvenient for comparison aims between neighborhoods. Instead, the vector \mathbf{q} is made up of one single value q per dimension, and summarizes the information of \mathbf{A} . In fact, it corresponds to the weighted average of the values of the K price classes in \mathbf{A} . To sum up, it consists in a synthetic value expressing the impact of each dimension on the price determination and allows easy comparison between neighborhoods.

Algorithm 1 illustrates all steps needed to obtain \mathbf{q} and \mathbf{A} . Computationally, for each set of independent variables $M \in \mathcal{M}$, B iterations are carried out. At each iteration b , firstly the bootstrap sample \mathbf{T}^b and the corresponding out-of-bag one \mathbf{V}^b are generated and used as inputs of the function $v(\cdot)$, setting \mathbf{y} as the dependent variable and the variables belonging to one of the sets M as independent ones. Then, the matrix of probabilities \mathbf{P}^{Mb} obtained from $v(\cdot)$ is used to compute the θ index for each observation. The true classes \mathbf{y}^b of the

dataset \mathbf{V}^b is compared to the predicted ones $\hat{\mathbf{y}}^b$ to compute the γ index. Next, the probabilities associated with the true classes and the predicted ones are used to compute the δ index. Successively, a matrix \mathbf{U} stores the sum of the values obtained after multiplying the three indexes, according to the true class associated with the observations and the iterations. For computational reasons, in the algorithm we introduce an indicator function $I_k(\mathbf{y}^b)$, that generates a vector with elements equal to 1 if $y^b = k$ and 0 otherwise. The matrix \mathbf{Y} , instead, stores the frequencies of the true classes in each iteration. Finally, the results collected in \mathbf{U} are summed by iterations and then, using the frequencies contained in \mathbf{Y} , averaged to obtain the matrix \mathbf{A} and the vector \mathbf{q} . They store the indicator values of all dimensions concerning, respectively, the K price classes and their weighted combination.

Algorithm 1 Indicators definition

Input: \mathbf{D}

- 1: Set a $K \times |\mathcal{M}|$ empty matrix \mathbf{A} , a vector \mathbf{q} of length $|\mathcal{M}|$
- 2: **for** M in \mathcal{M} **do**
- 3: Set two $K \times B$ empty matrices \mathbf{U} and \mathbf{Y}
- 4: **for** $b = 1$ to B **do**
- 5: to generate randomly \mathbf{T}^b and \mathbf{V}^b from \mathbf{D}
- 6: $\mathbf{P}^{Mb} = \nu(\mathbf{y} \sim M, \mathbf{T}^b, \mathbf{V}^b)$
- 7: $\theta^{Mb} = \frac{K}{K-1} \left[\mathbf{1} - (\mathbf{P}^{Mb})^{\circ 2} \mathbf{1} \right]$
- 8: $\gamma^{Mb} = \mathbf{1} - (|\hat{\mathbf{y}}^b - \mathbf{y}^b| / (K - 1))$
- 9: $\delta^{Mb} = \left(\delta_i^{Mb} = 1 - \left| \mathbf{P}_{iy_i}^{Mb} - \mathbf{P}_{iy_i}^{Mb} \right| \right)$
- 10: **for** $k = 1$ to K **do**
- 11: $\mathbf{U}_{kb} = \mathbf{1}' [\theta^{Mb} \circ \gamma^{Mb} \circ \delta^{Mb} \circ I_k(\mathbf{y}^b)]$
- 12: $\mathbf{Y}_{kb} = \mathbf{1}' I_k(\mathbf{Y})$
- 13: **end for**
- 14: **end for**
- 15: $\mathbf{A}_{.M} = (\mathbf{U}\mathbf{1}) / (\mathbf{Y}\mathbf{1})$
- 16: $\mathbf{q}_M = (\mathbf{1}'\mathbf{U}\mathbf{1}) / (\mathbf{1}'\mathbf{Y}\mathbf{1})$
- 17: **end for**

Output: \mathbf{A} , \mathbf{q}

For convenience reasons, the information of \mathbf{q} of all neighborhoods are then combined into a matrix \mathbf{Q} , such that each row corresponds to a vector \mathbf{q} concerning a neighborhood. In such a way, the matrix \mathbf{Q} groups together the importance of the dimensions for each neighborhood, showing the different levels of impact of the dimensions on the price with respect to the neighborhoods.

A second normalized price indicator r is obtained from \mathbf{Q} : it measures the relative impact that each dimension has on a specific neighborhood with respect to others. The price indicator r is estimated for each neighborhood and each dimension, and it is stored in a matrix \mathbf{R} , where each column i is computed as follows

$$\mathbf{R}_{.i} = \frac{\mathbf{Q}_j^0 - \min(\mathbf{Q}_j^0)}{\max(\mathbf{Q}_j^0) - \min(\mathbf{Q}_j^0)} \quad (6)$$

with $\mathbf{Q}^0 = \mathbf{Q} - (\mathbf{Q}\mathbf{1}/|\mathcal{M}|)\mathbf{1}'$ expressing the variations of \mathbf{Q} from its row means. It defines how much the first price indicator of the dimensions of each neighborhood differs from the average: a low (high) value represents a small (big) relative importance of the dimension in

the neighborhood. It can assume values in a range $[0, 1]$. It assumes a value equal to one in the neighborhood where the dimension has the highest impact concerning the other neighborhoods, the value zero with the lowest impact. The proximity of the value 1 identifies an elevated relative impact of the dimension. It is a sort of rank, where the value one identifies the highest relative impact.

In fact, the indicator allows defining a rank for each dimension that identifies where a specific aspect has a major relevance. The value 1 identifies the neighborhood where the dimension has the highest impact with respect to the other neighborhoods. The other values, ranging from 1 to 0, identify the relevance of the dimension with respect to the other neighborhoods. The value of r can be considered as a position occupied by the neighborhoods with respect to the importance of a specific dimension in the price determination.

4 Results and discussion

For each neighborhood, we have calculated firstly the impact of each dimension for the five different levels of price (i.e. a matrix \mathbf{A} for each neighborhood); secondly, the average impact of each dimension on the price (i.e. price indicator q , stored in \mathbf{Q}); and, finally, the relative importance of dimensions within the neighborhood (i.e. price indicator r , stored in \mathbf{R}).

