DO TOURISTS' VALUE URBAN GREEN SPACE?

A study into the effect of urban green space on Airbnb pricing in Amsterdam.

Abstract. In order to indicate to what extent urban green space is valuable to tourists, this study provides an overview of the effect of urban green space on the rental price of Airbnb accommodations in Amsterdam. As a first step, using ArcGIS, the distance between Airbnb accommodations and urban green space is measured. This measurement only includes attractive urban green space. Urban green space is particularly appreciated by residents which is possibly similar for tourists. Subsequently, to see how tourists appereciate these attractive green spaces, the hedonic price model was used to estimate and evaluate the effect of attractive urban green space on 7,770 Airbnb accommodations in Amsterdam. The findings indicate that the rental price of Airbnb accommodations increase with 6.92 percent for properties located within 0.25 km. The estimated price effect decays with distance from attractive urban green space and is negligible for Airbnb accommodations located further than 1 km away. The findings of this study may inform policy makers about the value of attractive urban green space for tourists, which can be considered in future policy making to counteract potential competition for attractive urban green space by both residents and tourists.

Keywords. Urban green space, Airbnb, rental price determinants, tourist value

Michelle de Jonge
Master Thesis Real Estate Studies
June 29, 2021

COLOFON

Title	Do tourists' value urban green space?
Subtitle	A study into the effect of urban green space on Airbnb pricing in Amsterdam
Version	29-06-2021
Author	M. (Michelle) de Jonge
Student number	S3549682
E-mail	m.de.jonge.26@student.rug.nl
Supervisor	dr. M.N. (Michiel) Daams

TABLE OF CONTENTS

1. INTRODUCTION

Urban green spaces provide numerous of benefits for the urban area. A green environment consisting of as parks, gardens and urban forests enhance the well-being of people, improve the air-quality and provide a healthy living environment. These benefits positively affect the quality of life in the city (Green cities.eu, 2018; Het Parool, 2020). A growing number of European cities focus on the presence of green space in their city as it is an essential contribution to the health and social well-being of people living in the city (Gemeente Amsterdam, 2015). The positive influence of urban green spaces is not primary limited to its residents. The presence of urban green space could be an attraction pole for tourists, because it creates a sense of place, and it improves the city's image. That both increase the attractiveness of the city, which is claimed by some to result in more tourism (CBRE, 2020; Kerr, 2017). A growing importance of urban green space, for both residents and visitors, could result in possible competition for urban green spaces in the city. It is therefore important with regard to policy makers, that studies provide deeper insights into how tourists value urban green spaces (Gemeente Amsterdam, 2020; Green cities.eu, 2018). How tourists' value urban green space closes to their place of overnight stay, may in particular depend on whether that green space has particular use value for them.

Prior academic literature regarding urban green spaces have examined the uses of urban green space by both local residents and tourists. For residents, urban parks are part of their daily life offering natural area for recreational activities or relaxation in spare time. Tourist's use urban green spaces as a place to experience a city's unique culture and in a number of European cities urban green spaces are a tourist attraction (Terkenli, et al., 2017; Song and Sim, 2021. Less evidence exists which specifically address the potential price effect of the presence of urban green space for tourists. Literature regarding price effects of urban green space, instead, center around the effect on nearby residential property prices (Conway, et al., 2010; Daams, et al., 2016; Daams, et al., 2019; Gibbons et al., 2014; Tajima, 2003). These studies show that residents have a higher willingness to pay for properties with closer proximity to urban green space. This means that a positive premium is paid for properties located near urban green space, but for tourism accommodations pricing this relationship remains unclear.

While tourists can choose extensively of accommodations to stay when visiting a city, Airbnb[1] has in recent years become particular relevant for our type of study. Derived from the fact that the use of Airbnb has become extensively larger, covering many residential areas in cities, unlike hotels, and that urban green space is playing a larger role in the urban area for tourists, it can be stated that this study includes relatively broad urban coverage. Yet, other studies have made limited use of this to assess how tourists value amenities.

[1] Airbnb is an online platform for residential rentals and tourist activities (Airbnb, 2021).

Airbnb pricing literature have so far investigated only driving forces other than green space behind rental prices of Airbnb accommodations. Physical, locational and host characteristics have a significant influence on the rental price of Airbnb accommodations (Gibbs, et al. 2018; Wang and Nicolau, 2017). Locational characteristics are often the location with regards to the city center and its general touristic amenities, but do not study the relationship to urban green space (Wang and Nicolau, 2017; Perez Sanchez, et al., 2018). As such, this study is to provide insights on the effect of urban green space on Airbnb pricing in order to find out to what extent urban green space is valuable to tourists. For policy makers evidence that tourists, equal to residents, value the presence of attractive urban green spaces is beneficial, as they try to ensure the right balance and distribution of green spaces in the city center (Gemeente Amsterdam, 2020). Regarding how tourists value green space in the city which they visit, we ask:

''What is the effect of urban green space on the rental price of nearby Airbnb accommodations in Amsterdam?''

The empirical analysis, to identify the effect of urban green space on Airbnb accommodations rental prices, is conducted using hedonic modelling. This type of model controls for quality differences between Airbnb accommodations (Chen and Xie, 2017; Gibbs, et al., 2017) and is also extensively adopted in both Airbnb and urban green space literature (Chen and Xie; Conway, et al., 2010; Daams, et al., 2016;). Using ArcGIS, the distance between Airbnb accommodations ($N = 7{,}770$) and urban green space in Amsterdam is measured. By following Daams et al. (2016) a decision is made to primary include attractive urban green space in the analysis because with general green space, derived from land use data, it is assumed that people perceive green space with similar characteristics the same. The analysis in this study could provide new insights to what extent tourists value attractive green space.

The remainder of this study is organized as follows. In section 2, a theoretical overview is established of the most important rental price determinants resulting in a conceptual model and section 3 gives an overview of the data and the empirical approach to modelling Airbnb prices. Section 4 presents the empirical results and section 5 offers the discussion. Section 6 concludes.

2. THEORETICAL FRAMEWORK

2.1 Airbnb price mechanism

Pricing is a critical factor in the long-term accomplishment of consumer-to-consumer business models, as Airbnb, and stands at the core of competitive advantage in the sharing economy (Cai, et al., 2019; Gibbs, et al., 2017). It is therefore essential to take a look at price construction and current implications in Airbnb pricing. The platform has no control over the final rental price and allows for customized daily prices, weekend prices and discounts (Gibbs, et al., 2017). Airbnb developed two types of algorithms, price tips and smart pricing, in attempt to give price directions to service providers of the platform. Price tips inform service providers how likely the accommodation will be booked on a particular day and will give prices of similar accommodations in the neighborhood. Smart pricing can be determined as setting a minimal and maximum rental price of the accommodation. The purpose of these partial algorithms is that Airbnb pricing is set more effectively (Ye, et al., 2018). However, service providers are not obligated to use these tools and therefore have service providers total control over the price setting. There is no clear way of constructing the price of an Airbnb accommodation and as result are Airbnb accommodations could be rented out higher than suggested (Gibbs, et al., 2017).

Further implications arise in the demand estimation. The reason for this is that the Airbnb pricing can be described as dynamic pricing[2] in which the focus particularly lies om homogeneous products (Ye, et al., 2018). As result, the exact rental price of an Airbnb accommodation is difficult to determine. Two causes of frequent adjustment of rental prices can be proposed. First, the time variety in the demand function. Due to seasonality and events is the demand exposed to large seasonal variation. Airbnb search quantity is higher during summertime which suggests a seasonal pattern (Ye, et al., 2018). However, seasonal patterns can differ between countries or be stronger present in certain countries. Furthermore, special events will result in a higher occupancy when a particular event is held. This results in higher demand on that specific date and possibly drive up the rental price (Ye, et al., 2018; Li, et al., 2015). Secondly, there is variation in Airbnb accommodations type. Every Airbnb accommodation is different and have their own identity (Ye, et al., 2018). Therefore, prices could differentiate. Resulting from the occurred implications in Airbnb pricing, it is difficult to construct the rental price. However, how rental prices of Airbnb accommodations are influenced by certain determinants is studied in previous academic literature (Chattopadhyay and Mitra, 2019; Gibbs, et al., 2017; Wang and Nicolau, 2017). The following section elaborates on how urban green space could be a possible determinant of Airbnb rental prices.

[2] Dynamic pricing can be determined as pricing of products and services that can frequently and effortlessly be adjusted (den Boer, 2013).

2.2 Urban green space as rental price determinant

Modern cities provide urban green spaces including public parks, trees and landscapes (Chan and Marafa, 2017). These urban green spaces can provide social and environmental services that positively affect the quality of life in the city, to the benefit of not only the city's residents but also its visitors (McPherson, 1992). In previous academic literature regarding price effects of urban green space is found that urban green spaces have a significant effect on residential property values, resulting in a higher willingness to pay of residents (Conway, et al., 2010; Daams, et al., 2016; Daams, et al., 2019). Daams et al. (2016) find a price premium of 16% for properties located within 0.5 km from green space. This price effect smoothly decays with distance from urban green space and is insignificant when a property is located further than 7 km away from urban green space. Therefore, capitalize urban green spaces not after this distance (Daams, et al., 2016). Czembrowski et al. (2019) similarly find that close proximity of urban green space in the city positvely effect property prices, however the effect-size crucially depends on green space's amenity-level. This means that the price effect is significantly stronger when the urban green space contains a high number of characteristics and activities. The price effect of urban green space can be related to the benefits urban green space provide to residents living in the city. Urban green spaces provide easy access to the natural environment and contribute to a healthy urban climate (Baur and Tynon, 2010; Leeuwen, et al., 2011). Residents further benefit from urban green space in a social context. Urban green spaces offer a location for physical activities which increases the well-being of people (Brown, et al., 2014). It also promotes healthy living environment and provides a place for social interaction (Keng Lee, et al., 2015).

The presence of urban green space could similarly provide benefits for tourists visiting a city, therefore a significant price effect could also be projected on the willingness to pay for an overnight stay of tourists. Urban parks namely enhance the recreational, aesthetic and cultural symbolization value of a city (Leeuwen, et al., 2011; McPherson, 1992). The aesthetic value of parks and landscapes increase the willingness to stay of tourists and promote the urban area as a tourist destination (Chan and Marafa 2017; Terkenli, et al., 2017). Urban green space could further be of benefit for tourists as incidental locations, perhaps to seek relaxation from the busy crowd, or as a suite of sights the visitors plan to visit (Terkenli, et al., 2020). A particular visit to a high-profile park, for example Vondelpark in Amsterdam, Hide park in London and Parc Güell in Barcelona are tourist attractions on their own and are therefore extremely popular to visit (Konijnendijk, 2013; Terkenli, et al., 2017). These parks symbolize the lifestyle and image of a city and tourists can temporily experience the unique culture a city has to offer (Song and Sim, 2021).

