Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n1p211

A statistical framework for Airbnb hosts and Superhosts

By Contu

Published: 20 May 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 01, May 2022, 211-231
DOI: 10.1285/i20705948v15n1p211

A statistical framework for Airbnb hosts and Superhosts

Giulia Contu*

University of Cagliari
Department of Economics and Management

Published: 20 May 2022

We propose a statistical framework in order to investigate the Airbnb hosts activities. We aim to propose an extended model able firstly to comprehend which variables can impact on the hosts' activity and; secondly to identify a guide that can support the hosts in the constant effort to reach the best performances and to become a Superhost. The framework uses two different models, the logistic regression model and the bivariate logistic regression model. Three groups of variables are taken into account. They are the attributes that Airbnb uses to assign the Superhost badge, the managerial aspects, and the characteristics of the accommodations. The analysis is focused on the hosts operating in the Italian most visited cities. Our findings show the capacity of the framework to identify the variables that affect the hosts' performance. They are the number of reviews, the services, and the typology of the rented accommodation, that affect the hosts' performance. The results show how the framework can be used as managerial support for the hosts.

keywords: Airbnb, hosts, superhost badge, VGAM, logistic regression model

1 Introduction

Airbnb is a company operating as an online marketplace for peer-to-peer accommodation rental services (Fagerstrøm et al., 2017, p. 123). It is one of the most famous online platforms where it is possible to book single rooms or whole apartments for one or more days (Edelman and Luca, 2014; Liu and Mattila, 2017; Malazizi et al., 2018; Choi et al.,

*Corresponding author: giulia.contu@unica.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

2015). It has been founded in 2008 in San Francisco and has grown significantly during the years, becoming an important tool for tourists who look for a different, often less expensive, kind of accommodation (Guttentag, 2015; Zervas et al., 2015; Lee and Kim, 2018).
Today, it operates in more than 65,000 cities and 191 countries and it sells millions of room nights for tourists around the globe (Aznar et al., 2016). In 2016, 160 million guests used the platform to book accommodation. In the same year, more than three million announcements were published. The number of hosts is equal to 2,9 million and an average of 800,000 overnight stays are spent each day in rooms or apartments booked on Airbnb. The number of people that subscribe as hosts increases at a rate of around 14,000 new hosts per month (`https://www.airbnb.com/host/homes?from\_footer=1\&locale=en`) (accessed on 20/07/2021). Becoming a host is easy. It is only necessary to decide to share a proper space (a room or an entire apartment) and to enroll on the Airbnb platform. The best hosts are awarded with the badge of Superhost. Since 2014, the Superhost program has recognized over 400,000 hosts (`https://www.airbnb.co.uk/host/homes`) (accessed on 20/07/2021). Airbnb declares that to be a Superhost it is necessary for the previous 12 months: (`https://www.airbnb.com/Superhost?locale=en`) (accessed on 27/07/2021): to host at least 10 stays a year; to honor every reservation unless there is an extenuating circumstance; to answer the guests' request within 24 hours at least 90% of the time; to achieve a 4.8+ overall rating.
Airbnb can not control all single hosts, but it can incentivize the hosts' activities through the badge in order to improve the hosts' performance and satisfy the guests' needs.
Few articles have focused on the study of the hosts' activity in itself, and of the elements that can influence their performance and improve the possibility to become a Superhost. Two studies written by Gunter (2018) and Contu et al. (2019) are crucial, as they give relevant support in the identification of the elements that can influence the probability to become Superhost. Gunter (2018) has been the first researcher that has investigated the elements that can influence the probability to become a Superhost. He has defined a model, which we will hereinafter call Gunter's model, to evaluate if the four criteria set by Airbnb actually have a direct impact on the probability to obtain the Superhost badge. The four variables included in the model are the number of bookings, cancellation policy, response rate, and overall rating, which we will call *the Airbnb variables*. Gunter (2018) has focused his study on the city of San Francisco. He has discovered that the four variables influence the attribution of the Superhost badge and that it is possible to identify a rank among the variables. Specifically, he has discovered that the most important criterion is "rating", followed by reliable cancellation behavior, responsiveness, and sufficient demand. Gunter's model has been extended at first in Contu et al. (2019). In the model, a new group of variables has been inserted. The group, which has been called the *managerial variables*, identifies the aspects that can be directly chosen by the hosts as the minimum and the maximum number of nights for reservations, how far in the future guests can book, and the cleaning fees. The results have shown how the managerial variables impact on the probability to become a Superhost.
To further contribute to this literature, our study aims to define a framework able to identify the elements that can guide the hosts in the management of their activity and the

effort to improve constantly their performance. The framework is defined by combining two different models, the logistic regression model, and the bivariate logistic regression model. We use firstly the logistic regression model. Specifically, we replicate the study of Gunter (2018) and Contu et al. (2019) in order to identify the elements that can directly influence the badge attribution, taking into account the Airbnb criteria and the aspects related to the accommodation management (Gunter, 2018; Contu et al., 2019). Later, two new models are proposed. The first aims to identify the impact that the characteristics of the accommodation could have on the probability to become a Superhost. The second aims to identify the elements that can impact on the hosts' performances. In the first case, we use the logistic regression model, in the second the bivariate logistic regression model. The use of different models and groups of variables allows defining an extended framework. Each model gives a different and crucial point of view in understanding the host's activity.
The analysis is focused on the most touristic Italian cities, i.e. the ones where the highest number of official tourists is recorded. They are Rome, Milan, Venice, and Florence. Analyzing different cities allows comprehending if differences in terms of variables' impact can be recorded. In fact, the cities present specific characteristics: the cities of Rome and Milan can be considered tourist and business cities, the others only as tourist cities. For this reason, it has been supposed that the differences in terms of variables' impact can be recorded, and that they should be considered for a correct management of the Airbnb accommodation.
From a theoretical point of view, the innovative aspect of this study is the definition of a statistical framework combining different models in order to study the host's activity, and the definition of a guide to support hosts in the effort to improve their performances and become Superhosts.
Four sections, besides the introduction, complete this study. The second and the Third Section is related to the research design: here methodology and data are described. The results and discussion are explained in the fourth part. Finally, the last Section focuses on concluding remarks, limitations, and future developments.

2 Methodology

As evidenced before, the framework is defined by combining different models: the logistic regression model and the bivariate logistic regression model.
It has been decided to use *Logistic regression model* to comprehend which variables can impact on the probability to become a Superhost, and the *Bivariate logistic regression model* to investigate the hosts' performances.
The logistic regression model (or Logit) has been chosen in order to replicate the same framework and model of Gunter (2018). The Logit model allows us to estimate the probability that a specific event occurs. In this specific case, it allows us to estimate the impact of each variable on the probability to become a Superhost. Furthermore, the model offers the possibility to estimate the marginal effects, providing more interpretable results, especially for the coefficients of discrete covariates (Powers and Xie, 2008).