4.1 Example of price indicators estimation for a single neighborhood

We are now going to explain how a single neighborhood can be deeply analyzed using the price indicators mentioned above. Specifically, we will be analyzing the neighborhood of *Trastevere* (R21), one of the traditional tourist quarters located in the center of the city. The first information that we obtain can be found in the matrix \mathbf{A} , which shows how the impact of the six dimensions on price classes changes within the neighborhood. By investigating the results shown in Table 2 we can observe how the values of all dimensions increase, moving from the lowest price to the highest one. The low values recorded for the lowest price classes show how these dimensions cannot explain these specific classes. In other words, the price of cheaper accommodations is not affected by our six dimensions. On the contrary, the highest values recorded for the highest price classes indicate that the dimensions influence the price. For instance, the dimension *property* assumes the value 0.0466 for the first class and the value 0.1768 for the fifth class, highlighting that the latter class is affected by that dimension, whilst the former is much less affected. From a practical point of view, it means that if hosts want to set a very-high price they have to offer accommodations with a high number of rooms and bathrooms. Nonetheless, the fact that the very-low price class is almost not affected by the *property* dimension also means that guaranteeing high standards with respect to this dimension is a necessary but not sufficient condition to ensure a very-high price class: it can happen that, despite high standards at the *property* dimension, other elements can negatively influence the price.

Moreover, the two main price indicators q and r (Table 2) confirm that, in order to control the price as efficiently as possible, it is worthwhile for hosts to focus on the *property* dimension, since its overall importance is the highest ($q = 0.1038$). Instead, the second indicator r highlights that in this neighborhood the effect of the dimension *culture* ($r = 0.9646$) is one of the highest among all Rome's neighborhoods: consequently, if hosts

with accommodations in several neighborhoods want to focus on this dimension, they will know that in *Trastevere* they will get the best results.

4.2 Price indicator q

The values of the first price indicator in Fig. 2 and Table 4, show how the six dimensions impact differently on the price of accommodations in different neighborhoods. The dimension *transports* has a high impact in the neighborhoods located in the center of Rome and in the suburbs. The highest impact is registered in the Suburbs of *Torre Maura* (Z3) ($q = 0.3051$), located in the East of the city and recently included in a process of urban regeneration, and *Torrino* (Z4) ($q = 0.2850$), one of the most popular and livable places in Rome, with a variety of civil buildings, such as shopping centers, sport facilities, recreational centers, and buildings of worship. These suburbs are far away from the city center, and for this reason the presence of train or subway stops is crucial.

The dimension *crowd* has a high impact on the price of accommodation located in the center such as in the neighborhood of *Colonna* (R6) ($q = 0.2060$), a tourist quarter with well-known historical monuments and government buildings, and in suburban neighborhoods such as *Ponte Mammolo* (Q20) ($q = 0.3388$), which is a residential area of the city. The results highlight also how the presence of other tourist facilities, like hotels or Airbnb accommodations, increases the possibility of a price raise. They suggest that the presence of other accommodations in the neighborhood tend to attract more guests, since tourists might think that an area with more accommodations is also a one with more attractions, monuments, and services.

The dimension *culture* is significant in some of the neighborhoods located in the outskirts and in the center of the city. Generally, the impact of this variable is not high. However, in some neighborhoods the proximity to monuments seems to affect the the price of the accommodations. This means that the cultural elements are important in defining the price, but only if they are close to the accommodation. The dimension is significant in the above mentioned neighborhood of *Torre Maura* (Z3) ($q = 0.2561$), and in Rome's city center, in the neighborhood *Colonna* (R6) ($q = 0.2040$).

Higher values have been recorded for the dimension *property*. They underline how the booking of an accommodation is influenced by the property characteristics, and how the number of bedrooms and bathrooms influence prices. The highest impact of this dimension has been recorded in *Torre Maura* (Z3) where the importance is 0.5592; and *Castel Giubileo* (Z1), a neighborhood located in the Northeast of the city, with a value of 0.5096. These areas have been recently interested by an urban regeneration, that apparently improved the attraction of new Airbnb guests.

On the contrary, the dimension *management* is more relevant in the outskirts and, specifically, in the neighborhoods *Giuliano-Dalmata* (Q12) located at south of the city, *Ponte Mammolo* (Q20), *Castel Giubileo* (Z1) and *Val Melaina* (Z5). Their estimated values are respectively 0.3133, 0.5456, 0.4609 and 0.4306. This high relevance can be justified by the fact that guests book accommodations in the suburbs only if the services somehow enhance their experience. Services become in this case the main element to influence the level guests' satisfaction and, consequently, the price determination of future rents.

Similar results have been recorded for the dimension *time*, which explains the impact of tourism high season on the accommodation rate. This aspect impacts only on the outskirts, which means that the rates of the accommodations located in the city center are not influenced by the different days or months in which they are booked. On the contrary, the