Recent years, various cities profile itself as green urban tourist destinations, for example Minneapolis in the United States, which have resulted in increased tourism (Terkenlis, et al., 2020; Dodds and Joppe, 2001). The establishment of a green city image has become an important city branding concept to promote sustainable development, improve the quality of life and increase competitive advantage to attract tourism recent years. Green cities are associated with an increased positive city image which equals increased intention to visit a city (Kalliopi, 2015). A possible explanation why the presence of urban green spaces result in a positive city image, is that it increases the attractiveness of the city, which will attract more tourism (Leeuwen, et al., 2011). It can therefore be concluded urban green spaces are an enrichment to urban tourism, which could significantly and positively contribute to the willingness to pay of tourists for their overnight stay (Airbnb) (Deng, et al., 2010).

2.3 Other rental price determinants Airbnb accommodations

The rental price of Airbnb accommodations could further be explained by a set of rental price determinants relating to property, host and site-specific characteristics. Property characteristics focus on explaining elements only related to the property itself. These are direct observable characteristics (Ert et al. 2016). Characteristics room type, property type and the number of bathrooms available correspond with a positive impact on Airbnb pricing (Chen and Xie, 2017; Chica-Olmo et al., 2020; Chattopadhyay and Mitra, 2019; Lorde, et al., 2019). When looking at how the accommodation is rented out, the rental price of an entire accommodation is higher, compared to when the accommodation is shared. When the accommodation offers a private room in comparison to a shared home, the rental price will similarly be higher (Abrate and Viglia, 2017; Chen and Xie, 2017; Dogru and Pekin, 2017). Magno et al. (2018) who use the number of accommodates as a measure of size, identify a price premium when the number of accommodates increase. Furthermore, state the researchers that visitors are more likely to pay a greater rental price for an apartment than other property types (Magno, et al., 2018). Chen and Xie (2017) indicate that the property type has an influence on the rental price of Airbnb accommodations. The researchers find that people are more likely to pay for an apartment than for a house or other property types. Lorde et al. (2019) report contrary findings, the authors conclude that larger property types, house or villa, have a positive significant effect on the rental price.

Host characteristics can be described as important rental price determinants of Airbnb accommodations. Wang and Nicolau (2017) identify host characteristics as quality signaling factors. Dogru and Pekin (2017) make the same identification and argue that a superhost status[3] corresponds with a higher charged rental price. As the superhost status has a positive relation with the rental price, the tourists are thus willing to pay a premium above the rental prices for a superhost status (Chica-Olmo, et al., 2020; Wang and Nicolau, 2017). An accommodation with higher review score correlates with a higher rental price (Chen and Xie, 2017; Gibbs, et al., 2018; Gutt and Hermann, 2015). Chen and Xie (2017) find a weak significant effect between the review score and the rental price. A possible explantion is that visitors of Airbnb accommodations wish social interactions with the host, however due to commercialization of the market to accomodate mass tourists this will diminish. As result, when less

[3] A superhost status can be described as a quality measure for Airbnb accommodations. Four criteria need to be satisfied to secure the Airbnb superhost status. These are (1) ten completed trips in the previous year, (2) five star ranking from at least 80% of the visitors, (3) response rate of at least 90% and (4) a low cancellation rate (Gunter, 2018).

social interaction exists, custumer reviews are lower (Chen and Xie, 2017). Furthermore, has the number of reviews per year a negative impact on the price. Each additional review on the accommodation results in a price decrease (Abrate and Viglia, 2017; Gibbs, et al., 2018; Wang and Nicolau, 2017). When the number of reviews is high, visitors possibly do not trust the reviews and are not willing to pay high prices (Magno, et al., 2018). Furthermore, is the rental price obtained by a professional host[4] higher compared to a non-professional host (Li, et al., 2015). A possible explanation is that professional hosts are more likely to obtain a superhost status and thus possibly higher rental prices (Gunter, 2018). The cancellation policy a host is adopting partly indicates the host effort. Host effort can be explained as an extension of functionality of Airbnb accommodations, this increases the utility to visitors and eventually improve consumer valuation. When a host adopts a moderate cancellation policy it positively contributes to the rental price (Chen and Xie, 2017). Another indicator of host effort is host response time which is the quickness of responding to reservation requests. The rental price decreases, as the response time increases (Mauri, et al., 2018).

Site-specific characteristics relate to the location of the Airbnb accommodation. Previous literature focusses on the proximity from the city center. The distance to the city center has a significant negative effect, therefore the further the Airbnb accommodation is located from the city center, the lower its rental price will be. This means that the rental price of an Airbnb accommodation decreases with distance from the city center (Gibbs, et al., 2017; Gunter, 2018; Wang and Nicolau, 2017). A possible explanation could be that tourists favor to rent accommodations that are close to areas where main tourist attractions take place. Therefore, a strong spatial association exists between the city center and the Airbnb accommodation. Consequently, most Airbnb accommodations are located in the center (Gutiérrez, et al., 2017). Cai et al. (2018) includes amenities in their research into rental price determinants of Airbnb accommodations. The researchers investigate, among others, if there is an effect present between the distance to a tourism attraction and the rental price of Airbnb accommodations. Results show that distance to tourism attractions has a significant negative effect on the rental price, meaning the further away from the tourism attraction, the lower the rental price will be. It can be concluded that amenities play a significant role in social structuring of the urban space (Tivadar and Jayet, 2006). It could therefore be suggested that urban green space, also determined as amenity, similarly could have an effect on Airbnb pricing.

2.4 Hypothesis

Urban green space is an important characteristic for the quality of life in the city center and provides benefits for both tourists and residents (Baur and Tynon, 2010; Konijnendijk, 2013; Leeuwen, et al., 2011; McPherson, 1992; Terkenli, et al., 2017; Song and Sim, 2021). In residential pricing

[4] Non-professionals are hosts with a single Airbnb accommodation listed and professionals as hosts with multiple Airbnb accommodations listed (Li, et al., 2015).

literature is found that urban green space positively contributes to property prices (Czembrowski et al., 2019; Conway, et al., 2010; Daams, et al., 2016; Daams, et al., 2019). Whereas urban green space is similarly providing benefits for tourists, a possible relation could exist between urban green space and the rental price of Airbnb accommodations. Based on the theoretical framework[5], the following hypothesis can be defined:

'The presence of attractive urban green space has a positive effect on the rental price of Airbnb accommodations'

This study's analysis will provide insights on the effect of urban green space and the rental price of Airbnb accommodations in order to find out to what extent urban green space is valuable to tourists.

[5] Based on the theoretical framework a conceptual model is established. A detailed description is given in appendix 1.

3. METHODOLOGY

3.1 Data overview and limitations

The data necessary to analyze whether urban green space has an effect on Airbnb rental prices consist of two different datasets. Airbnb data were obtained from Inside Airbnb (Inside Airbnb, 2019). Inside Airbnb is a non-commercial and independent organization that provides more detailed information on Airbnb accommodations. The dataset is publicly available and gives insights on how Airbnb is used in different cities around the world. Inside Airbnb focuses on facilitating research and making data about Airbnb transparent (Inside Airbnb, 2019; Benítez-Aurioles, 2017). For this particular research the dataset of Amsterdam is used (Inside Airbnb, 2020). This dataset is used as a source of data in previous academic studies (Benítez-Aurioles, 2017; Cai, et al., 2019; Sheppard and Udell, 2016; Wang and Nicolau, 2017). However, limitations of this dataset exist. The location of each Airbnb accommodation is anonymized. This means that the pointed location on the map is not the real location of the Airbnb accommodation. The actual location will be at approximately 0-150 meters distance from the pointed location of the Airbnb accommodation on the map. Likewise, are Airbnb accommodations located in the same building similarly anonymized and scattered in the surrounding area (Inside Airbnb, 2020). Furthermore, the dataset is a snapshot of the accommodations available at a particular time. Therefore, it is possible that not all accommodations captured in the analysis, because for example the Airbnb accommodation is not rented out at that time or hosts could have made the Airbnb accommodation unavailable at particular times. Sheppard and Udell (2016) similarly conclude this and suggest that hosts can easily enter the market by creating a listing and consequently exit the marktet. Therefore, a possibility exists that not all listings are captured in the analysis. This could possibly result in a measurement error (Sheppard and Udell, 2016). It is further unknown whether an airbnb accommodation is actually rent out. This means that it is unidentified if the rental price in the dataset is the actual price the consumer paid for when booked or whether it is an indicative rental price of an available airbnb accommodation. Further concerns, relating to the rental price of airbnb accommodations in the dataset, arise in the price mechanism of Airbnb. The rental price is not controlled by Airbnb, the hosts are therefore permitted to form the rental price. Therefore, could rental prices be higher than they should be.

Urban green space data were obtained from a dataset that combined Greenmapper data and land use data (Greenmapper, 2020; Daams, et al., 2016). This dataset provides detailed information about green spaces in the Netherlands. Greenmapper data is an open-access database viewer created by the university of Groningen. The dataset contains crowd-sourced data about attractive nature on different levels (local, regional, national and global level). The dataset uses a PSGIS-based survey in which survey respondents mark nature on a map that is attractive, valuable or important. Daams et al. (2016) developed a method in which land use data is combined with the Greenmapper data. This study focuses on the effect of natural spaces that are pointed out as attractive. The Greenmapper survey data identifies people's opinion on attractive urban green space, which is combined with the general identified green space data in the Netherlands. The researchers developed this method in order to counteract potential heterogeneity in preference of green spaces. The researchers argue that previous studies evaluate natural space with the same land use as homogenous, but there exists higher heterogeneity among natural spaces. Because it is possible that recreational infrastructure is different or diversity in landscape exists (Daams, et al., 2016). The combination of land use data and the survey on attractive green space have resulted in cluster areas and these are used in this study to locate attractive green spaces in Amsterdam.