Given the matrix $\mathbf{X}' = (X_0, X_1, X_2, \ldots, X_p)$, where $X_0 = (1, \ldots, 1)'$ has length n and the other elements are the covariates expressed by a collection of p independent variables of length n, the outcome (binomial) variable $Y = (y_1, \ldots, y_n)'$ can assume value 1 (the host is *Superhost*) or 0 (the host is not *Superhost*). The conditional probability of success is $F(\mathbf{X}'\boldsymbol{\beta}) = \Pr(Y = 1|\mathbf{X}, \boldsymbol{\beta}) = \pi(\mathbf{X})$. The vector $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)$ is an unknown vector of $p+1$ regression coefficients. Several functions $F(\cdot)$ can be considered for estimating the relationship between Y and $\mathbf{X}$. The logistic regression model is given by:

$$F(\mathbf{X}'\boldsymbol{\beta}) = \frac{e^{\sum_{i=0}^{p} \beta_i X_i}}{1 + e^{\sum_{i=0}^{p} \beta_i X_i}} \tag{1}$$

The probability of success $P(Y = 1|X)$ is related to $\boldsymbol{\beta}$ through $F^{-1}(\mathbf{X}'\boldsymbol{\beta})$, consequently

$$\log\left(\frac{\pi(\mathbf{X})}{1 - \pi(\mathbf{X})}\right) = \sum_{i=0}^{p} \beta_i X_i \tag{2}$$

The logistic regression model is estimated by maximum likelihood estimation. The likelihood function is

$$\mathcal{L}(\boldsymbol{\beta}|Y, \mathbf{X}) = \prod_{i=1}^{n} \left(F(\mathbf{X}'\boldsymbol{\beta})^{y_i} [1 - F(\mathbf{X}'\boldsymbol{\beta})]^{(1-y_i)}\right) \tag{3}$$

A marginal effect is defined as the rate at which y changes at a given point in the covariate space, with respect to one covariate dimension and holding all covariate values constant (Leeper, 2017, p. 7). It is obtained computing partial derivatives. Considering the other k independent variables, for continuous variables a unit change of x_{ik} determines a β_k change in y_i. Consequently, β_k can be considered as the marginal effect of x_{ik}

$$\frac{\partial y_i}{\partial x_{ik}} = \beta_k \tag{4}$$

Considering a discrete independent variable, the marginal effect is given by

$$E(y_i|x_{ik} = 1) - E(y_i|x_{ik} = 0) = \beta_k \tag{5}$$

Furthermore, the marginal effect can be calculated taking into account the probability of an event to occur. Specifically,

$$\frac{\partial \Pr(y_i = 1|\mathbf{x}_i)}{\partial x_{ik}} = \frac{\partial F(\mathbf{x}_i'\boldsymbol{\beta})}{\partial x_{ik}} = f(\mathbf{x}_i'\boldsymbol{\beta})\beta_k \tag{6}$$

where $f(\cdot)$ denotes the density function. This quantity is the rate of change in the success probability in the neighborhood of a particular value of x. In this study, the average marginal effect (AME) has been calculated. The AME estimates the marginal effects for each observed value of X and average across the effect estimates. It provides a natural summary measure that respects the distribution of the original data (Leeper, 2017).

IIn order to evaluate the variables that can impact on the host performances, the bivariate logistic regression model has been used. The model is a part of a Vector Generalized Additive Model (VGAM) family functions described by Yee (2008); Yee et al. (2010); Yee (2015). It can be defined as a logistic regression model with two different response variables, Y_1 and Y_2. Specifically, it is an extension of the logistic regression model and it estimates the probability that the two response variables, and the positive combination of both variables, assume a value equal to 1. Moreover, it models the marginal distributions of two Y_j and the odds ratio. It is defined through:

$$logit\, p_j(x) = \eta_j(x), j = 1, 2 \tag{7}$$

$$log\, \psi(x) = \eta_3(x) \tag{8}$$

where ψ identifies the odds ratio and describes the association between the two responses. The odds ratio is calculated as:

$$\psi(x) = \frac{p_{00}(x)p_{11}(x)}{p_{01}(x)p_{10}(x)} = \frac{P(Y_1 = 0, Y_2 = 0|X)P(Y_1 = 1, Y_2 = 1|X)}{P(Y_1 = 0, Y_2 = 1|X)P(Y_1 = 1, Y_2 = 0|X)} \tag{9}$$

in such a way that $\psi = 1$, if Y_1 and Y_2 are independent.

The model has been chosen because it is capable of estimating the probability of two binomial response variables and their combination. Thus it gives the possibility to study the association between the two response variables. It allows studying how the different variables can influence the presence of specific characteristics. Moreover, it is data-driven.

The response variables inserted in the model are *Overall rating* (OvR) and *Occupancy rate* (OcR). The first variable has been chosen because Gunter (2018) has discovered that this variable has the highest marginal effect on the probability to become a Superhost. It measures the level of satisfaction expressed by guests with respect to the Airbnb experience lived inside the accommodation. Investigating the aspects that can support the host to improve the OvR, and consequently, the probability to reach the badge of Superhost can support the host to improve their performance. The variable OcR has been investigated because it is an indicator used to evaluate the performance of hotels. In this case, it is calculated as the ratio between the *Count of Reservation Days* and the sum of *Count of Reservation Days + Count of Available Days*. It expresses the percentage of the reserved days' with respect to the available days. Since the aim of the host is to rent the apartment, the occupancy rate is a measure of his capacity to work better. Comprehending which variables can influence OcR is crucial to improving the activity of the host.

Both variables are considered a measure of the host's performance and crucial to evaluating an accommodation facility. It is believed that the two aspects define the most important results for the host activities: one in terms of the number of reservations and one in terms of the quality of the host services.

The two quantitative variables have been transformed into binary variables to become the response variable in the bivariate logistic regression model.

To conclude, it is possible to state that both proposed models are able to estimate the probability of specific events. The first is able to estimate the influence that each covariate has on the probability to obtain the status of Superhost. The second is able to estimate the probability to reach high results in terms of the two outcome variables and of their positive combination. The combination of the two models, joined with the use of different groups of independent variables, defines a useful framework to guide the hosts in the constant effort to improve their performances and become Superhosts.

3 Data

The study has been realized using a dataset provided by *Airdna*, a company that manages Airbnb data. The Airdna dataset includes different information, for instance, aspects related to the hosts and their activities, the revenues, the number of reservation days, the characteristics of accommodations, the reviews on the apartments, and the coordinates of the accommodations. A subset of data has been extracted. Three main groups of variables compose it, as evidenced in Table 1.