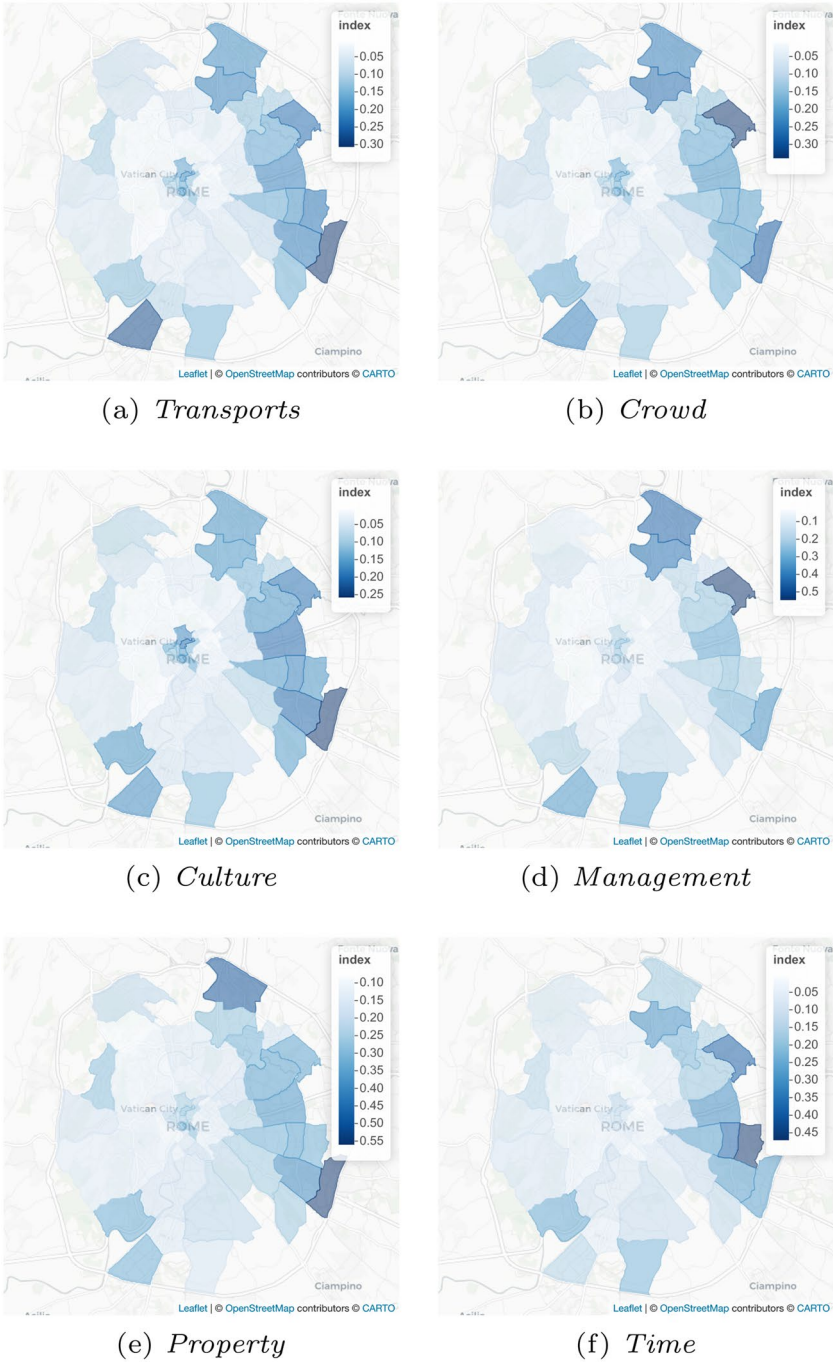


Fig. 2 Distribution of the first price indicator q of the six dimensions on the neighborhoods

tourist high season, which in Rome corresponds to all the weekends and to the months of March, April, May, September and October, causes a price increase in the Neighborhoods of *Alessandrino* (Q1) ($q = 0.4696$), *Ponte Mammolo* (Q20) ($q = 0.4011$) and *Preneestino Centocelle* (Q22) ($q = 0.3892$).

4.3 Price indicator r

Next, we have evaluated the relative importance of the dimensions among the neighborhoods. Figure 3 and Table 5 highlight how the dimensions *transports*, *crowd* and *culture* have a major relative importance in the central neighborhoods of the city and in the suburbs located at the east and south side. On the contrary, the dimensions *property*, *management* and *time* have major relative importance only in the outskirts located at the east and south side of the city.

Overall, we can state that the relative impact of the dimensions *property*, *management* and *time* is significant in the suburban neighborhoods. For instance, the dimension *property* assumes the maximum value ($r = 1.000$) in *Torre Maura* (Z3), where the distance from the center is overcome by staying in nice and bigger accommodations. The dimension *management* assumes the maximum value in the suburban neighborhood of *Ponte Mammolo* (Q20). In this case, the distance from the center is overcome providing a variety of services that justify higher prices. Instead, the dimension *time* assumes the maximum in *Alessandrino* (Q1), where the distance from the center allows increasing the price only in tourism high season.

The dimensions *transports*, *crowd* and *culture* affect price both in the central neighborhoods of Rome and in the outskirts. Although the dimension *transports* assumes the maximum in the suburb *Torrino* (Z4), one of the most popular and livable places in the capital of Italy, high values can be recorded even in central neighborhoods such as *Sant'Angelo* (R16) ($r = 0.8077$) and *Ponte* (R12) ($r = 0.7189$). Similarly, the dimension *crowd* assumes the maximum in the suburb *Monte Sacro Alto* (Q14), but high values can be recorded also in central neighborhoods such as *Campitelli* (R2) ($r = 0.9451$), one of the most visited neighborhoods for the presence of numerous cultural monuments. Finally, the dimension *culture* assumes its maximum value in *Ponte* (R12) and in the other central neighborhoods that are rich of monuments, museums and churches.

4.4 Cluster analysis

To conclude the investigation, we have used the information derived by the estimation of matrices of the indicators q and r to perform a hierarchical clustering using Ward's method to identify similarities among the neighborhoods. We have obtained four clusters, that have been characterized by the average of the two price indicators (hence \bar{q} and \bar{r}) of the six dimensions (Table 3). Figure 4 shows a contrast between the city center (*green* and *blu* clusters) and the suburbs (*yellow* and *red* clusters).

The *green* cluster includes tourist and residential neighborhoods that are rich in monuments. Furthermore, we can state that in this area the major impact on the price is generated by the dimensions *property* ($\bar{q} = 0.2193$) and *management* ($\bar{q} = 0.1597$). Instead, the major relative impact on the price is generated by the dimensions *transports* ($\bar{r} = 0.6665$), *crowd* ($\bar{r} = 0.8236$) and *culture* ($\bar{r} = 0.8743$). Consequently, all dimensions should be considered by hosts while fixing the price for their accommodations.