3.2 Geographic information system

In order to perform an analysis on the relation between presence of urban green space and rental price of Airbnb accommodations, an analysis in the spatial program ArcGIS is conducted to measure the distance between Airbnb accommodations and urban green space. After spatial analysis, the datasets are combined in order to do the statistical analysis. Using the spatial program ArcGIS, the distance between the location of Airbnb accommodations and the presence of attractive urban green space is measured. Both datasets were used for this purpose. As result, a new variable was created indicating the distance between the locations of Airbnb accommodations and attractive urban green space. Figure 1 indicates

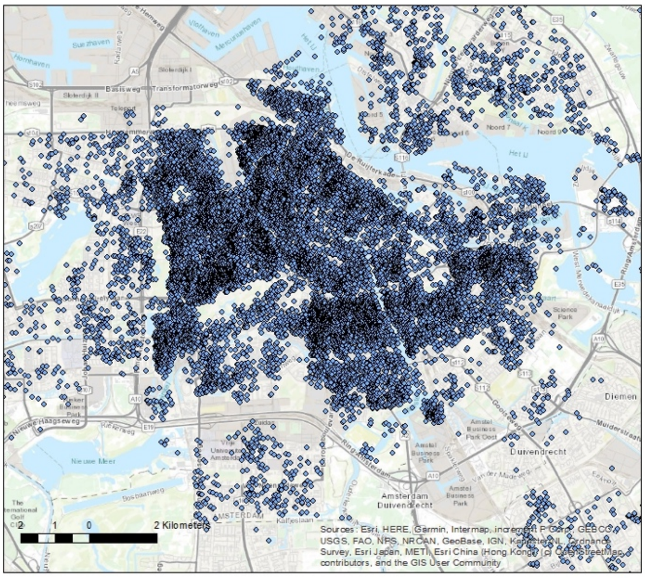

Figure 1: Overview Airbnb accommodations Amsterdam.
Source: own work

Airbnb accommodations located in Amsterdam[6]. The longitude and latitude coordinates from the Airbnb dataset are used to point out the locations of Airbnb accommodations on the map. The longitude and latitude coordinates are added into the display XY function in ArcGIS and the exact coordinates are shown on the ArcGIS map. As result, a good overview of Airbnb accommodations in Amsterdam is

[6] Not all Airbnb accommodations projected in the figure were used in the analysis due to data operationalization.

created. As shown in figure 1 are most Airbnb accommodations concentrated around the city center. Figure 2 indicates attractive urban green spaces in Amsterdam, using cluster areas (Daams, et al., 2016). The distance between attractive green space and the Airbnb accommodations has been measured using the NEAR tool in ArcGIS. It is a multipoint to polygon measurement in which the nearest distance between the input feature and near feature is measured. The input feature is indicated as the Airbnb accommodations and the near feature is the attractive urban green space.

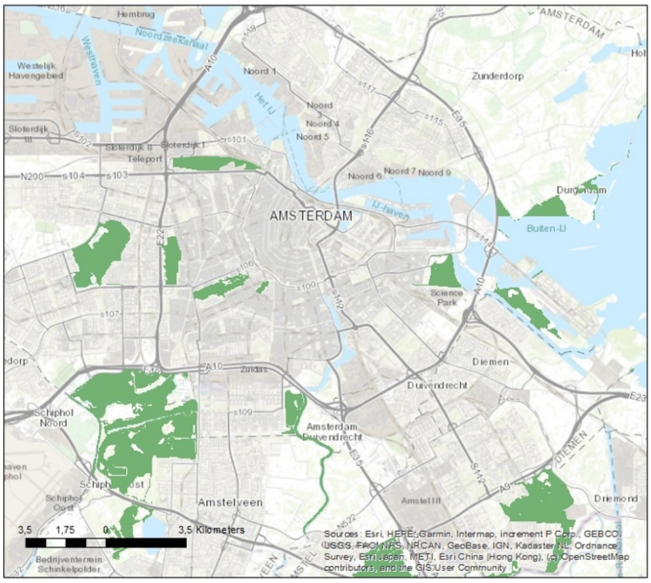

Figure 2: Attractive urban green space in Amsterdam.
Source: own work

Additionally, the distance between Airbnb accommodations and the city center is measured. This relationship is studied in previous academic studies (Zhang, et al., 2017; Gibbs, et al., 2017; Wang and Nicolau, 2017; Qiu, et al., 2018). This variable will not be used as interest variable but is indicated as control variable. A different approach to measure the distance between Airbnb accommodations and the city center has been taken into account compared to the previous distance measurement. A measurement between the Airbnb accommodations and the dam square has been carried out. This resulted in a point-to-point measurement in which the input feature is the Airbnb accommodations and the near feature the dam square. In previous academic studies this same approach has been carried out. Gibbs et al. (2017) used the city's city hall to mark a point as city center. Cai et al. (2018) chose an important metro station located in city center to measure the distance between Airbnb accommodations and the city and Chica-Olmo et al. (2020) marked an important and busy street as city center point.

3.3 Descriptive statistics

Table 1 presents the descriptive statistics for the variables[7] used in this study. The descriptive statistics describe the most important characteristics of the data using the mean, standard deviation,

[7] An overview of the dependent, independent and control variables is given in appendix 2.

minimum and maximum of all the variables. The table is divided in three categories: dependent variable, independent variables and control variables. The category control variables are subdivided into property characteristics, host characteristics and site-specific characteristics. The dependent variable is a log transformation of the rental price of Airbnb accommodations. The independent variable is the distance from an Airbnb accommodation to attractive green space in kilometers. The distance between Airbnb accommodations and attractive urban green space ranges from 0.00182 and 3.915 kilometer. The mean is 0.996 kilometer. To extend the analysis on the price effect of urban green space on Airbnb accommodations, measures of proximity to attractive urban green space were included. The effect of distance to the nearest attractive urban green space on Airbnb pricing is estimated using discrete distance intervals following Daams et al. (2016). Dummy variables indicate these discrete distance intervals and are categorized into 0-250m, 250-500m, 500-750m, 750-1000m, 1000-1250m, 1250-1500m and >1500m.

Table 1: Descriptive Statistics

Variable	Definition	Mean	*SD*	Min.	Max.
Dependent variable					
Log price	Log rental price of an Airbnb accommodation	4.939	.520	2.197	6.907
Independent variables					
Distance attractive urban green space	Distance Airbnb accommodation from attractive green urban space in km	.996	.633	.00182	3.915
Attractive green space category	Distance attractive to nearest attractive urban green space in categories	2.836	1.969	0	6
	0-250	.089	.285	0	1
	250-500m	.152	.359	0	1
	500-750m	.161	.367	0	1
	750-1000m	.162	.368	0	1
	1000-1250m	.139	.346	0	1
	1250-1500m	.103	.304	0	1
	>1500m	.194	.396	0	1
Control variables					
Property characteristics					
Property type	Property type of an Airbnb accommodation (1 = apartment, 2 = house 3 = other)	1.429	.733	1	3
Room type	Room type of an Airbnb accommodation (0 = entire home/apartment, 1= shared/private room)	.306	.461	0	1
Accommodates	The number of accommodates	2.903	1.445	1	17
Bathrooms	The number of bathrooms	1.206	.405	.5	8
Host characteristics					
Host response time	The host's response speed (1 = within a day 2=within a few hours 3= within an hour)	.819	.499	1	3
Host response rate	The host response rate (0 = less than 100%, 1 = 100%)	.818	.386	0	1
Host is superhost	Whether a host is a superhost (0 = no, 1= yes,)	.287	.453	0	1

Host listings count	Host's number of accommodations listed on Airbnb	3.19	16.71	1	751
Host identity verified	Whether the host's identity is verified (0 = no, 1 = yes)	.416	.493	0	1
Number of reviews	The number of reviews of an Airbnb accommodation	43.838	71.837	0	821
Review scores value	Average review score of an Airbnb accommodation	9.197	.722	2	10
Cancellation policy	Strictness of cancellation policy (0= less strict, 1 = strict)	.468	.499	0	1
Site-specific characteristics					
Distance city center	Distance Airbnb accommodation from dam square in KM	2.611	1.686	.020	11.539
N = 7,770					

3.4 Data selection

The total dataset consists of 19.643 observations which are all Airbnb accommodations rented out in Amsterdam. However, not all of this data is fit for analysis. Corrections were necessary to prepare the dataset for analysis. The first step in preparing the dataset was to convert string variables to numeric variables. In doing so, 155 observations were lost. The second step was correcting for missing values in the dataset. These missing values were found in the variables rental price, host response rate, host response time, host is superhost, host listings count and bathrooms. After correcting for missing values in the dependent, independent and control variables, a total number of 8.942 observations were present in the dataset. After correcting for missing values and the transformation of the string variables into numeric variables further analysis of the variables is carried out. Resulting in a total number of 7.770 observations for the analysis, this is 39.55% over the total sample. Brooks and Tsolacos (2010) state a rule of thumb that a sample size of at least 100 is desirable for statistical analysis. The sample size is therefore considerate to be large enough as the number of observations of 7.770 is larger than the stated 100 observations. Further corrections to prepare the dataset for analysis are described in appendix 3.

3.5 Empirical model

The aim of this study is to indicate to what extent urban green space has an effect on the rental price of Airbnb accommodations. For this purpose, hedonic modelling is used to describe and evaluate a possible effect between the dependent variable, the rental price of Airbnb accommodations and the independent variable, related to urban green space and control variables. In previous research regarding Airbnb rental price determinants (Cai, et al., 2019; Chen ans Xie, 2017; Perez-Sanchez , et al., 2018) and urban green space (Czembrowski, et al. 2016; Daams, et al., 2019) hedonic modelling is similarly adopted. The hedonic price model was developed by Rosen (1974) in order to estimate implicit prices and demand for attributes of heterogeneous goods. Sheppard (1998) indicates that hedonic price functions are estimated to develop price indices accounting for changes in quality of goods and to analyze attributes for heterogenous goods. The hedonic price function is written as:

$$P = f\ (z, \varepsilon) \tag{0}$$

Where P is the rental price of an Airbnb accommodation, z is the vector of the urban green space, property characteristics, host characteristics, site-specific characteristics and ε is the error term. The hedonic price model is adopted to determine a relationship between urban green space and the rental price of Airbnb accommodations. To estimate how the characteristics, influence the rental price of Airbnb accommodations, the model is estimated using Ordinary least squares (Perez-Sanchez , et al., 2018). In order to use the Ordinary least squares, several assumptions have to be satisfied. These assumptions are linearity (average value of the errors is zero), homoscedasticity (The variance of the errors is constant and finite over all values), no autocorrelation (the errors are statistically independent of one another, independence (independent variables are not correlated with the error term of the estimated equation), normality (the errors are normally distributed) (Brooks and Tsolacos, 2010). A detailed description of testing the OLS assumptions can be found in appendix 9. In order to use the OLS method, the hedonic price function has to be specified as a parametric model (OECD, 2013). The hedonic linear specification model is written as:

$$P = \beta_0 + \sum_{k=1}^{k} \beta_k z_{nk} + \varepsilon_n \tag{1}$$

Where, β_0 is the intercept and β_k are the characteristics parameters to be estimate. Using equation (1) as starting point, the following baseline specification only including property characteristics, host characteristics and site-specific characteristics, is specified:

$$\begin{aligned} Ln(Price) = {} & \beta_0 + \beta_1 PropertyType + \beta_2\ RoomType + \beta_3\ Accommodates + \ \beta_4\ Bathrooms \\ & + \ \beta_5\ HostResponseTime + \beta_6\ HostResponseRate + \beta_7 Superhost \\ & + \beta_8\ HostListingCount + \ \beta_9\ IdentityVerified + \ \beta_{10}\ NumberOfReviews \\ & + \ \beta_{11}\ ReviewScoreValue + \beta_{12}\ CancellationPolicy + \beta_{13}\ DistanceCityCenter \\ & + \ \varepsilon \end{aligned} \tag{2}$$