The first group includes the variables used by Airbnb to attribute the badge of Superhost. The second group is composed of the *managerial variables.* Before appearing on the platform, the host decides, and declares, specific aspects related to her activity, such as the minimum and the maximum number of nights for reservations, maximum guests, cleaning fees, and other aspects related to the activity management (`https://www.airbnb.com/b/setup`) (accessed on 15/02/2019). All these aspects are related to the services offered, the additional fees to be applied, and how to build and conduct the relationship with the guest. They are directly decided and managed by the hosts. Three more variables have been included in this group. These variables are *Occupancy rate, Number of reviews, and Superhost.* They cannot be totally controlled by the host, but nonetheless, the host can influence them. Finally, the third group of variables includes the specific characteristics of the accommodation like *Listing Type* and the number of bathrooms and bedrooms.

Some variables have been transformed to be inserted into the models. For instance, following the model defined by Gunter (2018), the variable cancellation policy is introduced in the model as a binary variable. It has been attributed the value equal to 0 if the host chooses the flexible policy and a value equal to 1 if the host chooses the strict or moderate policy. This transformation aims to differentiate the hosts that have a specific policy of cancellation from the others. Moreover, the variables Property Type and Listing Type have been transformed into binary variables. In the first case, the value 1 is attributed to the accommodation type "apartment", whereas the value 0 is attributed to all the other types. In the second case, the variables equal 1 when the entire accommodation is offered and 0 when only a room is rented.

The time period is the year 2016. We focus on the four cities in Italy: Rome, Milan, Venice, and Florence. The choice of analyzing these Italian cities is due to their importance as touristic destinations: most tourists visiting Italy stay in one of these cities. Following the official tourism data of 2016, the four most visited cities in Italy were Rome, with more than 29 million tourists, followed by Milan, Venice and Florence (see

Table 2). At the same time, analyzing the Airbnb data, it has been discovered that these cities present the highest number of Airbnb accommodations. This means that in Italy, Airbnb accommodations are located in the well-known and most visited tourist cities. Moreover, since different researchers have shown that short-term rentals happen mostly in the most touristic areas of urban centers, (see for instance Quattrone et al. (2016)), it has been decided to study the location of the Airbnb accommodations. Analyzing the Figure 2, it is possible to evidence again that Airbnb accommodations are located in the most visited cities and near the tourist attractions. These findings are totally in line with the statement of Gutiérrez et al. (2017) that has argued that the Airbnb accommodations are located close to the main attractions, and with the statement of Dudás et al. (2017) that confirmed the fact that Airbnb accommodations are concentrated in specific areas of the tourist cities.

Table 1: The variables

Group of variables	Name	description
Airbnb	Cancellation policy	Airbnb gives the possibility to choose among three main different cancellation policies: *flexible*, when the guest can cancel until one day before his arrivals; *moderate*, when the cancellation can be made until 5 days before his arrivals; and *strict*, when the reservation can be canceled until seven days before the arrival. Additionally, the host can choose among different stricter cancellation policies, as the *Super Strict 30 Days*, *the Super Strict 60 Days* and *Long Term*.
	Number of Bookings	the number of reservations recorded by the host
	Response Rate	the percentage of times a host responds to potential guests within 24 hours
	Overall Rating	the score value that the guests express to judge the whole Airbnb experience.
Managerial	Max Guests	the maximum number of guests that can be hosted in the accommodation
	Response Time	the amount of time that the guests must wait on average to obtain an answer to their questions
	Extra People fee	the price to be paid to add one or more persons to the reservation
	Minimum Stay	the minimum number of days that must be booked
	Security Deposit	the payment (or not) of a deposit
	Business Ready	i.e. the services that can support business travelers, such as Wi-Fi, a laptop-friendly workspace, an iron, hangers, shampoo, a hairdryer and all the other essentials such as toilet paper and clean towels
	Instantbook Enabled	a service that allows guests to book accommodation without an explicit host approval, facilitating the reservation process
	Occupancy rate	he ratio between the *Count of Reservation Days* and the sum of *Count of Reservation Days + Count of Available Days*.
	Number of reviews	the number of reviews published by guests
	Superhost	
Accommodation	Property Type	different types of accommodation can be rent on Airbnb. Specifically, it is possible to rent more than 25 different types of accommodation, as for instance apartments, castles, boats, and caravans
	Listing Type	three different listing types are offered on the Airbnb platform: entire accommodation, private room, and shared room
	Bedrooms	the number of bedrooms in the accommodation
	Bathrooms	the number of bathrooms in the accommodation

Rank of tourism official flow		
position	**City**	**Bed nights**
1	Rome	25,191,580
2	Milan	10,976,244
3	Venice	10,511,788
4	Florence	9,334,085

Rank of Airbnb accommodation		
position	**City**	**Bed nights**
1	Rome	19,042
2	Milan	11,057
3	Florence	7,753
4	Venice	5,625