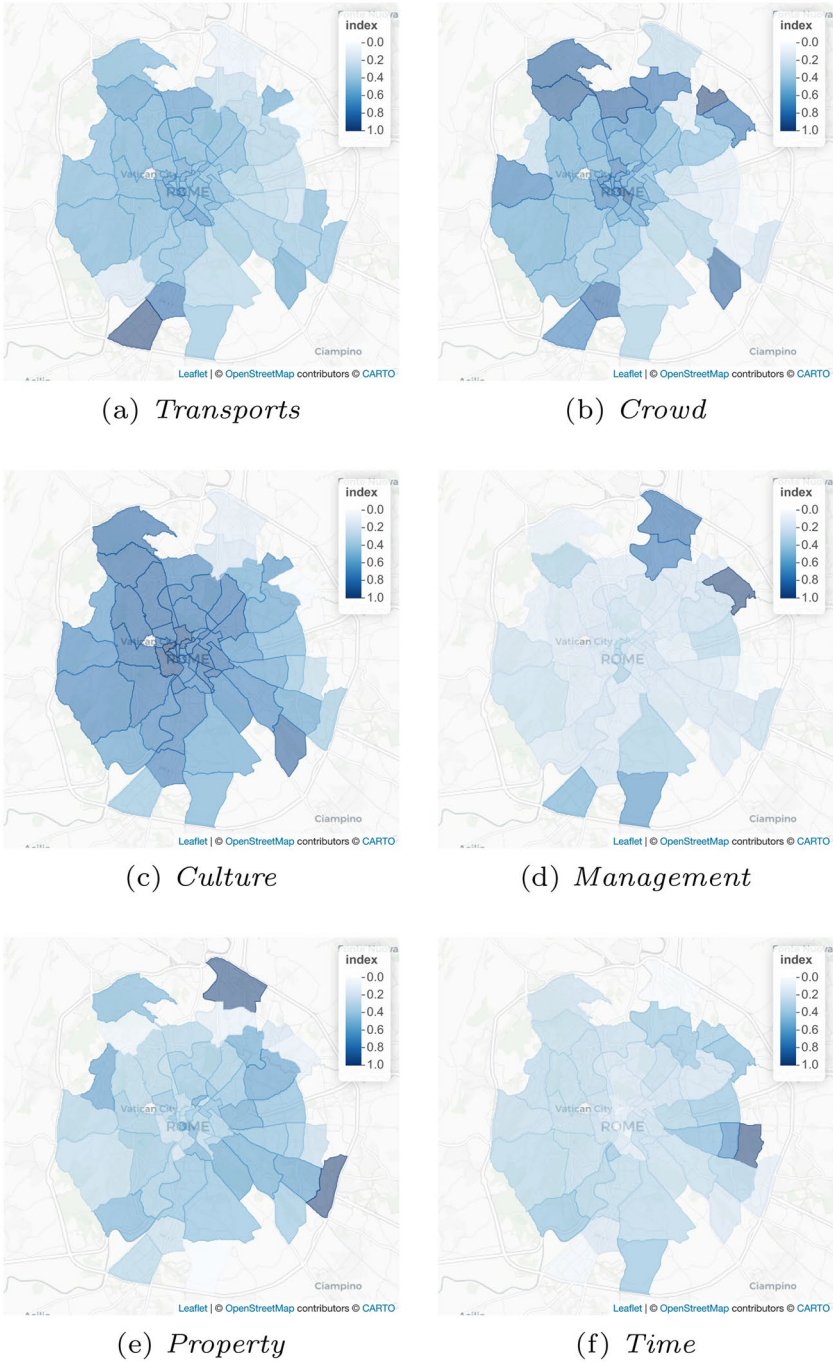
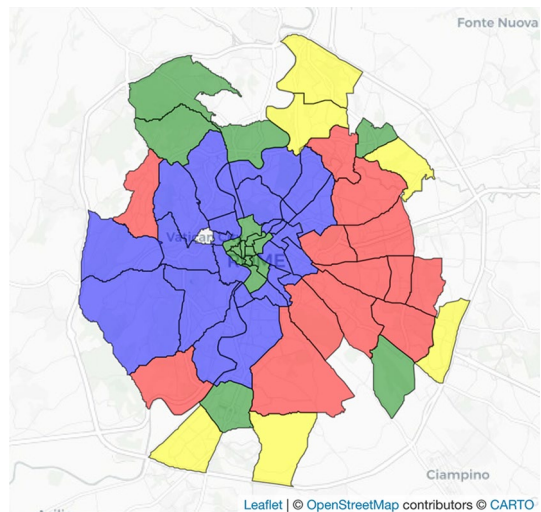


Fig. 3 Distribution of the second price indicator r of the six dimensions on the neighborhoods

Table 2 Matrix *A* and the two price indicators *q* and *r* estimated for the neighborhood *Trastevere* (R21)

Price	Transports	Crowd	Property	Management	Culture	Time
A						
1	0.0103	0.0107	0.0466	0.0162	0.0098	0.0163
2	0.0224	0.0229	0.0626	0.0285	0.0219	0.0308
3	0.0370	0.0380	0.0661	0.0493	0.0362	0.0470
4	0.0531	0.0537	0.0982	0.0741	0.0516	0.0466
5	0.0378	0.0389	0.1768	0.0579	0.0372	0.0295
<i>q</i>						
	0.0385	0.0392	0.1038	0.0542	0.0376	0.0392
<i>r</i>						
	0.6771	0.7684	0.2994	0.2170	0.9646	0.3159

Fig. 4 Distribution of the four clusters in the neighborhoods**Table 3** Average of the dimensions of the indicators *q* and *r* by cluster belonging neighborhoods

Indicator	Cluster	Dimensions					
		Transports	Crowd	Property	Management	Culture	Time
\bar{q}	Blue	0.0346	0.0362	0.1425	0.0658	0.0293	0.0459
	Green	0.1285	0.1347	0.2193	0.1597	0.1168	0.1023
	Red	0.1456	0.1467	0.3104	0.1997	0.1350	0.2317
	Yellow	0.2343	0.2582	0.3670	0.4110	0.1816	0.2586
\bar{r}	Blue	0.5949	0.6285	0.4337	0.2354	0.8526	0.3152
	Green	0.6665	0.8236	0.4012	0.2732	0.8743	0.2256
	Red	0.4068	0.2753	0.5697	0.2280	0.6316	0.4742
	Yellow	0.3950	0.5721	0.4271	0.7118	0.3105	0.2726

The *blue* cluster have similar characteristics with respect to the previous group. Government offices and embassies are located in the neighborhoods included in this cluster. The dimensions that affect prices the most are *property* ($\bar{q} = 0.1425$) and *management* ($\bar{q} = 0.0658$). The dimensions with the highest relative importance are, in descending

order, *culture* ($\bar{r} = 0.8526$), *crowd* ($\bar{r} = 0.6285$) and *transports* ($\bar{r} = 0.5949$). In this case, the proximity to cultural monuments gives hosts the possibility to increase prices.

The *red* cluster consists of residential neighborhoods, where also university buildings and the *Cinecittà* film studios are located. The dimensions *property* ($\bar{q} = 0.3104$) and *time* ($\bar{q} = 0.2317$) are the most relevant in this area. The relevance of the dimension *time* is confirmed also by the results of \bar{r} , which assumes the highest value in this cluster. The type (apartment or single room) and size of the accommodation strongly influence price. The same happens for high tourism seasons, since during this days or months hosts have the possibility to increase the rate for their apartments.