Where, β_0 is the intercept; Ln (Price) is the logarithm of the rental price of an Airbnb accommodation; $PropertyType$ presents the property types (apartment, house, and other); $RoomType$ refers to the type of rooms (entire home/apartment or private room); $Accommodates$ refers to the number of accommodates; $Bathrooms$ is the number of bathrooms in an Airbnb accommodation; $HostResponseTime$ refers to the host's response speed (within a day, within a few hours and within an hour); $HostResponseRate$ refers to the rate of response of a host; $Superhost$; refers to whether a host is a superhost (yes or no); $HostListingCount$ represents the number of accommodations a host rents out on Airbnb; $IdentityVerified$ represents whether the identity of a host is identified (yes or no); $NumberOfReviews$ refers to the number of reviews an Airbnb accommodation has; $ReviewScoreValue$;

is the average review score of an Airbnb accommodation; $CancellationPolicy$ represents the strictness in cancellation policy (strict or less strict); $DistanceCityCenter$ refers to the distance from an Airbnb accommodation to Dam square and ε presents the error term.

The baseline specification indicates the variables found in the literature that are important rental price determinants of Airbnb accommodations relating to property characteristics, host characteristics and site-specific characteristics. Whereas specification (3), in order to determine to what extent urban green space has an effect on the rental price of Airbnb accommodations, includes one important urban green space variable. This results in the following main specification, referred to as urban green space model.

$$\begin{aligned} Ln(Price) = {} & \beta_0 + +\beta_1\, DistanceAttractiveGreenSpace + \beta_2 PropertyType + \beta_3\, RoomType \\ & + \beta_4\, Accommodates + \beta_5\, Bathrooms + \beta_6\, Bedrooms + \beta_7\, Beds \\ & + \beta_8\, HostResponseTime + \beta_9\, HostResponseRate + \beta_{10}\, Superhost \\ & + \beta_{11}\, HostListingCount + \beta_{12}\, IdentityVerified + \beta_{13}\, NumberOfReviews \\ & + \beta_{14}\, ReviewScoreValue + \beta_{15}\, CancellationPolicy + \beta_{16}\, DistanceCityCenter \\ & + \varepsilon \end{aligned} \tag{3}$$

Where, β_0 is the intercept; Ln (Price) is the logarithm of the rental price of an Airbnb accommodation; $DistanceAttractiveGreenSpace$ is the distance from Airbnb accommodations to attractive green space.

The baseline and the main both address the problem of spatial autocorrelation. Wilhelmsson (2002) argues that samples of real estate prices strongly relate to the location of the property. This study includes locational variables (e.g., distance attractive urban green space and city center) which leads to spatial effects in the hedonic model, as they represent spatial interactions and spatial structures. Spatial effects include spatial dependence[8] which can be explained as the dependency between observations based on location**.** To be more precise, an observation at a certain location depends on observations at another location (Conway, et al., 2008). Spatial dependence arises from spillover effects and spatially correlated variables that are omitted. For example, the inclusion of all neighborhood attributes is rarely fulfilled in previous studies resulting in omitted variable bias (Wilhelmsson, 2002). This is similarly concluded in research of (Abbot and Klaiber, 2011) as unobserved variables often occur at varying spatial scales. Czembrowski et al. (2016) suggest when using spatial data, one has to assume the presence of spatial autocorrelation. Spatial autocorrelation could be caused by imperfections of the model, meaning omission of a variable related to price. Or prices might depend directly on the

[7] Griffith (1992) p. 267 defines the concept of spatial dependency as: ''The dependencies that exist among observations that are attributable to the relative locations or variable values in geographic space''.

surrounding area, because the neighborhood is more expensive. For the estimation technique OLS[9] to be accurate, meaning unbiased, efficient and consistent, spatial dependence cannot be present (Wilhelmsson, 2002). When the analysis contains spatially autocorrelated data the assumptions are no longer valid and OLS produce inefficient coefficient estimates (Gangodagamage, et al., 2008; Conway, et al., 2008; Ham, et al. 2012). Following Daams et al. (2016) by including clustered standard errors at street-level, six-digit zip codes, will address the problem of spatial autocorrelation. Furthermore, include both model specifications spatial controls at neighborhood level. In total, 7 neighborhoods are included.

It is expected that urban green space will have an influence on the rental price of Airbnb accommodations, as similarities exists with previous literature on the price effect on property prices (Daams et al., 2016; Daams et al., 2019). It is therefore expected that the rental price of an Airbnb accommodation will increase when located nearby attractive urban green space. For the control variables it is expected that property characteristics, the number of bathrooms and accommodates have a positive effect on the rental price. Property type apartment is preferred over house or other property types and the possibility of renting the entire home positively effects the rental price. It is further expected that the rental price of Airbnb accommodations increases when the host response time is fast, the host response rate is 100% and if the host has the super host status. The host is assumed to be more professional when managing more listings, it is therefore expected that the rental price increases when the listing count is high. A high number of reviews negatively impacts the rental price whereas a high review score value positively affects the rental price. What cancellation policy a host implements partly indicates the host effort. When the cancellation policy is strict, the rental price is expected to increase. At last, the rental price is higher in the city center, therefore the rental price will decrease with distance from the city center.

In this study, the chow test is conducted as part of a robustness analysis (Chow, 1960). This robustness check examines whether the coefficients of different groups are equal. The different groups are characterized by Airbnb accommodations with a view on attractive urban green space (located within 100m)[10] and Airbnb accommodations with no view on attractive urban green space (located beyond 100m). The chow test examines to what extent a difference in the parameters of Airbnb accommodations with a view or no view on attractive urban green space is present.

[9] Using the OLS estimation method it is necessary that the multicollinearity is detected because in a case of multicollinearity, changing a variable will cause the value of the coefficient on other variables to be different (Brooks and Tsolacos, 2010). According to Lin (2008) is multicollinearity not a violation of the assumptions of regression analysis, but it can cause other serious problems such as the parameters of the estimates could be unclearly large, not be significant or have another sign than expected. Chen and Rothschild (2010) state that multicollinearity is often a problem in hedonic price models. Besides using the correlation matrix (appendix 7) to investigate multicollinearity, the variance inflation factor (VIF) is used to verify multicollinearity. The VIF is a method for detecting multicollinearity in which a VIF greater than 10 is an indicator of multicollinearity (Chen and Rothschild, 2010; Lin, 2008). Magno et al. (2018) proposes a cut-off level of 5. In this study should multicollinearity not be an issue. All variance inflation factors (VIF) are below level 5, suggesting that no multicollinearity problem exists in the model specifications. An overview of the VIF values is given in appendix 8.

[10] A variable concerning view on urban green space is not present in the data, therefore by following Daams et al. (2016) a proxy measure is included in the analysis.

4. EMPIRICAL RESULTS AND DISCUSSION

4.1 Baseline specification results

Table 2 reports the regression results for the model specifications. Model (1), the baseline model, indicates regression results of the most important variables found in the literature in relation to rental price determinants of Airbnb accommodations. These variables consider property, host and site-specific characteristics. The baseline model does not consider the distance to urban green space variables. The null hypothesis of model specification 1 is property characteristics, host characteristics, site-specific characteristics have no effect on the rental price of Airbnb accommodations. The coefficients are tested whether they are far enough away from zero to conclude that it is significantly different from zero. The R-squared is used to find out how well the regression model fits the data. The R-squared in the baseline model is 0.469. This means that 46.9 percent of the variation in rental price of Airbnb accommodations is explained by the variation of the independent variables. The regression results for the baseline model are as expected and in line with previous literature.

The results show that variables property type, room type, the number of accommodates, the number of bathrooms is significantly different zero at a 99 percent probability level. The sign and the size of the coefficient explain positive or negative impact and the exact value of the coefficient. Both coefficients on property type have a positive sign. Suggesting that the rental price of an Airbnb accommodation in a house or other property type is higher than the rental price of an apartment. The exact rental price increase for property type house is 12.19 percent[11] and for other property types 16.18 percent. These results are in line with research of Chen and Xie (2017) into Airbnb rental price determinants in which similarly a higher rental price is found for property type house. The variable room type shows a negative sign on the coefficient. The rental price declines with 28.82 percent when the Airbnb is a shared or private room. This result supports the research of both Abrate &Viglia (2017) and Chen & Xie (2017). Consistent with research of Magno et al. (2018), the number of accommodates and bathrooms have a positive sign on the coefficient. The rental price of an Airbnb accommodation will increase with 14.68 percent, when the number of accommodates increases. Therefore, is an Airbnb accommodation that fits more people, more expensive. When more bathrooms are present, the rental price increases with 12.52 percent. Similar results are presented in research of Lorde et al. (2019) and Chica-Olmo et al. (2020).

The variables host is superhost, host listings count, number of reviews, number of reviews and cancellation policy are significantly different from zero at a 99 percent probability level. Host response time is significantly different from zero at a 95 percent probability level. Variables host response rate, host identity verified and review score value are not significantly different from zero. The coefficient

[11] $(e^{0.108} - 1) * 100$ following (Woolridge, 2013).