Table 2: Rank of tourism flow and Airbnb accommodation

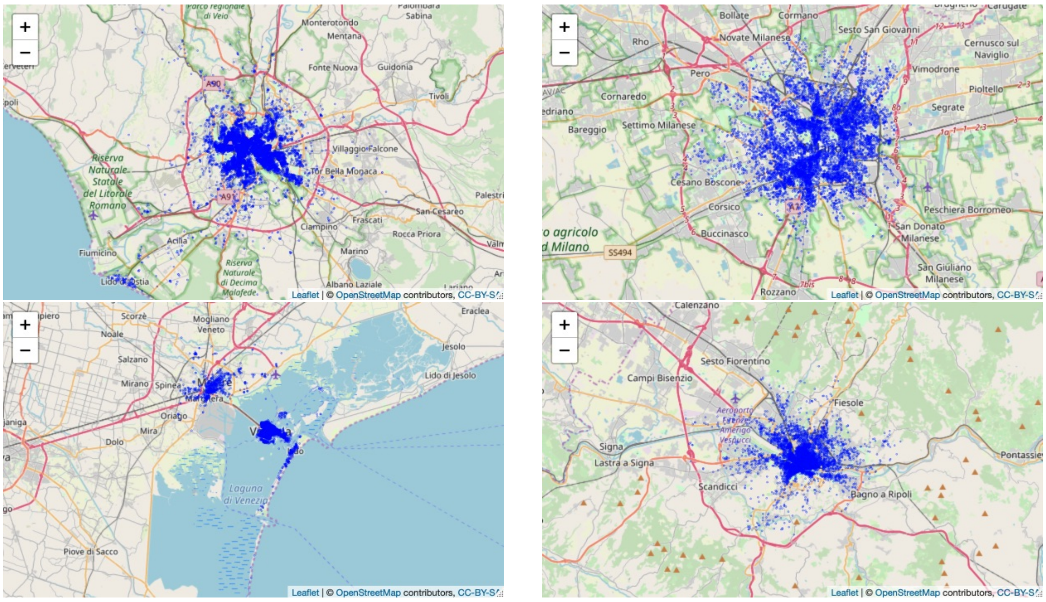

Figure 1: Airbnb location in Rome, Milan, Venice and Florence

4 Results and discussion

4.1 The Superhosts

It has been firstly applied the Logit model and investigated if the Airbnb variables can have an impact on the probability to become a Superhost for the hosts operating in the main Italian tourist cities. Observing the results reported in Table 3, it emerges that all variables that Airbnb declares to use to award hosts with the badge of Superhost have a significant and positive influence on the probability to become Superhost in the Italian cities. The p-value identified with the stars in Table 3 (the stars correspond to a significance level of, respectively, less than 0.1% ($'***'$), from 0.1% to 1% ($'**'$) from 1% to 5% ($'*'$), and from 5% to 10% $'.'$) is, in most of the cases, close to the zero, evidencing statistical significance of all variables. This means that Gunter's model is valid for Italian tourist cities also, and not only for the city of San Francisco.
Findings emerge from the estimation of average marginal effect (AME) for each variable. Specifically, it measures the change in the probability to become a Superhost generated by an increase of covariate by one unit. This information is relevant to better comprehend the effect that each variable can have on the achievement of the Superhost badge. It can support the host to focus on the variables that can really influence their possibility to change their status. Analyzing the values of the estimated AMEs, it is possible to observe that *Overall rating* is the variable that influences the most the probability to become a Superhost in all Italian cities. This result is in line with the one obtained from Gunter's model where the estimated marginal effect of *Overall rating* is estimated to be equal to 0.28 in San Francisco. However, estimated coefficients are higher for the Italian cities. They move from 0.39 for the city of Milan to 0.64 for the city of Florence.

The impact is significant from a statistical point of view: the probability to become Superhost increases significantly for an increase in the value of *Overall rating.* Hosts operating in the Italian cities should work to the overall rating of their accommodations in order to improve the probability to become a Superhost.
The other findings concerning the ranking of the most important variables affecting the possibility to become Superhosts are equally interesting. The second most important variable is *Cancellation policy* for Rome, Venice, and Florence, followed by *Response rate* and *Number of bookings.* For these cities, the rank of the variables is exactly the same rank estimated for the city of San Francisco by Gunter (2018). For this reason, it is possible to re-state that Gunter's model is valid also in other cities. However, some differences emerge for the city of Milan. Specifically, in Milan, the most important variable is still the Overall rating, but the second most important one is *Response rate*, followed by Cancellation policy and Number of bookings. This means that Gunter's model is not completely valid in Milan, as our results suggest that the contact among hosts and guests (expressed by *Response rate*) is more important than the presence of a strict cancellation policy. This difference is motivated by arguing that people traveling to Milan are usually doing it for job reasons more often than in other Italian cities, so they are probably looking for a direct and fast relationship with the host.
To sum up, it is possible to state that Gunter's model is extendible to the Italian cities as well: the four Airbnb variables impact the probability to become a Superhost. Moreover, it is possible to state that in order to obtain and maintain the Superhost badge hosts should improve their results in terms of OvR. The general review expressed by guests is the element that determines the attribution of the badge. We believe that comprehending which aspects can improve the OvR level is crucial for a host in order to achieve the status of Superhost.
Next, we investigate if even other variables have a specific impact on the probability to become a Superhost. For this purpose, two different kinds of variables have been taken into account: the managerial aspects and the characteristics of the accommodations. Results in Table 4 show that all the above-mentioned managerial variables have a significant impact on the probability to become a Superhost. The p-values show the significant impact that these covariates have on the badge attribution. Only the variable *Max Guests* is not significant for the cities of Rome and Florence. Generally, similar results emerge for the cities of Rome, Venice, and Florence and different behavior for the city of Milan.
Particularly interesting is also the positive impact of the variables *Occupancy rate* and *Business ready.* The two variables have a low p-value and the highest values of the AME. Occupancy rate is a measure of the capacity of the host to rent the house for as many days as possible: an increase of one point in occupancy rate causes an increase of the probability to become a Superhost. This means that positive results in terms of occupancy rate can improve the probability to become a Superhost. This positive impact is in line with the statement of Xie and Mao (2017) that has discovered that being a Superhost, having a longer operating experience and a higher response rate have an impact on reservations. The Business Ready variable, on the other side, describes some of the services supplied by hosts. The value of the marginal effect is higher than 10%

in the cities of Rome, Venice, and Florence. The positive impact suggests that guests like to receive the same services that hotels normally offer. The results are confirmed by the statement of Guttentag et al. (2018) and Lalicic and Weismayer (2018), who have argued that *Airbnb users' are primarily attracted to the service by its practical advantages*(Guttentag et al., 2018, p. 354); and that *hosts should be aware of the fact that guests request a specific standard, besides the interactive part of the hosting experience* (Lalicic and Weismayer, 2018, p. 89). To sum up, guests seem to prefer living their holidays in comfortable, family-like accommodations that offer more or less the same facilities that could be found in hotels. Therefore, the results suggest a sort of substitutability (or interchangeability) of hotels and Airbnb accommodations.

It is also interesting to note the negative impact of the variable *Max Guests* for the cities of Milan and Venice. It seems to be in contrast with the positive, and significant, impact of the variables *Extra People Fee.* This contraposition suggests that guests prefer small accommodations with the possibility to invite other people after paying an extra cost.

To summarize, it is possible to state that the managerial variables have an impact on the probability to become a Superhost.

Finally, the impact of the characteristics of the house has been investigated in order to evaluate the possible influence of these aspects. The analysis of Table 5 shows a significant and positive impact of renting an entire accommodation in Rome, Florence, and Venice. The marginal effect doesn't have a high value, but it's still positive. This confirms the preference of guests for private places where they can feel at home. Moreover, the probability to be a Superhost is influenced by the variable Type of places in Milan, where apartments are preferred w.r.t. other kinds of accommodations. Finally, the high number of bedrooms has a significant but negative impact on the city of Florence only. To sum up, the characteristics of the house have an impact on the probability to become a Superhost. These aspects are not considered by Airbnb in the attribution of the Superhost badge. Nonetheless, Airbnb has created an online section called "*Airbnb plus stays*", where the most beautiful and charming accommodations are available for booking. This choice from Airbnb reinforces the necessity to consider these aspects related to the accommodation in the attribution of the badge of Superhost.

The results suggest firstly that the Superhost status is affected by the services supplied and the capacity to offer high-quality service. For this reason, the host to become a Superhost should offer different services to improve the quality of the Airbnb experience, maintain a direct relationship with the guests in the effort to support them during their stay, answering as soon as possible to their request.

Secondly, the results highlight the existence of some differences in terms of variables' impact among the cities. For instance, the hosts that operate in Rome should rent an entire apartment and reduce the number of guests in the accommodation in order to reach the badge of Superhost. On the contrary, the hosts that operate in the cities of Milan, Florence, and Venice can improve the probability of becoming a Superhost, if they rent an apartment instead of the other types of accommodation. Moreover, the size of the accommodation become relevant only for the host operating in the city of Florence, the unique city where the number of bathrooms and bedrooms influences the probability to become a Superhost.