Finally, the *yellow* cluster includes the suburban neighborhoods recently redeveloped and far from the city center. The price of accommodations is generally lower in this areas and the dimensions that influence the rates are *property* ($\bar{q} = 0.3670$) and *management* ($\bar{q} = 0.4110$). However, also the other dimensions have a strong relevance and their impact ranges from the $\bar{q} = 0.1816$ of the dimension *culture* to the $\bar{q} = 0.2586$ of the dimension *time*. The major relative impact is recorded for the dimensions *management* ($\bar{r} = 0.7118$) and *crowd* ($\bar{r} = 0.5721$).

5 Conclusions

5.1 Theoretical implication

The new forms of accommodations and the necessity to manage them properly generate a new challenge: the determination of the correct price and the evaluation of its determinants. For this reason, we have defined two price indicators, called q and r , to evaluate the relevance of different dimensions on the price and to highlight the differences among geographical areas.

The first indicator q allows identifying the relevance of each dimension on price determination, attributing a weight to each dimension. This weight allows to rank the different dimensions and to measure the impact of each group of variables on the price. On the other side, the indicator r measures the relative importance that each dimension has in specific neighborhoods.

Some innovative elements can be identified in the estimation of the two indicators. Firstly, we have estimated the indicators combining different methodologies and operating in three steps. In the first step, we have used the Proportional Odds Model to evaluate the impact that each dimension have on different levels of price. In this way, for each geographical area taken into account, we have estimated the impact of dimensions on five different levels of price. In the second step, the estimated probability is used to evaluate the ability of the model to predict correctly the price class of accommodations by three indexes defined in $[0, 1]$. Finally, the three indexes are combined into two price indicators.

Secondly, we have obtained different outputs that can be used to analyze the Airbnb price determination. Specifically, the matrix **A** summarizes the impact that each dimension has on different levels of price and for each analyzed neighborhood. The matrix is able to create a useful framework to comprehend the price behavior in a specific geographical area. The two indicators q and r allow estimating the relevance of each dimension on price determination, offering two specific pieces of information: more in detail, the indicator q measures the different impact of the dimensions on the price with respect to the single neighborhoods; whereas the indicator r identifies the relative impact that

each dimension has on a specific neighborhood with respect to the others. In other words, the two indicators allow evaluating the relevance of each dimension inside the neighborhood and with respect to the other neighborhoods. In this way, we can analyze the price determination for a specific area and, at the same time, we can also compare the results in different neighborhoods, in order to comprehend where a specific dimension have a major relevance.

Finally, the proposed price indicators have not been developed according to the features of a specific location, and therefore they can be generalized. Consequently, they can be used for all kinds of locations such as countries, regions, cities, neighborhoods, etc., allowing also the comparison among them. In the specific case of Rome, the results have demonstrated that the indicators record the differences among neighborhoods: that prices change with respect to the location of the accommodation: and that there are differences between the city center and the outskirts.

5.2 Practical implications

From a practical point of view, it is crucial to comprehend which aspects really influence the price and which are irrelevant. Moreover, knowing the relevance of dimensions for each neighborhood supports hosts in the correct definition of prices, in order to become more competitive on online platforms such as Airbnb.

Our results can indicate to each host which dimensions really impact on the price of accommodations, based on where they are, on the services included and on the time of the year in which they are booked. This way, hosts have a tool to help them establish a price in line with the market. Specifically, the *old* hosts, that is those who already have experience on the Airbnb platform, can better understand the dimensions that affect the rate and, eventually, modify it accordingly. On the other hand, *new* hosts can use the results of the model to establish the best renting class price for their accommodation. Moreover, the estimation for the neighborhood allows to better explain the price and its determinants taking into account the geographical proximity. We suppose that accommodations close to each other are subject to the same impacts. These externalities can be also perceived by tourists, and they can similarly impact on the price. Defining a model able to represent the importance of proximity is relevant in order to evaluate the impact with respect to geographical characteristics. Then, the identification of different rankings for each neighborhood allows to discover how the rate of Airbnb accommodation is influenced by the local environment and the elements located in the area. Specifically, transportation facilities, monuments, and the proximity of other accommodations affect the management of an accommodation and its price. This impact demonstrates the need to insert in the model variables that are not directly related to accommodation management.

To conclude, we have defined two indicators able to estimate the price determinants and the differences among different neighborhoods, offering an important tool to support hosts in the definition of one of the most important aspects in the management of an Airbnb accommodation, i.e. its price.

5.3 Limitations

The study has some limitations. Firstly, we have considered six dimensions in the price determination. We believe that more aspects can be included in the model. For instance, a variable related to weather conditions could be also taken into account. We cannot be

sure, for instance, that the impact of the proximity of public transportation facilities is the same in all seasons: we might suppose, on the contrary, that a bus or a train stop nearby an accommodation is more relevant in winter, when rain and cold push tourists to spend as little time as possible on the outside. Secondly, we have analyzed the impact of the determinants of price only for one city. Taking into account other cities might be interesting in order to comprehend if the weights of the dimensions can be different in other tourist areas. Future research should therefore consider more dimensions and more destinations.

A Matrices Q and R

See Table 4.