Table 2: Regression results

Variables	(1)	(2)	(3)
Distance attractive urban green space (km)		-0.0396***	
		(0.0093)	
Distance attractive green space in categories			
0-250m			0.0669***
			(0.0208)
250-500m			0.0442**
			(0.0181)
500-750m			0.0423**
			(0.0191)
750-1000m			0.0291*
			(0.0172)
1000-1250m			0.0289
			(0.0178)
1250-1500m			0.00355
			(0.0201)
Property type			
House	0.115***	0.121***	0.119***
	(0.0175)	(0.0174)	(0.0175)
Other	0.150***	0.151***	0.151***
	(0.0182)	(0.0181)	(0.0184)
Room type	-0.340***	-0.339***	-0.340***
	(0.0162)	(0.0162)	(0.0163)
Accommodates	0.137***	0.137***	0.138***
	(0.0089)	(0.0090)	(0.00896)
Bathrooms	0.118***	0.116***	0.116***
	(0.0174)	(0.0174)	(0.0174)
Host response time			
Within a few hours	-0.030**	-0.0292**	-0.0291**
	(0.0135)	(0.0135)	(0.0135)
Within an hour	-0.036**	-0.0364**	-0.0362***
	(0.0133)	(0.0133)	(0.0133)
Host response rate	0.0218	0.0220	0.0213
	(0.01638)	(0.0164)	(0.0164)
Host is superhost	0.065***	0.0663***	0.0660***
	(0.0119)	(0.0119)	(0.0120)
Host listings count	0.001077***	0.00106***	0.00107***
	(0.000243)	(0.000246)	(0.000246)
Host identity verified	0.0157	0.0158	0.0160
	(0.0104)	(0.0104)	(0.0104)
Number of reviews	-0.00089***	-0.000896***	-0.000895***
	(0.0000792)	(0.0000799)	(0.0000801)
Review score value	0.00895	0.00103	0.0101
	(0.0070623)	(0.00705)	(0.00707)
Cancellation policy	0.0354***	0.0337***	0.0337***
	(0.00946)	(0.00946)	(0.00946)
Distance to city center (km)	-0.0798***	-0.0805***	-0.0816***
	(0.00729)	(0.0067)	(0.00703)
Constant	4.619***	4.665***	4.606***
	(0.1071)	(0.1033)	(0.107)
Spatial controls	Neighborhood	Neighborhood	Neighborhood
Observations	7,770	7,770	7,770
R-squared	0.4690	0.4700	0.4923

Note: The dependent variable is the log of the rental price of an Airbnb accommodation. Specification (1) includes the baseline model and specification (2) includes the urban green space model. Specification (3) is the urban green space model including distance intervals. The reference categories include property type is apartment; room type is entire home/apartment; host response time is within a day; host response rate is less than 100%; host is not a superhost; host identity is not verified; cancellation policy is less strict; >1500. Specifications (1), (2), (3) include spatial controls at neighborhood level and include spatially clustered standard errors at six-digit zip code level (n=3,797).

*** p<0.01, ** p<0.05, * p<0.1

on whether a host is a superhost is positive. Therefore, is the rental price higher when the host is a superhost. This corresponds with a rental price increase of 6.72 percent. This finding is in line with research of Wang and Nicolau (2017) in which the authors find that a superhost status corresponds with a higher rentel price. The coefficient on the number of accommodations a host manages is similarly positive. Consistent with research of Li et al. (2015), the results show that if the listing count of a host is larger, than the rental price increases with 0.12 percent. This result supports research of Gunter (2018) in which is stated that professional hosts are more likely to obtain a superhost status which results in higher rental prices. The coefficient on cancellation policy is as well positive. A strict cancellation policy is associated with a rental price increase of 3.60 percent. This is in contrast to the results of Chen and Xie (2017) who indicate higher rental prices when a moderate cancellation policy is adapted. Furthermore, in line with research Gibbs et al. (2018), the number of reviews is associated with a small rental price decrease of 0.09 percent. Surprisingly, the coefficient on host response time is negative. This results in a rental price decrease of 2.95 percent, when the host response time is within an hour, and 3.54 percent within a few hours, in relation to a long response time. The finding is contrary to previous studies which have suggested that when the host response time is fast, the rental price will increase (Mauri, et al., 2018; Li, et al., 2015; Lorde, et al., 2019). The variable distance to city center is significantly different from zero at 99 percent probability level. The distance to the city center is negatively related to rental price of Airbnb accommodations. This implies that the rental price of an Airbnb accommodation decreases with distance (in km) from the city center. This associated with 8.32 percent rental price decrease. This is in line with research of Gibbs et al. (2017) and Gunter (2018). From the results it can be concluded that deviations with previous studies are modest. It can therefore be suggested that the baseline specification model produces sensible results for the control variables.

4.2 Urban green space specification results

Table 2, model (2) indicates the results for the main specification. This model specifications examines the effect of urban green space, considering attractive urban green space[12], on Airbnb pricing. Model (2) includes spatial controls at neighborhood level and clustered standard errors at six-digit zip code level and model. The null hypothesis of model specification 2 is the presence of urban green space has no effect on the rental price of Airbnb accommodations. The R-squared of the main specification is 0.470. Meaning, that 47.00 percent variation in rental price of Airbnb accommodations is explained by the variation of the independent variables. The R-squared increased compared to model (1) and therefore the model has a slightly better fit.

[12] This study particularly focusses on attractive urban green space, because when only including general green space data, indicated as fixed, it is assummed that people perceive urban green space the same. This is however not the case (Czembrowski et al., 2016; Daams, et al., 2016). This study therefore elaborates on the notion that higher heterogeneity exists among urban green spaces as differences in infrastructure and diversity in landscape occur (Daams, et al., 2016; Ham et al., 2012).

Distance from attractive urban green space is significantly different from zero at a 99 percent probability level. The sign on the coefficient is negative, this implies that the rental price of Airbnb accommodations declines when located further away (in km) from the attractive urban green space. This associated with 3.88 percent rental price decrease. This effect signifies that tourist's value being nearby attractive urban green space when visiting a city (Chan and Marafa, 2017; Konijnendijk, 2013; Terkenli, et al., 2017; Song and Sim, 2021). The results are further consistent with previous academic literature concerning the effect of urban green space on property prices. In previous academic literature is found that urban green spaces have a significant effect on residential property values, resulting in a higher willingness to pay of residents (Conway, et al., 2010; Daams, et al., 2016; Daams, et al., 2019). The results of this study are therefore in line with those of previous studies, as the findings show that attractive urban green spaces have a significant effect, indicating a price premium, on the rental price of Airbnb accommodations.

For the control variables relating to property characteristics, host characteristics and site-specific characteristics the significance level almost remained the same, only small difference can be observed. Property type, room type, the number of accommodates, the number of bathrooms, host is superhost, host listings count, number of reviews, number of reviews, cancellation policy and distance from city center are significantly different from zero at a 99 percent probability level. Host response time is significantly different from zero at a 95 percent probability level. The variables host response rate, host identity verified and the review score value remained insignificantly different from zero. The coefficients on all variables, with exception of room type, bathrooms, host response time, host is superhost, host listings count and cancellation policy, slightly increased. Suggesting that their influence on the rental price of Airbnb accommodations slightly increased. Furthermore, the sign on the coefficients for all variables remained the same.

In order to further evaluate the effect of urban green space on Airbnb pricing, one alternative model specification was developed. Model (3) table 2, include measures of proximity to attractive urban green space with discrete distance intervals. When including these proximity measures, a clearer and more accurate overview of the effect of urban green space on Airbnb pricing is constructed. The R-squared of this model is 0.4923, suggesting that 49.23 percent of the variation in rental price of Airbnb accommodations is explained by the variation of the independent variables. As expected, the results show that the estimated price effect smoothly decay with distance from attractive urban green space. The finding of distance decay is in line with research of Daams et al. (2019) in which property prices decay with distance from the attractive green space. For Airbnb accommodations located within 0-250m from attractive urban green space the price effect is 6.92 percent. This price effect falls to 4.52 percent for Airbnb accommodations located with 250-500m distance from attractive urban green space. A price effect of 4.32 percent is found for Airbnb accommodations located 500-750m from urban green space and 2.95 percent for a 750-1000m distance. There is no significant price effect for Airbnb accommodations located further than 1 km away from attractive urban green space. This finding is similar to that of Daams et al. (2019) in which the price effect of urban green space on property prices becomes negligible after 1 km. On the contrary, Daams et al. (2016) indicate an estimated price effect for residential properties located up to 7 km from attractive urban green space. A possible explanation could be that tourists find attractive urban green space only valuable when it is located within walking distance, as 1 km can be determined as a feasible walking distance.

4.3 Robustness analysis

The analysis only considers the distance between urban green space and Airbnb accommodations. However, one could suggest that view on urban green space is similarly important (Daams, et al., 2016). Daams et al. (2016) suggest that the measures of proximity could result in inaccurate estimates when view is not controlled for. Therefore, using a chow test it is analyzed whether a difference exists between Airbnb accommodations with a view and no view on attractive urban green space. The result of the chow test, see appendix 10, shows no significant effect ($F=0.56$, $p=0.5721$) indicating a robust model. Based on the chow test, the null hypothesis that the coefficients of Airbnb accommodations with a view and without a view are equal cannot be rejected as the calculated p-value is larger than 0.05. For the regression analysis it is consequently irrelevant if the Airbnb accommodation has a view or no view. A possible explanation could be that tourists, unlike residents, spend fewer time inside their accommodation and therefore attach more importance to accessibility than specifically view on attractive urban green space.

5. DISCUSSION

5.1 The 'touristic value' of attractive urban green space

A growing literature shows the importance of green space in the urban environment. Green spaces provide benefits for the urban area and positively affect the quality of life in the city (Brown, et al., 2014; Chan and Marafa, 2017; Keng Lee, et al., 2015; Leeuwen, et al., 2011; McPherson, 1992). The growing importance of green space in the urban environment is similarly important for tourists (Deng, et al., 2010). Unlike other studies of green space, this study indicates to what extent urban green space is valuable to tourists, by focusing on the price effect of urban green space on Airbnb accommodations. In doing so, the analysis carried out in this study include a proximity measurement for urban green spaces perceived as attractive (Daams et al., 2016; Daams et al., 2019). This study therefore elaborates on the notion that when attractiveness of urban green space is not considered, the distance over which urban green spaces are capitalized is misconceived. For the reason that with general green space it is assumed that people perceive green space with similar characteristics the same. Consequently, when only using general green space, derived from land use data, the possible price effect of urban green space on Airbnb pricing is not captured. Few studies apply green space perceived as attractive in their analysis while it appears to be relevant to include a perception-based green space measurement (Daams et al., 2016). By including attractive urban green space, this study captures as precisely as possible the potential price effect on Airbnb pricing. The findings of this study provide evidence that attractive green space is indeed valuable to tourists whereas tourists' pay higher rental prices for Airbnb accommodations located nearby attractive urban green space. This implies that Airbnb accommodations located nearby attractive green space provide benefits for tourists in a way it becomes valuable to them. It is important to note that the findings of this study, as expected, show that the estimated effect of attractive green space on Airbnb pricing smoothly decays with distance. This finding is consistent with previous residential pricing literature on attractive urban green spaces (Daams et al., 2016; Daams et al., 2019). The estimated price effect becomes negligible after 1 km which indicates that attractive green spaces do not capitalize in the rental price of Airbnb accommodations further than 1 km away. This implies that when the distance from attractive green space becomes too large, an effect on the rental price is not observed. A possible explanation could be that tourists find attractive urban green space only valuable when it is located within feasible walking distance.