Table 3: The impact of Airbnb variables on the probability to be a Superhost

	Milan			Rome			Florence			Venice		
	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME
Intercept	-32.05	***		-31.19	***		-40.14	***		-42.55	***	
Cancellation.Policy (1)	0.24	**	0.00	0.16	**	0.02	0.29	**	0.03	0.41	**	0.04
Number of Bookings	0.02	***	0.00	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
Response Rate	0.08	***	0.01	0.05	***	0.01	0.11	***	0.01	0.10	***	0.01
Overall Rating	4.56	***	0.39	5.01	***	0.54	5.60	***	0.64	6.45	***	0.62

Table 4: The impact of managerial variables on the probability to be a Superhost

	Milan			Rome			Florence			Venice		
	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME
Intercept	-2.5752	***		-2.5808	***		-2.3819	***		-2.7049	***	
Occupancy Rate	1.2972	***	0.12	1.5574	***	0.19	1.3677	***	0.22	1.8937	***	0.22
Max Guests	-0.0605	**	-0.01	-0.0078		-0.00	0.02		0,00	-0.0389	.	-0.00
Response Time min	-0.0015	***	-0.00	-0.0011	***	-0.00	-0.0011	***	-0.00	-0.0010	***	-0.00
Security Deposit (Yes)	0.2778	***	0.03	0.1904	***	0.02	0.1809	**	0.02	0.5205	***	0.06
Extra.People.Fee (Yes)	0.2598	***	0.02	0.3224	***	0.04	0.2203	***	0.02	0.2662	**	0.03
Business.Ready (Yes)	0.5985	***	0.06	0.7704	***	0.11	0.72	***	0.12	0.82	***	0.12
Instantbook Enabled (Yes)	-0.8048	***	-0.06	-0.2832	***	-0.03	-0.3661	***	-0.05	-0.5508	***	-0.06
Number of Reviews	0.0076	***	0.00	0.0024	***	0.00	0.0023	***	0.00	0.0032	***	0.00

Table 5: The impact of characteristics of the accommodation on the probability to be a Superhost

	Milan			Rome			Florence			Venice		
	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME	Coef.	*pvalue*	AME
Intercept	-2.2046	***		-1.9293	***		-1.9812	***		-2.0386	***	
Listing Type (1)	-0.0060		-0.00	0.25	***	0.03	0.4873	***	0.06	0.3787	***	0.04
Property Type (1)	0.1845	*	0.02	-0.0269		-0.00	0.0547		0.01	-0.0092		-0.00
Bedrooms	-0.0666		-0.01	0.0124		0.00	-0.1141	*	-0.02	-0.0105		-0.00
Bathrooms	0.0344		0.00	0.0172		0.00	0.1151	.	0.02	0.0122		0.00

4.2 Occupancy rate and overall rating

The second part of the framework aims to identify the variables that directly influence the probability to obtain good results in terms of OvR and OcR. The two variables have a relevant impact on the probability to become a Superhost and also the highest marginal effects on it. Comprehending which aspects can influence OvR and OcR can support hosts in the effort to improve their results and become Superhost.
Firstly, the relationship between OcR and the OvR has been investigated. Figure 2 shows the presence of an elevate number of host that, at the same time, reach high results in terms of OvR and OcR.

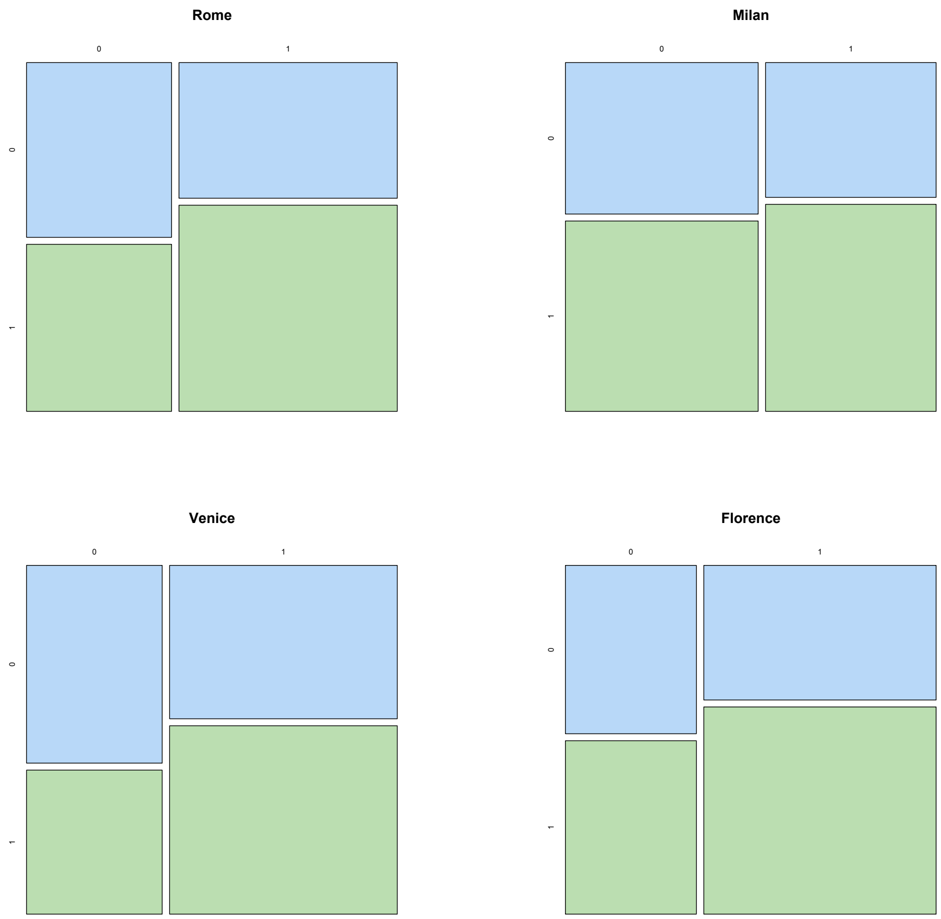

Figure 2: The mosaic plot of OcR and OvR for Rome, Milan, Venice and Florence

This means that hosts have accommodations occupied for most of the open days and offer a quality service that allows reaching high results in terms of positive reviews. This is confirmed also by the odds ratio. Generally, the odds ratio (OR) quantifies the strength of the association between two events. The OR is equal to one if the two events are independent. It is greater than one when two events are associated. In this study, the odds ratio describes the association between the two response variables. It assumes a value equal to one whether *Occupancy rate and Overall rating* are independent, otherwise, it assumes a value larger than one when the two variables are positively associated, and it is lower than one if the two variables are negatively associated. The highest value of odds ratios has been recorded for the city of Venice (odds ratio=1.69), followed by the city of Rome (odds ratio=1.59), Florence (odds ratio=1.47), and Milan (odds ratio=1.22). The estimated OR for Venice shows how the odds to reach a high value of OcR for a host with a high value of OvR is about 1.69 times the odds to reach a high value of OcR for the host with the lowest value of OvR. This highest-value shows firstly the existence of a positive connection between the two variables; secondly, they suggest the capacity of the hosts in Venice to work as professionals. It seems that these hosts are able to occupy the accommodation for the available days and to offer high-quality services so much to generate positive judgment of the entire Airbnb experience. A positive association between OcR and OvR has been also recorded in the cities of Rome and Florence, where high results of odds ratio has been estimated. On the contrary, a lowest value has been recorded for the city of Milan. It suggests a minor connection between the two variables in Milan and the reduced probability for the host with high value of OvR to obtain also high value of OcR. This result seems to be in line with the lowest impact of the OvR on the probability to become a Superhost among the other cities suggesting a minor impact of this variable in the evaluation of the Airbnb phenomenon in Milan.