Table 4 Matrix Q

	Neighborhood	Label	Transports	Crowd	Property	Management	Culture	Time
<i>Quartieri</i>	Alessandrino	Q1	0.2349	0.1998	0.3212	0.2156	0.1650	0.4696
	Appio Claudio	Q2	0.1591	0.1752	0.2676	0.1736	0.1567	0.1360
	Appio Latino	Q3	0.0514	0.0512	0.2232	0.0992	0.0501	0.0751
	Ardeatino	Q4	0.0742	0.0811	0.2247	0.1790	0.0666	0.1295
	Aurelio	Q5	0.0126	0.0119	0.1239	0.0346	0.0111	0.0338
	Collatino	Q6	0.2199	0.2213	0.3839	0.3198	0.2084	0.3159
	Della Vittoria	Q7	0.0148	0.0155	0.1121	0.0376	0.0134	0.0353
	Don Bosco	Q8	0.2252	0.2084	0.4055	0.2784	0.2099	0.2786
	Europa	Q9	0.1022	0.0893	0.1491	0.1057	0.0580	0.0536
	Flaminio	Q10	0.0197	0.0205	0.1058	0.0462	0.0108	0.0144
	Gianicolense	Q11	0.0121	0.0129	0.1316	0.0398	0.0108	0.0390
	Giuliano Dalmata	Q12	0.1481	0.1510	0.1701	0.3133	0.1240	0.2282
	Monte Sacro	Q13	0.1350	0.1312	0.3182	0.1908	0.1280	0.2298
	Monte Sacro Alto	Q14	0.1200	0.1449	0.1775	0.1384	0.0948	0.1712
	Nomentano	Q15	0.0575	0.0585	0.1827	0.0931	0.0575	0.0662
	Ostiense	Q16	0.0444	0.0451	0.1760	0.0749	0.0422	0.0771
	Parioli	Q17	0.0434	0.0502	0.1527	0.0716	0.0334	0.0370
	Pianciano	Q18	0.0196	0.0252	0.1373	0.0670	0.0142	0.0106
	Pietralata	Q19	0.1743	0.1928	0.3627	0.2456	0.1682	0.1957
	Ponte Mammolo	Q20	0.2448	0.3388	0.3539	0.5456	0.2019	0.4011
	Portuense	Q21	0.0474	0.0482	0.1681	0.0982	0.0455	0.0997
	Prenestino Centocelle	Q22	0.2022	0.2096	0.3577	0.2586	0.1856	0.3892
	Prenestino Labicano	Q23	0.1678	0.1662	0.3220	0.2261	0.1644	0.2784
	Primavalle	Q24	0.1181	0.1037	0.2889	0.1325	0.0981	0.1509
	Salario	Q25	0.0237	0.0235	0.1531	0.0346	0.0112	0.0224
	Tiburtino	Q26	0.0544	0.0553	0.2324	0.0900	0.0529	0.1187
	Tor Di Quinto	Q27	0.0629	0.0714	0.1478	0.0648	0.0373	0.0630
	Trieste	Q28	0.0285	0.0302	0.1418	0.0631	0.0264	0.0515
	Trionfale	Q29	0.0142	0.0143	0.1104	0.0391	0.0131	0.0399
	Tuscolano	Q30	0.0925	0.0952	0.2622	0.1412	0.0931	0.1223

Table 4 (continued)

	Neighborhood	Label	Transports	Crowd	Property	Management	Culture	Time
<i>Rioni</i>	Borgo	R1	0.0305	0.0317	0.1315	0.0661	0.0257	0.0348
	Campitelli	R2	0.1487	0.1726	0.2951	0.1911	0.1424	0.0869
	Campo Marzio	R3	0.1781	0.1833	0.2604	0.2086	0.1778	0.1476
	Castro Pretorio	R4	0.0104	0.0110	0.1437	0.0585	0.0092	0.0077
	Celio	R5	0.0387	0.0422	0.1243	0.0666	0.0355	0.0537
	Colonna	R6	0.2055	0.2060	0.2917	0.2667	0.2040	0.1427
	Esquilino	R7	0.0266	0.0266	0.1358	0.0683	0.0257	0.0364
	Ludovisi	R8	0.0520	0.0513	0.2097	0.0814	0.0429	0.0465
	Monti	R9	0.0284	0.0293	0.1283	0.0566	0.0275	0.0301
	Parione	R10	0.1068	0.1131	0.2345	0.1471	0.1074	0.0930
	Pigna	R11	0.1624	0.1730	0.2551	0.2384	0.1518	0.1326
	Ponte	R12	0.1365	0.1346	0.2119	0.1472	0.1349	0.1026
	Prati	R13	0.0180	0.0188	0.1385	0.0556	0.0171	0.0353
	Regola	R14	0.1040	0.1076	0.1845	0.1308	0.1027	0.0693
	Ripa	R15	0.0833	0.0759	0.1639	0.1493	0.0582	0.0499
	Sant'Angelo	R16	0.1955	0.1805	0.3162	0.2002	0.1597	0.0992
	Sant'Eustachio	R17	0.1275	0.1304	0.2181	0.1518	0.1234	0.1127
	San Saba	R18	0.0749	0.0582	0.2020	0.0767	0.0412	0.0253
	Sallustiano	R19	0.0300	0.0355	0.1111	0.0630	0.0223	0.0109
	Testaccio	R20	0.0554	0.0578	0.1238	0.0764	0.0549	0.1185
	Trastevere	R21	0.0385	0.0392	0.1038	0.0542	0.0376	0.0392
	Trevi	R22	0.1378	0.1377	0.2504	0.1820	0.1262	0.1045
<i>Suburbi</i>	Aurelio	S1	0.0674	0.0872	0.1716	0.1153	0.0570	0.0920
	Della Vittoria	S2	0.0737	0.0854	0.0842	0.1366	0.0632	0.0823
	Gianicolense	S3	0.0560	0.0593	0.1426	0.1053	0.0459	0.0901
	Portuense	S4	0.1429	0.1910	0.3329	0.2197	0.1648	0.2578
<i>Zone</i>	Castel Giubileo	Z1	0.2050	0.2511	0.5096	0.4609	0.1687	0.1845
	Tomba Di Nerone	Z2	0.0810	0.1083	0.2202	0.0829	0.0860	0.0912
	Torre Maura	Z3	0.3051	0.2934	0.5592	0.3649	0.2561	0.2733
	Torrino	Z4	0.2850	0.2435	0.3255	0.3506	0.1749	0.1568
	Val Melania	Z5	0.2178	0.2714	0.2837	0.4306	0.1639	0.3077