5.2 Policy implications

The empirical findings discussed in this paper are relevant to public policy with regard to green spaces in the city. The local government acknowledges the importance of urban green spaces as it is an essential contribution to the health and social well-being of people in the city, but explicitly information about how tourists' value is not present in policy documentation (Gemeente Amsterdam, 2015). Also on larger scale, the European government is encouraging the presence of green spaces in cities by promoting several projects (Green cities.eu, 2018). This study provides evidence that, similar to residents, tourists value being nearby attractive green space when visiting a city. Consequently, the growing importance of urban green space, for both residents and tourists, could result in possible competition for green spaces in the city. To ensure the right balance and distribution of urban green spaces, it is necessary with regard to policy makers that the importance of green spaces in cities is considered in future urban planning.

5.3 Limitations and recommendations for further research

This study also highlights a number of limitations in data availability and research design. Airbnb data used in the analysis was publicly available, this type of data could therefore be incomplete or unreliable. Further concerns could arise in the calculated distance between attractive urban green space and Airbnb accommodations. The exact location of the Airbnb accommodations is anonymized. The actual location will be at approximately 0-150 meters distance from the pointed location of the Airbnb accommodation on the map. Additionally, is the dataset a snapshot of a particular time, it is possible that not all Airbnb accommodations on the market are captured. It is also unknown whether the rental price is the actual paid price of a booked Airbnb accommodation or an indicative rental price of an available Airbnb accommodation.

It is further unknown whether hosts follow the by Airbnb platform provided, price tips and smart pricing. Therefore, a difference between Airbnb accommodations that follow the price directions and Airbnb accommodations that are not could exist, which could influence the given rental price in the dataset. Further limitations occur relating to the research design. Firstly, a lot of observations were lost as the number of observations declined from 19.634 to 7.770 when preparing the data for analysis. Secondly, beside the variables relating urban green space, numerous control variables were included in the analysis. When using control-based research designs, a possibility exists that not all relevant factors are included in the analysis, simply because the data is not available. This is similarly the case in this study, although a number of control variables were included, variables relating to site-specific characteristics are still limited. It is further important to note that attractive urban green space in Amsterdam is located nearby areas with a high amenity level. It is therefore possible that this is partly picked up in the proximity measurements.

Furthermore, is spatial autocorrelation a main concern as this study includes spatial data (Czembrowski et al. 2016). Variables relating to location specifically lead to spatial effects in the hedonic model, because they represent spatial interactions and dependency (Conway, et al., 2008; Wilhelmsson, 2002). By following Daams et al. (2016) spatially clustered standard errors at street-level, six-digit zip codes, were included in the analysis to address spatial autocorrelation. Furthermore, were spatial controls included at neighborhood level. Daams et al. (2019) suggests that after inclusion of spatial controls estimates show the expected pattern of distance decay and show the expected sign. Six-digit zip code level are very restrictive and could absorb the actual price effect of the distance to urban green space (Daams, et al., 2019). This level of spatial control was therefore not included in the analysis. Daams et al. (2019) further suggests that spatial controls lead to more precise estimates, however at a certain point these spatial controls will become over-precise and give an exhaustible representation of the actual effect. Previous studies show that varying scales of spatial controls should be applied (Abbot and Klaiber, 2011; Daams, et al., 2019). However, this was not possible as the data was not available. It is therefore usefull that future studies focus on applying varying scales of spatial controls in this type of study. Further recommondations for future research center around to what extent urban green space have touristic value. This study only captures a single city, it would however be interesting to extent this by considering multiple cities in the analysis. This gives the opportunity to make comparisons between cities in order to find out to what extent differences exist in effect of urban green space among cities.

6. CONCLUSION

The aim of this study was to provide insights on the effect of attractive urban green space and the rental price of Airbnb accommodations in order to find out to what extent attractive urban green space is valuable to tourists in Amsterdam. This study therefore contributes to previous literature on Airbnb pricing and the price effects of urban green space. The hedonic price approach was used to estimate and evaluate the influence of attractive urban green space on Airbnb pricing. By using the spatial program ArcGIS, the distance between attractive urban green space and the location of Airbnb accommodations in Amsterdam was measured.

The findings show that urban green spaces have a significant effect on the rental price of Airbnb accommodations. For attractive urban green space, the rental price of Airbnb accommodations declines with 3.88 percent, when located further away (in km) from attractive urban green space. In order to establish a more precise measurement of the estimated price effect of attractive urban green space on Airbnb pricing, a measure of proximity to attractive urban green space with discrete distance intervals was included. The results show that the estimated price effect decays with distance from attractive urban green space and is negligible for Airbnb accommodations located further than 1 km away. The estimated price effect varies from 6.92 percent for properties located within 0-250m distance to attractive urban green space and falls to 4.52 percent for Airbnb accommodations located in 250-500m distance to attractive urban green space. When an Airbnb accommodation is located within 500-750m to attractive urban green space, the price effect is estimated to be 4.32 percent and a 2.95 percent price effect was found for a 750-1000m distance.

From the findings of this study, it can be concluded that tourists are willing to pay a price premium on the rental price of their overnight stay, when the Airbnb accommodation is located nearby attractive urban green space. It can therefore be suggested that tourists value the presence of nearby attractive urban green space. Since urban green spaces are highly valued by local residents, evidence that tourists similarly value the attractive urban green spaces is beneficial to policy makers. Therefore, should the growing importance of urban green space for tourists be taken into consideration in future policy making.

LITERATURE

Abbot, J. K. & Klaiber, A. (2011). An embarrassment of riches: confronting omitted variable bias and multiscale capitalization in hedonic price models. *The review of Economics and Statistics,* 93(4), 1331-1342.

Abrate, G. & Viglia, G. (2017). Personal or Product Reputation? Optimizing Revenues in the Sharing Economy. *Journal of travel research,* 58(1), 136-148.

Airbnb (2021). *About us.* Retrieved on 10 February, 2021 from https://news.airbnb.com//about-us/

Baur, J. W. R. & Tynon, J. F. (2010). Small-scale urban nature parks: Why should we care?. *Leisure science,* 32(2), 195-200.

Benítez-Aurioles, B. (2017). The role of distance in the peer-to-peer market for tourist accommodation. *Tourism economics,* 24(3), 237-250.

Berry, W. D. (1993). *Understanding regression assumption*s. 1st editon. Newbury Park, CA: SAGE publications.

Brooks , C. & Tsolacos, S. (2010). Real Estate Modelling and Forecasting. 1st editon. Cambridge: Cambridge University Press.

Boer, A. den, (2013). Dynamic Pricing and Learning: Historical Origins, Current Research, and New Directions. *Surveys in Operations Research and Management Science,* 20(1), 1-18.

Brown, G., Schebella, M. F. & Weber, D. (2014). Using participatory GIS to measure physical activity andurban park benefits. *Landscape and Urban Planning,* 121(1), 32-44.

Cai, Y., Zhou, Y., Ma, J. & Scott, N. (2019). Price Determinants of Airbnb Listings: Evidence from Hong Kong. *Tourism Analysis,* 24(1), 227-242.

CBRE (2020). *The role of green space in cities.* Amsterdam: CBRE.

CBS (2008). *Bestand Bodemgebruik.* Voorburg: Centraal Bureau voor de Statistiek.

CBS (2020). *Quality statement.* Retrieved on 10 August, 2020 from https://www.cbs.nl/en-gb/about-us/organisation/quality

Chan, C. S. & Marafa, L. (2017). How a green city brand determines the willingness to stay in a city: the case of Hong Kong. *Journal of Travel and Tourism Marketing,* 34(6), 719-731.

Chattopadhyaya, M. & Kumar Mitra, S. (2019). Do airbnb host listing attributes influence room pricing homogenously. *International Journal of Hospitality Management,* 81(19), 54-64.

Chen , C. & Rothschild, R. (2010). An application of hedonic pricing analysis to the case of hotel rooms in Taipei. *Tourism Economics,* 16(3), 685-694..

Chen, Y. & Xie, K. (2017). Consumer valuation of Airbnb listings: a hedonic pricing approach. *International Journal of Contemporary Hospitality Management,* 29(9), 2405-2424.

Chica-Olmo, J., González-Morales, J. G. & Zafra-Gómez, J. L. (2020). Effects of location on Airbnb apartment pricing in Málaga. *Tourism Management,* 77(1), 1-13.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica,* 28(3), 591-605.

Conway, D., Li, C. & Wolch, J. (2008). A Spatial Autocorrelation Approach for Examining the Effects of Urban Greenspace. *Real Estate Finance Economics,* 41(1), 150-169.

Czembrowski, P., Kronenberg, J. & Czepkiewicz, M. (2016). Integrating non-monetary and monetary valuation methods – SoftGIS and hedonic pricing. *Ecological Economics,* 130(1), 166-175.

Czembrowski, P., Laskiewicz, E., Kronenberg, J., Engström, G & Andersson, E. (2019). Valuing individual characteristics and the multifunctionality of urban green spaces: The integration of sociotope mapping and hedonic pricing. *PLoS one,* 14(3), 1-16

Daams, M. N., Sijtsma, F. J. & Veneri, P. (2019). Mixed monetary and non-monetary valuation of attractive urban green space: A case study using Amsterdam house prices. *Ecological Economics,* 166(1), 1-12.

Daams, M., Sijtsma, F. J. & Vlist, J. van der, (2016). The Effect of Natural Space on Nearby Property Prices: Accounting for Perceived Attractiveness. *Land Economics,* 92(3), 389-410.

Deng, J., Arano, K., Pierskalla, C. & McNeel, J. (2010). Linking Urban Forests and Urban Tourism: A Case of Savannah, Georgia. *Tourism Analysis,* 30(2), 167-181.

Dodds, R. & Joppe, M., (2001). Promoting urban green tourism: The development of the other map of Toronto. *Green Tourism Association,* 7(3), 261-267.

Dogru, T. & Pekin, O. (2017). What do guests value most in Airbnb accommodations? An application of the hedonic pricing approach. *Boston Hospitality Review,* 5(2), 1-13.

Ert, E., Fleischer, A. & Magen, N. (2016). Trust and reputation in the sharing economy: The role of personal photos in Airbnb. *Tourism management,* Volume 55(1), 62-73.