Furthermore, the bivariate logistic regression model has been applied in order to evaluate the impact of the managerial variables and the characteristics of the accommodation on OvR and OcR. Firstly, the managerial variables are taken into account. The results in Table 6 and 7 show that managerial variables are, in most cases, statistically significant (the $p-values$ are close to zero). This means that managerial variables have an impact on the probability to reach high results for both response variables and are able to explain the success of Airbnb accommodations.

Table 6: Milan and Rome: the Airbnb impact on Occupancy rate and overall rating, managerial variables

	Milan						Rome					
	Occ.		Over.		(Occ=1,Over=1)		Occ.		Over.		(Occ=1,Over=1)	
Intercept	−1.80	***	−0.22		0.73	·	−1.62	***	−0.54	***	0.48	
Max Guests	−0.07	***	−0.09	***	0.05	·	−0.04	***	−0.05	***	0.02	
Response Time min	−0.00	***	0.00		−0.00		−0.00	***	−0.00		−0.00	
Security Deposit (Yes)	−0.07	·	0.23	***	−0.02		−0.18	***	0.19	***	−0.03	
Extra People Fee (Yes)	0.09	·	0.03		0.12		−0.04		0.13	***	−0.06	
Business Ready (Yes)	−0.06		0.41	***	0.15		0.16	*	0.89	***	0.29	·
Instantbook Enabled (Yes)	0.81	***	−0.47	***	−0.28	*	0.58	***	−0.20	***	0.28	***
Number of Reviews	0.04	***	−0.00		−0.00		0.05	***	0.00	**	0.01	*
Response Rate	0.014	***	0.01	***	−0.01	*	0.01	***	0.00	***	−0.00	
Superhost (=1)	0.62	***	2.87	***	0.39		0.68	***	3.01	***	0.88	***
Cancellation Policy (=1)	0.08	·	0.02		−0.09		0.09	*	0.04		−0.06	

Table 7: Florence and Venice: the Airbnb impact on Occupancy rate and overall rating, managerial variables

	Florence						Venice					
	Occ.		Over.		(Occ=1,Over=1)		Occ.		Over.		(Occ=1,Over=1)	
Intercept	-1.02	**	-0.61	*	-0.07		-0.09		-1.38	***	-0.65	
Max Guests	-0.10	***	-0.03	*	0.07	*	-0.10	***	-0.07	***	0.10	**
Response Time min	-0.00	***	0.00		0.00	·	-0.00	***	-0.00		-0.00	
Security Deposit (Yes)	- 0.17	**	0.15	**	0.09		-0.14	*	0.25	***	0.13	
Extra People Fee (Yes)	0.19	**	0.16	**	-0.08		0.21	**	0.17	**	-0.09	
Business Ready (Yes)	0.07		0.53	***	-0.24		0.18		0.27	*	0.51	
Instantbook Enabled (Yes)	0.43	***	-0.40	***	-0.07		0.20	**	-0.30	***	0.25	
Number of Reviews	0.05	***	0.00		0.00		0.04	***	0.00	**	0.01	*
Response Rate	0.01	·	0.01	**	0.00		-0.00		0.01	***	0.00	
Superhost (=1)	0.61	***	3.28	***	0.65	·	0.60	***	3.38	***	0.46	
Cancellation Policy (=1)	-0.04		-0.09		-0.11		-0.29	***	-0.07		0.03	

It is also interesting to highlight the positive and negative impact of specific covariates on OcR and OvR. For instance, the variable *Max Guests* has a negative influence on the level of OcR. This aspect is confirmed by the significant and negative impact that the variables *Bedrooms* and *Bathrooms* have on the probability to have a high level of *Occupancy rate* (Table 8 and 9). The increase in the size of the house does not have a positive impact on the rate of occupied days of Airbnb accommodations in the Italian cities. This result suggests that generally speaking guests prefer smaller accommodations. However, in Milan, Florence, and Venice a positive and significant effect of the variable *Extra People fee* is also observed. This means that guests prefer to have the possibility to share the accommodation with other people. This result suggests that it is not important to rent a big house, but rather to give the possibility to invite people to the accommodation: in this way, guests will feel as if they were in their own homes. Consequently, if a host wants to succeed on the Airbnb platform she should offer a small house and give the possibility to share the space with other people freely.

Moreover, a positive impact has been also observed for the variable *Instantbook Enabled* on the OcR. This variable identifies the possibility to book without a previous confirmation by the host. Gunter and Önder (2018) and Benítez-Aurioles (2018) have stated that this positive impact is related to the preference of the guests for hotel-like booking processes. It is believed that this impact is also related to a sense of bigger freedom that guests look for while planning their holidays. In fact, Airbnb accommodations have fewer constraints and schedules than hotels, and this feeling of freedom and ease gets stronger when one does not need to wait for the host's confirmation. However, the variable *Instantbook Enabled* has a negative impact on the variable Overall rating. This suggests that Overall rating is affected by the relationship created between hosts and guests. This is particularly interesting because a negative impact was registered also on the probability to become a Superhost, showing another time the strong relationship between being a Superhost and high review scores.

The variable *Response Time* and *Response rate* have, respectively, a negative and a positive impact on *Occupancy rate*. *Response rate* has a positive impact also on *Overall rating* and on the positive combination of the two variables in Rome, Florence, and Venice. As evidenced by Xie and Mao (2017), Response Rate is perceived as a signal of reliability and honesty and it is crucial to building a trustful relationship between host and guests from the very beginning. Results suggest that if communication is required, it has to be quick and immediate. This aspect seems to suggest that guests look for a careful host, able to support them during her Airbnb experience.

Furthermore, the variable *Cancellation Policy* has a positive impact on *Occupancy rate*. This suggests the preference of the guests for the presence of strict cancellation policies. In addition, the positive and significant impact of the variable *Business Ready* on the response variable OvR shows another time that the reviews of guests are related to the quality and quantity of the services supplied.