Table 5 Matrix **R**

	Neighborhood	Label	Transports	Crowd	Property	Management	Culture	Time
<i>Quartieri</i>	Alessandrino	Q1	0.5315	0.0000	0.3071	0.0000	0.3170	1.0000
	Appio Claudio	Q2	0.6364	0.9087	0.4595	0.1907	0.9145	0.2233
	Appio Latino	Q3	0.4744	0.3828	0.6371	0.2384	0.7654	0.3043
	Ardeatino	Q4	0.3884	0.3234	0.4988	0.4211	0.6361	0.3686
	Aurelio	Q5	0.5880	0.5832	0.4440	0.1949	0.8738	0.3437
	Collatino	Q6	0.3380	0.1535	0.5280	0.3750	0.5581	0.4772
	Della Vittoria	Q7	0.6030	0.6320	0.3938	0.2065	0.8892	0.3481
	Don Bosco	Q8	0.4585	0.1209	0.6638	0.2513	0.6465	0.3920
	Europa	Q9	0.8502	0.8958	0.3178	0.2594	0.8143	0.2317
	Flaminio	Q10	0.6547	0.7287	0.3748	0.2482	0.8844	0.2876
	Gianicolense	Q11	0.5610	0.5553	0.4640	0.2034	0.8488	0.3505
	Giuliano Dalmata	Q12	0.4690	0.4161	0.0000	0.7051	0.5928	0.4814
	Monte Sacro	Q13	0.3721	0.1429	0.6282	0.2163	0.6243	0.4875
	Monte Sacro Alto	Q14	0.6201	1.0000	0.2344	0.1974	0.7309	0.4528
	Nomentano	Q15	0.5647	0.5649	0.4901	0.2374	0.8623	0.2942
	Ostiense	Q16	0.5358	0.5074	0.5012	0.2017	0.8182	0.3587
	Parioli	Q17	0.6187	0.7446	0.4530	0.2360	0.8409	0.2688
	Pianciano	Q18	0.5826	0.6626	0.4683	0.2938	0.8402	0.2455
	Pietralata	Q19	0.4096	0.5228	0.6708	0.2980	0.6670	0.2694
	Ponte Mammolo	Q20	0.0000	0.8238	0.1068	1.0000	0.0000	0.5271
	Portuense	Q21	0.4990	0.4402	0.4343	0.2630	0.7848	0.4055
	Prenestino Centocelle	Q22	0.2879	0.1446	0.4637	0.1742	0.4722	0.7456
	Prenestino Labicano	Q23	0.3781	0.1853	0.5088	0.2297	0.6567	0.5404
	Primavalle	Q24	0.5485	0.3192	0.6739	0.1438	0.6994	0.3640
	Salario	Q25	0.6206	0.6516	0.5390	0.1680	0.8247	0.2858
	Tiburtino	Q26	0.4301	0.3148	0.6384	0.1659	0.7207	0.4146
	Tor Di Quinto	Q27	0.6918	0.9039	0.3906	0.1694	0.7978	0.3204
	Trieste	Q28	0.5648	0.5746	0.4397	0.2331	0.8471	0.3399
	Trionfale	Q29	0.5960	0.6095	0.3848	0.2109	0.8844	0.3616
	Tuscolano	Q30	0.4620	0.4007	0.6214	0.2357	0.7678	0.3184

Table 5 (continued)

	Neighborhood	Label	Transports	Crowd	Property	Management	Culture	Time
<i>Rioni</i>	Borgo	R1	0.6067	0.6448	0.4111	0.2593	0.8679	0.2979
	Campitelli	R2	0.5976	0.9451	0.5982	0.2817	0.8475	0.0835
	Campo Marzio	R3	0.6697	0.8176	0.3674	0.2722	0.9621	0.2138
	Castro Pretorio	R4	0.5552	0.5413	0.5190	0.2823	0.8444	0.2538
	Celio	R5	0.6174	0.6968	0.3521	0.2342	0.8897	0.3364
	Colonna	R6	0.6745	0.7599	0.3864	0.3975	0.9575	0.1129
	Esquilino	R7	0.5781	0.5762	0.4301	0.2687	0.8687	0.3035
	Ludovisi	R8	0.5630	0.5385	0.6270	0.2114	0.7940	0.2483
	Monti	R9	0.6161	0.6581	0.4118	0.2348	0.9054	0.2937
	Parione	R10	0.5766	0.6611	0.5073	0.2623	0.8784	0.2275
	Pigna	R11	0.6047	0.7716	0.3748	0.4199	0.8233	0.1885
	Ponte	R12	0.7189	0.8072	0.3654	0.2187	1.0000	0.2232
	Prati	R13	0.5588	0.5508	0.4666	0.2420	0.8500	0.3192
	Regola	R14	0.6857	0.8230	0.3683	0.2657	0.9701	0.2069
	Ripa	R15	0.6786	0.6572	0.3646	0.4185	0.7879	0.2078
	Sant'Angelo	R16	0.8077	0.7889	0.6066	0.2417	0.8348	0.0620
	Sant'Eustachio	R17	0.6552	0.7583	0.3940	0.2399	0.9200	0.2574
	San Saba	R18	0.7439	0.6477	0.5980	0.1962	0.7883	0.1838
	Sallustiano	R19	0.6628	0.8086	0.3585	0.2787	0.9008	0.2469
	Testaccio	R20	0.5850	0.6220	0.2612	0.1895	0.8784	0.4760
	Trastevere	R21	0.6771	0.7684	0.2994	0.2170	0.9646	0.3159
	Trevi	R22	0.6390	0.6857	0.4782	0.3107	0.8493	0.1917
<i>Suburbi</i>	Aurelio	S1	0.5450	0.7913	0.3903	0.2758	0.7668	0.3367
	Della Vittoria	S2	0.6752	0.9173	0.0661	0.4047	0.8922	0.3401
	Gianicolense	S3	0.5741	0.6140	0.3317	0.2969	0.7973	0.3790
	Portuense	S4	0.2092	0.5686	0.5660	0.2146	0.6790	0.4832
<i>Zone</i>	Castel Giubileo	Z1	0.0854	0.3125	0.9820	0.8655	0.1311	0.0000
	Tomba Di Nerone	Z2	0.5483	0.9007	0.5403	0.0938	0.8832	0.2920
	Torre Maura	Z3	0.5004	0.2688	1.0000	0.3001	0.4401	0.1384
	Torrino	Z4	1.0000	0.7725	0.3743	0.5864	0.4752	0.0410
	Val Melania	Z5	0.3148	0.8390	0.0995	0.8139	0.2240	0.4479

Funding Open access funding provided by Università degli Studi di Cagliari within the CRUI-CARE Agreement. Research funding was provided by the Regione Autonoma della Sardegna under the Grant Pacchetti Integrati di Agevolazione Industria, Artigianato e Servizi, PIA–2013 No. 282/13.