Gangodagamage, C., Zhou, X. & Lin, H. (2008). Spatial autocorrelation, in: Shekhar, S., Xiong, H. & Zhou, X. *Encyclopedia of GIS* (pp. 32-37). Boston: Springer.

Gemeente Amsterdam, Onderzoek, Informatie en Statistiek (2016). *Basisbestand Gebieden Amsterdam (BBGA).* Retrieved on 15 January, 2021 from https://data.amsterdam.nl/datasets/G5JpqNbhweXZSw/basisbestand-gebieden-amsterdam-bbga/

Gemeente Amsterdam - Ruimte en Duurzaamheid (2017). *Explanation land use map.* Retrieved on 1 April, 2020 https://maps.amsterdam.nl/grondgebruik/?LANG=en

Gemeente Amsterdam - Ruimte en Duurzaamheid (2017). *Land use 2017.* Retrieved on 1 April, 2020 from https://maps.amsterdam.nl/open_geodata/

Gemeente Amsterdam (2015). *Agenda Groen.* Amsterdam: Ruimte en Duurzaamheid van de Gemeente Amsterdam.

Gemeente Amsterdam (2020). *Groenvisie 2020 - 2050.* Amsterdam: Gemeente Amsterdam.

Gemeente Amsterdam (2020). *Stadsdelen.* Retrieved on 20 November, 2020 from https://www.amsterdam.nl/bestuur-organisatie/organisatie/stadsdelen/

Gibbons, S., Mourato, S. & Resende, G. M. (2014). The Amenity Value of English Nature: A Hedonic Price Approach. *Environment Resource Economics,* 57(1), 175-196.

Gibbs, C., GretzeL, U., Guttentag, D. & Yao, L. (2017). Use of dynamic pricing strategies by Airbnb hosts. *Journal of contemporary hospitality management,* 30(1), 2-20.

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. & Goodwill, A. (2018) Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings. *Journal of travel and tourism marketing,* 35(1), 46-56.

Green cities.eu (2018). *Green Cities for a Sustainable Europe.* Retrieved on 12 January, 2021 from https://nl.thegreencity.eu

Greenmapper (2020). *Greenmapper data.* Retrieved on 20 April, 2020 from http://data.greenmapper.org

Griffith, D. A. (1992). What is Spatial Autocorrelation? Reflections on the Past 25 Years of Spatial Statistics. *L'Espace géographique ,* 21(3), 265-280.

Gunter, U. (2018). What makes an Airbnb host a superhost? Empirical evidence from San Francisco and the Bay Area. *Tourism Management,* 66(1), 26-37.

Gutiérrez, J., García-Palomares, J. C., Romanillos, G. & Salas-Olmedo, M. H. (2017). The eruption of Airbnb in tourist cities: Comparing spatial patterns of hotels and peer-to-peer accommodation in Barcelona. *Tourism management,* 62(1), 278-291.

Gutt, D. & Hermann, P. (2015). *Sharing means caring? Hosts' price reaction to rating visibility.* Münster: ECIS 2015 Research-in-Progress Papers.

Ham, C., Champ, P. A., Loomis, J. B. & Reich, R. M. (2012). Accounting for Heterogeneity of Public Lands in Hedonic Property Models. *Land Economics,* 88(3), 444-456.

Het Parool (2020). Amsterdam wil radicaal vergroenen: nieuw stadsbos, grotere parken. *Het Parool*,

13-05-2020
Inside Airbnb (2019). *Get the data.* Retrieved on 1 december, 2019 from http://insideairbnb.com/get-the-data.html
Inside Airbnb (2020). *About Inside Airbnb.* Retrieved on 1 May, 2020 from http://insideairbnb.com/about.html
Kalliopi, S. (2015). Does "green" matter? Exploring the effects of Green City Branding on the intention to visit a destination. *University of Amsterdam. 1-45*
Keng Lee, A. C., Jordan, H. C. & Horsley, J. (2015). Value of urban green spaces in promoting healthy living and wellbeing: prospects for planning. *Risk Management Health Policy,* 8(1), 131-137.
Kerr, N. (2017). Picking a Travel Spot? Go for the Green. *AARP*, 20-09-21
Konijnendijk, C., 2013. Benefits of Urban Parks a Systematic Review. *Urban Forestry and Urban Greening,* 12(1), 1-44
Leeuwen, E. van, Nijkamp, P. & Noronha Vaz, T. (2011). The multifunctional use of urban greenspace. *International Journal of Agricultural Sustainability,* 8(1), 20-25.
Li, J., Moreno, A. & Zhang, D. (2015). Agent Behavior in the Sharing Economy: Evidence from Airbnb. *Manufacturing & Service Operations Management,* 1298(1), 1-30.
Lin, F. J. (2008). Solving Multicollinearity in the Process of Fitting Regression Model Using the Nested Estimate Procedure. *Quality & Quantity,* 42(1), 417-426.
Lorde, T., Jacob, J. & Weekes, Q. (2019). Price-setting behavior in a tourism sharing economy accommodation market: A hedonic price analysis of AirBnB hosts in the caribbean. *Tourism Management Perspectives,* 30(19), 251-261.
Magno, F., Cassia, F. & Ugolini, M. (2018). Accommodation prices on Airbnb: effects of host experience and market demand. *TQM Journal,* 30(5), 608-620.
Mauri, A., Minazzi, R., Nieto-García, M. & Viglia, G. (2018). "Humanize your business: the role of personal reputation in the sharing economy. *International Journal of Hospitality Management,* 73(1), 36-43.
McPherson, G. E. (1992). Accounting for benefits and costs of urban greenspace. *Landscape and Urban Planning,* 22(1), 41-51.
OECD (2013). *Handbook on Residential Property Prices Indices (RPPIs)*. 2013 Edition. Luxembourg: Publications Office of the European Union
Pdok (2020). *Dataset: CBS Bestand Bodemgebruik.* Retrieved on 4 august, 2020 from https://www.pdok.nl/introductie/-/article/cbs-bestand-bodemgebruik
Perez-Sanchez, P., Serrano-Estrada , L., Marti, P. & Mora-Garcia , R. T. (2018). The What, Where, and Why of Airbnb Price Determinants. *Sustainability,* 10(12), 4596-4626.
Qiu, R., Xuefeng Fan, D. & Liu, A. (2018). Exploring the Booking Determinants of the Airbnb Properties: An Example of the Listings of London. *Information and communication technologies in tourism,* 31(12), 44-51.
Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy,* 55(1), 34-55.
Royston, P., 2020. *swilk — Shapiro–Wilk and Shapiro–Francia tests for normality.* Retrieved on 1 September, 2020 from https://www.stata.com/manuals/rswilk.pdf
Sheppard, Stephen (1999). Hedonic analysis of housing markets, in: P. C. Cheshire & E. S. Mills, *Handbook of Regional and Urban Economic*s (pp. 1595-1635), Amsterdam: Elsevier.
Sheppard, S. & Udell, A. (2016). Do Airbnb properties affect house prices. *Department of Economics Working Papers*, 1-10.
Song, H. & Sim, C. (2021). Comparing resident and tourist perceptions of an urban park: a latent profile analysis of perceived place value. *Journal of Sustainable Tourism,* 5(1), 1-13.

StataCorp (2019). *Regress postestimation diagnostic plots.* College Station,TX: StataCorp LLC.
Statistics Netherlands (2015). *CBS bestand bodemgebruik 2015.* Retrieved on 1 April, 2020 from https://www.pdok.nl/imdtroductie/-/article/cbs-bestand-bodemgebruik

Tajima, K. (2003). New Estimates of the Demand for Urban Green Space: Implications for Valuing the Environmental Benefits of Boston's Big Dig Project. *Journal of Urban Affairs,* 25(5), 1-20.

Terkenli, T. S., Bell, S., Toskovic, O., Dubljevic-Tomicevic, J., Panagopoulos, T., Straupe, I., Kristianova, K., Straigyte, L., O'Brien, L. & Zivojinovíc, I. (2020.) Tourist perceptions and uses of urban green infrastructure: An exploratory cross-cultural investigation. *Urban Forestry & Urban Greening,* 49(1), 1-12

Terkenli, T. S., Bell, S., Tomicevic, J., & Zivojinovíc, I. (2017). Recreational use of Urban Green infrastructure: the tourist's perspective. *Urban forest,* 7(1), 191-216.

Tivadar, M. & Jayet, H. (2006). *Endougenous amenities and spatial structure of cities.* Lille: International Conference on Management Science and Engineering.

Wang, D. & Nicolau, J. L.(2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management,* 62(17), 120-130.

Wilhelmsson, M. (2002). Spatial Models in Real Estate Economics. *Housing, Theory and Society,* 19(2), 92-101.

Woolridge, J. M. (2013). *Introductory econometrics: a modern approach*. 1st edition. Boston, MA: South-Western Cengage Learning.

Ye, P., Qian, J., Chen, J., Wu, C., Zhou, Y., Mars, S., Yang, F. & Zhang, L. (2018). Customized Regression Model for Airbnb Dynamic Pricing. *Applied Data Science Track Paper,* 1(1), 19-23.

Zhang, Z., Chen, R., Han, L. & Yang, L. (2017). Key factors affecting the price of Airbnb listings: a geographically weighted approach. *Sustainability,* 9(9), 12-24.

APPENDICES

Appendix 1: Conceptual model

This study is to provide insights on the effect of urban green space on Airbnb pricing in order to find out to what extent urban green space is valuable to tourists. Based on the theoretical framework a conceptual model (figure 1) was established. The conceptual model indicates to what extent urban green space, the independent variable, has an effect on the rental price of Airbnb accommodations, the dependent variable. Besides this possible effect, the literature shows that the rental price of Airbnb accommodations is influenced by property, host and site-specific characteristics. These are included in this research as control variables. Previous academic research further suggests that in Airbnb pricing limitations exists in price construction and demand estimation which is important to take into account in analysis.