The variable *Superhost* has a positive impact on both response variables OcR and OvR. Also, Xie and Mao (2017) have shown the important impact that becoming a Superhost has on the occupancy rate. They have stated that *having more reservations in the subsequent month depends in part on becoming a Superhost* (Xie and Mao, 2017, p. 16). In

fact, the Superhost badge is a guarantee for guests, that normally associate the badge with honesty and reliability, which mitigates the perceived risk of staying in the private house of a stranger (Xie and Mao, 2017). Neumann and Gutt (2017) have suggested that being Superhost gives hosts the chance to improve their results on the Airbnb platform faster and more easily. It is also interesting that Airbnb declares that, in order to become a Superhost, high values in the OvR are needed. The results of this research show that the value of the OvR is actually influenced by the Superhost badge. This is easily explained by the fact that hosts become Superhosts by offering high-quality services and by supporting guests during their stays: these are the reasons why they get good reviews in the first place, and therefore improve their results in terms of OvR. It is possible to state that a positive and bidirectional relationship exists between the variable *OvR* and the badge of *Superhost*.
The analyzed results show a significant impact of the managerial variables on the OcR and OvR. However, some differences have been recorded among the cities. For instance, in Rome, the host should offer all services included in Business ready to improve the positive results in both response variables. Moreover, they should choose a strict cancellation to improve the OcR. Similarly, a strict cancellation policy can be a relevant choice for the hosts operating in Milan. Additionally, the Milanese hosts should not introduce an extra fee for the presence of more people, because the extra fees are negatively perceived, and they reduce the probability to have high results in terms of OcR and OvR. On the contrary, this extra fee presents a positive impact on the hosts operating in Rome, Florence, and Venice. For this reason, they should consider introducing this fee in the phase of management planning.
To sum up, the hosts should evaluate the impact of different variables considering the city where the accommodation is located to improve their performances.
Later, the impact that the characteristics of the accommodation have on host performances has been investigated. Specifically, the analysis of the impact of these variables can support the hosts in understanding what kinds of accommodation travelers prefer, their size, and the optimal number of rooms: evidence on these topics can help define the best offer.
The first aspect taken into account is *Listing Type*. Lutz and Newlands (2018) have shown that choosing to rent a shared or entire accommodation is affected by different aspects for instance: business constraints, travel circumstances, environmental factors, and guests' discomfort. The findings shown in Table 8 and 9 demonstrate that renting an entire accommodation improves the probability to both increase the occupation rate, in the cities of Milan and Rome, and obtain better reviews, in the cities of Rome, Florence, and Venice. In fact, results prove that guests prefer to rent entire accommodations rather than shared ones. This result is in line with the statement of Varma et al. (2016) that have underlined that guests choose Airbnb accommodations for the *relatively more personal atmosphere than hotels* (Varma et al., 2016, p. 233). Additionally, it is in line with the statement of Guttentag et al. (2018) that have identified the household amenities and the homely feel as important motivations of the guests.
This means that reviews depend not only on the available services but also on the characteristics of the house. This result, which never emerged in previous studies, is extremely

relevant in the study of the Airbnb phenomenon.
Moreover, *Property Type* has a positive influence on the hosts' performances. Renting an apartment, instead of a room, improves the *OcR* in Milan, Rome, and Venice. Additionally, it affects the *Overall rating* in Milan. It is once more evidenced the importance that the house characteristics have on the Airbnb activity and on the achievement of good performances.
Additionally, as evidenced before, the increase in the number of *Bedrooms* and *Bathrooms* negatively affects the *OcR*. Interestingly, the two variables negatively influence also the *Overall rating* in Venice. This means that, in this city, the characteristics of the house are crucial to getting good results.
The results have shown the presence of differences in terms of impact of the characteristics of the accommodation among the cities. For instance, the hosts operating in Venice should prefer to rent accommodation with an elevated number of bedrooms and bathrooms if they want to improve the probability to reach high results of OcR. The hosts operating in Florence and Rome should rent an entire accommodation to improve their performance. Finally, hosts operating in Milan should rent an apartment with respect to other kinds of structures to improve the OvR. These differences should be considered by host in the management of the activity in the continuous effort to improve their activity.
Finally, the results have shown how the bivariate logistic regression model is able to evaluate the impact of the managerial variables and the house characteristics on the *OcR* and *OvR*. This is an innovative result that can support new hosts in the identification of the best accommodation in order to successfully operate in the most visited Italian cities.

Table 8: The Airbnb impact on Occupancy rate and overall rating, characteristics of the accommodations - Milan and Rome

	Milan						Rome					
	Occ.		Over.		(Occ=1,Over=1)		Occ.		Over.		(Occ=1,Over=1)	
Intercept	-0.09		0.18	*	0.07		0.30	***	0.12	**	0.37	***
Listing Type (1)	0.31	***	-0.07		-0.01		0.36	***	0.07	*.	0.22	**
Property Type (1)	0.16	**	0.15	*	0.10		0.34	***	0.03		-0.07	
Bathrooms	-0.15	**	0.08	.	0.05		-0.20	***	0.01		-0.03	
Bedrooms	-0.20	***	-0.02		0.00		-0.10	***	0.02		0.02	

Table 9: The Airbnb impact on Occupancy rate and overall rating, characteristics of the accommodations - Florence and Venice

Florence							Venice					
	Occ.		Over.		(Occ=1,Over=1)		Occ.		Over.		(Occ=1,Over=1)	
Intercept	0.68	***	0.07		0.18		1.02	***	-0.15	.	0.26	
Listing Type (1)	0.03		0.24	***	0.33	**	-0.01		0.32	***	0.36	*
Property Type (1)	0.44	***	0.02		-0.16		0.13		-0.12		-0.04	
Bedrooms	-0.18	***	-0.07	.	0.08		-0.11	**	-0.00		-0.03	
Bathrooms	-0.19	***	0.09	.	-0.01		-0.32	***	0.02		0.08	

5 Concluding remarks

Few researchers have investigated the elements that can support the hosts in becoming Superhosts and improving their performances. In order to improve the literature, we propose a statistical framework to discover which aspects can affect the hosts' activity. We defined a framework combining different models. We used the logistic regression model and the bivariate logistic regression model. Firstly, the use of the logistic regression model has allowed demonstrating how Gunter's model can be extended to other geographic areas. The four Airbnb variables can explain the Superhost badge attribution in tourist cities. We have confirmed that the variable used by Airbnb to attribute the badge of Superhost impact really and significantly on the probability to become a Superhost.
Secondly, we have demonstrated that the Airbnb variables are not the unique variables that can impact on the probability to become a Superhost. Additional variables, such as the managerial variables and the characteristics of the accommodation, can improve this probability. The specification of a new model with these variables has allowed us to enrich the analysis and define a complete and more exhaustive framework.
Thirdly, the choice of a complex model, such as the bivariate odds model, has allowed investigating the host's performances taking into account not only single variables but their positive combination.
Additionally, we introduce to the literature defining a statistical framework useful to better comprehend the elements that can support the hosts in the constant effort to improve their performances and to become a Superhost. It identifies the aspects that the hosts should focus on to obtain, and maintain good results, and work professionally. Moreover, the framework explains the host activity taking into account the specific characteristics of tourism destinations. It evidences differences in terms of relevance and significance of variables among the cities.
In terms of managerial implication, our findings offer a guide for the hosts to improve their performances and become Superhost. We identify different aspects that the hosts should consider improving their performances and the probability to become a Superhost.
Firstly, the host should operate in order to improve the overall rating. The guests' opinions, expressed through the overall rating and the reviews, have a crucial impact on the probability to become a Superhost. Our findings evidence how hosts can improve the overall rating by offering many different types of services (as a laptop-friendly workspace and all the other essentials such as toilet paper and clean towels), supporting guests during holiday, and offering nice accommodation. Specifically, answering the guests quickly, offering services similar to the hotels, and creating a familiar accommodation encourage the guests to book an accommodation.
Secondly, since the aim of Airbnb renting is to occupy the highest number of available days, the host should work to improve the OcR. This goal can be reached by combining high-quality services and continuous support to the guest. In fact, the rental of Airbnb accommodation is based on the relationship between host and guest as evidenced by Guttentag et al. (2018). For this reason, the response time becomes a relevant key for