Declarations

Conflict of interest All the authors declare they have no financial interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not

permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aznar, J.P., Sayeras, J.M., Segarra, G., et al.: Airbnb landlords and price strategy: have they learnt price discrimination from the hotel industry? evidence from barcelona. *Int. J. Tour. Sci.* **18**(1), 16–28 (2018)
- Barnes, S.J., Kirshner, S.N.: Understanding the impact of host facial characteristics on airbnb pricing: Integrating facial image analytics into tourism research. *Tour. Manag.* **83**(104), 235 (2021)
- Cai, Y., Zhou, Y., Scott, N., et al.: Price determinants of airbnb listings: evidence from hong kong. *Tour. Anal.* **24**(2), 227–242 (2019)
- Celata, F., Romano, A.: Overtourism and online short-term rental platforms in italian cities. *J. Sustain. Tour.* **30**(5), 1020–1039 (2022)
- Chattopadhyay, M., Mitra, S.K.: Do airbnb host listing attributes influence room pricing homogenously? *Int. J. Hosp. Manag.* **81**, 54–64 (2019)
- Chen, Y., Xie, K.: Consumer valuation of airbnb listings: a hedonic pricing approach. *Int. J. Contemp. Hosp. Manag.* **29**(9), 2405–2424 (2017)
- Dudás, G., Boros, L., Kovalcsik, T., et al.: The visualisation of the spatiality of airbnb in budapest using 3-band raster representation. *Geogr. Tech.* **12**(1), 23–30 (2017)
- Dudás, G., Kovalcsik, T., Vida, G., et al.: Price determinants of airbnb listing prices in lake balaton touristic region, hungary. *Eur. J. Tour. Res.* **24**(10), 1–18 (2020)
- Falk, M., Larpin, B., Scaglione, M.: The role of specific attributes in determining prices of airbnb listings in rural and urban locations. *Int. J. Hosp. Manag.* **83**, 132–140 (2019)
- Faye, B.: Methodological discussion of airbnb's hedonic study: a review of the problems and some proposals tested on bordeaux city data. *Ann. Tour. Res.* **86**(103), 079 (2021)
- Gibbs, C., Guttentag, D., Gretzel, U., et al.: Pricing in the sharing economy: a hedonic pricing model applied to airbnb listings. *J. Travel Tour. Mark.* **35**(1), 46–56 (2018)
- Gunter, U.: What makes an airbnb host a superhost? empirical evidence from san francisco and the bay area. *Tour. Manag.* **66**, 26–37 (2018)
- Gutiérrez, J., García-Palomares, J.C., Romanillos, G., et al.: The eruption of airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in barcelona. *Tour. Manag.* **62**, 278–291 (2017)
- Kakar, V., Franco, J., Voelz, J., et al.: Effects of host race information on airbnb listing prices in san francisco. *Munich Personal RePEc Archive (MPRA)* (2016)
- Kalehbasti, P.R., Nikolenko, L., Rezaei, H.: Airbnb price prediction using machine learning and sentiment analysis. *arXiv preprint arXiv:1907.12665* (2019)
- Lawani, A., Reed, M.R., Mark, T., et al.: Reviews and price on online platforms: evidence from sentiment analysis of airbnb reviews in boston. *Reg. Sci. Urban Econ.* **75**, 22–34 (2019)
- Lorde, T., Jacob, J., Weekes, Q.: Price-setting behavior in a tourism sharing economy accommodation market: a hedonic price analysis of airbnb hosts in the caribbean. *Tour. Manag. Perspect.* **30**, 251–261 (2018)
- Ma, X., Hancock, J.T., Lim Mingjie, K., et al.: Self-disclosure and perceived trustworthiness of airbnb host profiles. In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pp. 2397–2409 (2017)
- Moreno-Izquierdo, L., Egorova, G., Peretó Rovira, A., et al.: Exploring the use of artificial intelligence in price maximisation in the tourism sector: its application in the case of airbnb in the valencian community. *Investigaciones Regionales - J. Reg. Res.* **42**, 113–128 (2018)
- Neumann, J., Gutt, D.: A homeowner's guide to airbnb: theory and empirical evidence for optimal pricing conditional on online ratings. *Association for Information Systems AIS Electronic Library (AISeL)* (2017)
- Önder, I., Weismayer, C., Gunter, U.: Spatial price dependencies between the traditional accommodation sector and the sharing economy. *Tour. Econ.* **25**(8), 1150–1166 (2019)
- Perez-Sanchez, V.R., Serrano-Estrada, L., Marti, P., et al.: The what, where, and why of airbnb price determinants. *Sustainability* **10**(12), 4596 (2018)
- Quattrone, G., Proserpio, D., Quercia, D., et al.: Who benefits from the sharing economy of airbnb? In: *Proceedings of the 25th international conference on world wide web, International World Wide Web Conferences Steering Committee*, pp. 1385–1394 (2016)

- Rodríguez-Pérez de Arenaza, D., Hierro, L.Á., Patiño, D.: Airbnb, sun-and-beach tourism and residential rental prices. the case of the coast of andalusia (spain). *Current Issues in Tourism* pp 1–18 (2019)
- Romano, A.: The shifting geographies of digital intermediation: the effects of the covid-19 pandemic on short-term rentals in italian cities. *Dig. Geogr. Soc.* **2**(100), 019 (2021)
- Sainaghi, R., Abrate, G., Mauri, A.: Price and revpar determinants of airbnb listings: convergent and divergent evidence. *Int. J. Hosp. Manag.* **92**(102), 709 (2021)
- Skalska, T.: Sharing economy in the tourism market: opportunities and threats. *KNUV* **4**(54), 248–260 (2017)
- Tang, L.R., Kim, J., Wang, X.: Estimating spatial effects on peer-to-peer accommodation prices: towards an innovative hedonic model approach. *Int. J. Hosp. Manag.* **81**, 43–53 (2019)
- Teubner, T., Hawlitschek, F., Dann, D.: Price determinants on airbnb: How reputation pays off in the sharing economy. *J. Self Gov. Manag. Econ.* **5**(4), 53 (2017)
- Voltes-Dorta, A., Sánchez-Medina, A.: Drivers of airbnb prices according to property/room type, season and location: A regression approach. *J. Hosp. Tour. Manag.* **45**, 266–275 (2020)
- Wang, D., Nicolau, J.L.: Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on airbnb. com. *Int. J. Hosp. Manag.* **62**, 120–131 (2017)
- Ye, P., Qian, J., Chen, J., et al.: Customized regression model for airbnb dynamic pricing. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 932–940 (2018)
- Yee, T.W.: *Vector generalized linear and additive models: with an implementation in R*. Springer (2015)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.