the host to improve the positive perception of the guests and to increase the occupancy rate.
Thirdly, the host should rent entire apartments, when available, because it gives more probability to become a Superhost. Even the number of bedrooms and bathrooms can influence the OcR and the general judgment of the guests. The hosts do not have many possibilities to change their accommodation, but they can in time modify it to offer a place in line with the guests' necessity.
Moreover, the analysis has highlighted the existence of differences in terms of significance and impact signs among the cities. For this reason, the hosts should consider the different impacts of the variables with respect to the destination.
To sum up, the identification of a list of significant elements has allowed defining a guide for hosts operating in tourist cities. The guide suggests which aspects are more important to reach high performances and obtain the badge of Superhost.
The study, of course, has also some limitations. The first is a temporal limitation. The analysis has been realized considering data from 2016 only. It would be interesting to extend the analysis in order to comprehend if the findings remain valid through time, or if it is possible to identify some differences. The second limitation is a geographical limitation. The study focuses on the most visited Italian cities. It is not possible to know whether the results we obtained for the cities could be extended to smaller or less touristic cities; or whether the Airbnb activity is also influenced by the location of the city. For instance, it could be interesting to understand if cities located by the sea behave differently with respect to the cities located far from the sea.
Finally, additional research could be made for more cities in order to comprehend if the findings of this study can be extended to other territories. Cities in other countries should be analyzed in order to understand if important differences emerge: for instance, it would be very interesting to study the host activity in different European capitals. Besides, the variable that influences the rental activities should be analyzed for different tourist destinations, to verify if the findings of this study could be extended to other areas.

References

Aznar, J. P., Sayeras, J. M., Rocafort, A., and Galiana, J. (2016). The irruption of airbnb and its effects on hotel profitability: An analysis of barcelona's hotel sector. *Intangible Capital,*, 13(1):147–159.

Benítez-Aurioles, B. (2018). Why are flexible booking policies priced negatively? *Tourism Management*, 67:312–325.

Choi, K. H., Jung, J. H., Ryu, S. Y., Do Kim, S., and Yoon, S. M. (2015). The relationship between airbnb and the hotel revenue: in the case of korea. *Indian Journal of Science and Technology*, 8(26).

Contu, G., Conversano, C., Frigau, L., and Mola, F. (2019). Identifying factors affecting the status of superhost: evidence from sardinia and sicily. *Quality & Quantity*, pages 1–21.

Dudás, G., Boros, L., Kovalcsik, T., and Kovalcsik, B. (2017). The visualisation of the spatiality of airbnb in budapest using 3-band raster represeantation. *Geographia Technica*, 12(1):23–30.

Edelman, B. G. and Luca, M. (2014). Digital discrimination: The case of airbnb. com. *Harvard Business School*, 14-054.

Fagerstrøm, A., Pawar, S., Sigurdsson, V., Foxall, G. R., and Yani-de Soriano, M. (2017). That personal profile image might jeopardize your rental opportunity! on the relative impact of the seller's facial expressions upon buying behavior on airbnb™. *Computers in Human Behavior*, 72:123–131.

Gunter, U. (2018). What makes an airbnb host a superhost? empirical evidence from san francisco and the bay area. *Tourism Management*, 66:26–37.

Gunter, U. and Önder, I. (2018). Determinants of airbnb demand in vienna and their implications for the traditional accommodation industry. *Tourism Economics*, 24(3):270–293.

Gutiérrez, J., García-Palomares, J. C., Romanillos, G., and 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.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current issues in Tourism*, 18(12):1192–1217.

Guttentag, D., Smith, S., Potwarka, L., and Havitz, M. (2018). Why tourists choose airbnb: A motivation-based segmentation study. *Journal of Travel Research*, 57(3):342–359.

Lalicic, L. and Weismayer, C. (2018). A model of tourists' loyalty: the case of airbnb. *Journal of Hospitality and Tourism Technology*, 9(1):80–93.

Lee, S. and Kim, D.-Y. (2018). The effect of hedonic and utilitarian values on satisfaction and loyalty of airbnb users. *International Journal of Contemporary Hospitality Management*, 30(3):1332–1351.

Leeper, T. J. (2017). Interpreting regression results using average marginal effects with r's margins. *Available at the comprehensive R Archive Network (CRAN).*

Liu, S. Q. and Mattila, A. S. (2017). Airbnb: Online targeted advertising, sense of power, and consumer decisions. *International Journal of Hospitality Management*, 60:33–41.

Lutz, C. and Newlands, G. (2018). Consumer segmentation within the sharing economy: The case of airbnb. *Journal of Business Research*, 88:187–196.

Malazizi, N., Alipour, H., and Olya, H. (2018). Risk perceptions of airbnb hosts: Evidence from a mediterranean island. *Sustainability (2071-1050)*, 10(5).

Neumann, J. and Gutt, D. (2017). A homeowner's guide to airbnb: theory and empirical evidence for optimal pricing conditional on online ratings. *Association for Information SystemsAIS Electronic Library (AISeL).*

Powers, D. and Xie, Y. (2008). *Statistical methods for categorical data analysis.* Emerald Group Publishing.

Quattrone, G., Proserpio, D., Quercia, D., Capra, L., and Musolesi, M. (2016). Who benefits from the sharing economy of airbnb? In *Proceedings of the 25th international conference on world wide web*, pages 1385–1394. International World Wide Web Conferences Steering Committee.

Varma, A., Jukic, N., Pestek, A., Shultz, C. J., and Nestorov, S. (2016). Airbnb: Exciting innovation or passing fad? *Tourism Management Perspectives*, 20:228–237.

Xie, K. and 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.

Yee, T. W. (2008). vgam family functions for bivariate binomial responses.

Yee, T. W. (2015). *Vector generalized linear and additive models: with an implementation in R*. Springer.

Yee, T. W. et al. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, 32(10):1–34.

Zervas, G., Proserpio, D., and Byers, J. (2015). A first look at online reputation on airbnb, where every stay is above average.