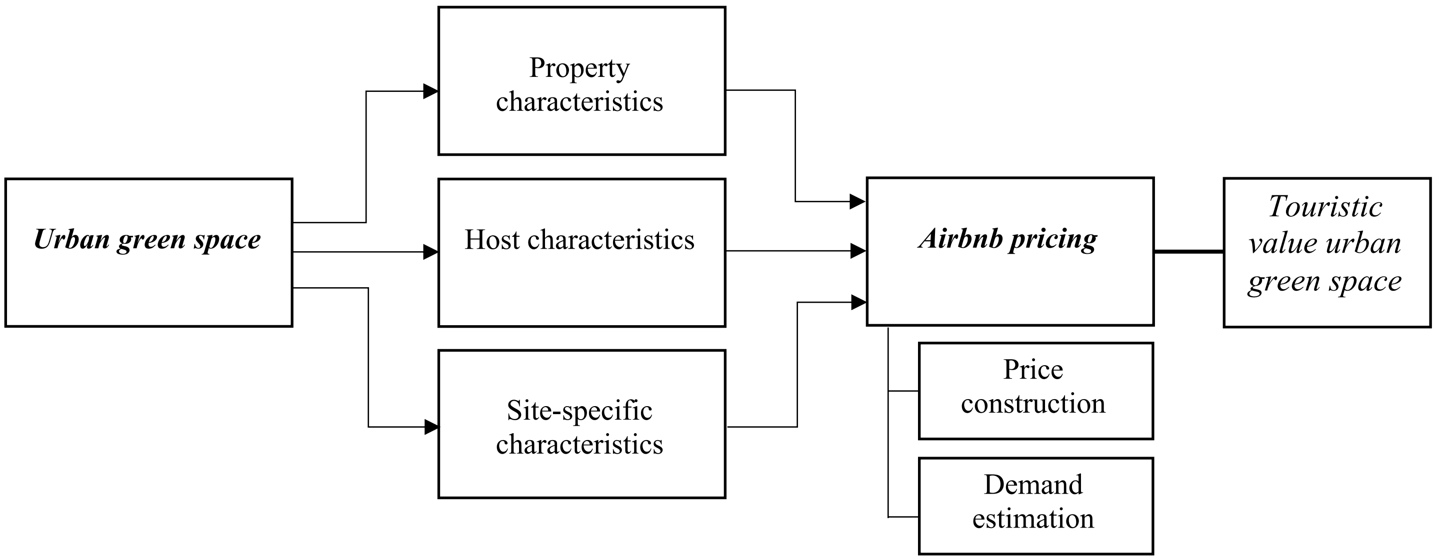

Figure 1: conceptual model (source: own work)

Appendix 2: Definition of variables

Table 1: Overview independent, dependent and control variables

Name	Category	Label	Type	Definition	Source
Ln(price)	Dependent variable	log_price	Continuous	Log rental price of an Airbnb accommodation	(Inside Airbnb, 2019)
Distance attractive urban green space	Independent variable	distance_cluster_areas_KM	Categorical	Distance Airbnb accommodation from attractive green space in km	(Daams, et al., 2016) (Inside Airbnb, 2019)
Distance attractive urban green space category	Independent variable	i.attractive_greenspace_250_1500	Categorical	Distance attractive green space in categories (0=>1500m, 1=0-250m, 2=250-500m, 3 =500-750m, 4=750-1000m, 5=1000-1250m, 6=1250-1500m)	(Inside Airbnb, 2019) (Daams, et al., 2016)
Property type	Control variable	property_type_category2	Categorical	Property type of an Airbnb accommodation (1 = apartment, 2 = house 3 = other)	(Inside Airbnb, 2019)
Room type	Control variable	ROOM_TYPE	Categorical	Room type of an Airbnb accommodation (0 = entire home/apartment, 1= shared/private room)	(Inside Airbnb, 2019)
Accommodates	Control variable	accommodates	Continuous	The number of accommodates	(Inside Airbnb, 2019)
Bathrooms	Control variable	bathrooms	Continuous	The number of bathrooms	(Inside Airbnb, 2019)
Host response time	Control variable	HOST_RESPONSE_TIME	Categorical	The host's response speed (1 = within a day 2=within a few hours 3= within an hour)	(Inside Airbnb, 2019)
Host response rate	Control variable	HOST_RESPONSE_RATE	Categorical	The host response rate (0 = less than 100%, 1 = 100%)	(Inside Airbnb, 2019)
Host is superhost	Control variable	HOST_IS_SUPERHOST	Categorical	Whether a host is a superhost (0 = no, 1= yes,)	(Inside Airbnb, 2019)
Host listings count	Control variable	host_listings_count	Continuous	Host's number of accommodations listed on Airbnb	(Inside Airbnb, 2019)
Host identity verified	Control variable	HOST_IDENTITY_VERIFIED	Categorical	Whether the host's identity is verified (0 = no, 1 = yes)	(Inside Airbnb, 2019)
Number of reviews	Control variable	number_of_reviews	Continuous	The number of reviews of an Airbnb accommodation	(Inside Airbnb, 2019)
Review scores value	Control variable	review_scores_value	Continuous	Average review score of an Airbnb accommodation	(Inside Airbnb, 2019)

Cancellation policy	Control variable	cancellation_policy_category2	Categorical	Strictness of cancellation policy (0= less strict, 1 = strict)	(Inside Airbnb, 2019)
Distance city center	Control variable	distance_city_center_KM	Continuous	Distance Airbnb accommodation From dam square in KM	(Inside Airbnb, 2019)
Tourist activities	Spatial control variable	BHVEST_TOER	Categorical	Spatial controls concerning tourist activities in Amsterdam (number of hotel and catering industry, cultural and recreational activities)	(Inside Airbnb, 2019) (Gemeente Amsterdam, Onderzoek, Informatie en Statistiek, 2016) (BBGA)
View on attractive green space	Chow test	attractive_greenspace_view	Categorical	View or no view on attractive urban green space (0=100m/max, 1=0/100m)	(Inside Airbnb, 2019)

Appendix 3: Data preparation

This section explains the operationalization of the variables and the construction of a dataset which is sufficient for analysis. The exact steps are described in the DO-file (appendix 11). In the operationalization of the variables are the dependent variable and the independent variables and control variables analyzed. In this study is the dependent variable is explained the independent variable and a set of control variables which are based on the theoretical framework of this study. The dependent variable is the rental price of Airbnb accommodations. This study includes one independent variable which is distance from Airbnb accommodation to attractive urban green space. The control variables are divided in three categories which are property characteristics, host characteristics and site-specific characteristics. The category property characteristics consists of the following control variables: property type, room type, number of accommodates and number of bathrooms. The category host characteristics exist of control variables host response time, host response rate, host is superhost, host listings count, host identity verified, number of reviews, review scores value and cancellation policy. The site-specific characteristics consists of one control variable the distance city center.

The distribution of the dependent variable is examined through a histogram, boxplot, QQ-plot and summary statistics (appendix 4). The dependent variable is not symmetric around its mean value. Therefore, a logarithmic transformation is necessary. A logarithmic transformation helps the skewed distribution to come closer to a normal distribution. A logarithmic transformation has been applied in previous academic research into rental price determinants of Airbnb accommodations (Cai, et al., 2019; Chen and Xie, 2017; Gibbs, et al., 2017; Magno, et al., 2018). Magno et al. (2018) argue that a logarithmic transformation results in a more straightforward interpretation of the OLS regression. Cai et al. (2018) similarly uses log transformation of the rental price into their study into price determinants of Airbnb accommodations. The researchers use the logarithm of the rental price to estimate semi elasticity of rental price in relation to the independent variables. After the logarithmic transformation of the rental price, a boxplot and QQ-plot indicated a few outliers. Therefore, all values under 2.5 were removed from the dataset.

The independent variable distance attractive green space was similarly investigated through a histogram, boxplot, QQ-plot and summary statistics. To ensure a normal distribution an additional numerical normality test is carried out, the skewness and kurtosis test for normality. Because of the large sample size, this numerical normality test is chosen instead of the Shapiro Wilkinson (Royston, 2020). The graphical normality tests (appendix 5) show a non-normal distribution. The same conclusion can be drawn from the numerical normality test (appendix 5). Therefore, is the variable corrected for extreme values. After this correction the graphical normality tests and skewness and kurtosis test is carried out again. Unfortunately, it did not help to bring the distribution closer to a normal distribution. Then the variable was transformed into a logarithmic variable. This did not totally bring the distributions closer to a normal distribution. When considering previous literature, the distance between Airbnb accommodations and urban green space is not studied before. However, the distance between the city center and Airbnb accommodations are indeed investigated in which the distance to the city center is considered a continuous variable (Cai, et al., 2018; Chica-Olmo, et al. (2020). Gibbs, et al., 2017; Wang and Nicolau, 2017; Qiu, et al., 2018; Zhang, et al., 2017).

The control variables consist of categorical variables and continuous variables. The categories are based on previous academic literature regarding Airbnb rental price determinants. The continuous variables were analyzed using a histogram, boxplot and summary statistics. The variable property type is a categorical variable and consists of a large number of categories. Categories that are not considered as property were removed from the dataset. These were barns, campers, campsite, casa particular, dome house, farm stay, tent and yurt. This downsized the number of categories but there is still a large number of categories left. Therefore, the existing categories were combined based on existing literature regarding rental price determinants of Airbnb accommodations. In the paper of Chen and Xie (2017) three property type categories are proposed 1"apartment" 2"house" 3"other". These categories are proposed in this study. The categorical variable room type contained 4 categories with uneven groups. The variable is re-categorized in order to make the groupsize more even among the categories. Following Magno et al. (2018) a distinction is made between an entire home or appartment and a private room. The original categories of the variable host response time were definend as 1“a few days or more”, 2”within a day”, 3”within a few hours” and 4”within an hour”. Both Lorde et al. (2019) and Chen and Xie (2017) use this type of categorization. However, in this study is the groupsize of these categories uneven. Following Mauri et al. (2018) the categories were defined as 1” within a day” 2 ”within” a few hours” 3 ”within an hour”. This categorization more even and therefore chosen for the statistical analysis. The categories for cancellation policy are 0 "less strict" 1 "strict" in which less strict is a formation of flexible and moderate cancellation policies. These categories are proposed in the study of Cai et al. (2018). Furthermore, the dataset includes two dummy variables. These are host is superhost and host identity is verified. These dummy variables are recoded and labelled as 0 "no" 1 "yes".

The remaining control variables that are contineous are analyzed using a histogram and boxplot to examine the distribution of the variables. The variables accommodates, bathrooms, host listings count, number of reviews and review score value are not normally distributed. In research of Cai et al. (2018) and Wang and Nicolau (2017) in which the same Airbnb data is used, are these variables not log transformed or categorized but considered as contineous variables. Therefore, no transformations of these variables are made.

In order to investigate whether a linear relationship exists between the dependent variable and the independent variable, a two-way scatterplot has been made between the dependent variable, log rental price Airbnb accommodations, and the independent variable, distance attractive green space. An overview is given in appendix 6. A correlation matrix is used to investigate potential multicollinearity. The correlation matrix (appendix 7) gives an overview of the linear relationship between the variables given by the correlation coefficient (Brooks and Tsolacos, 2010).

Appendix 4: Operationalization dependent variable

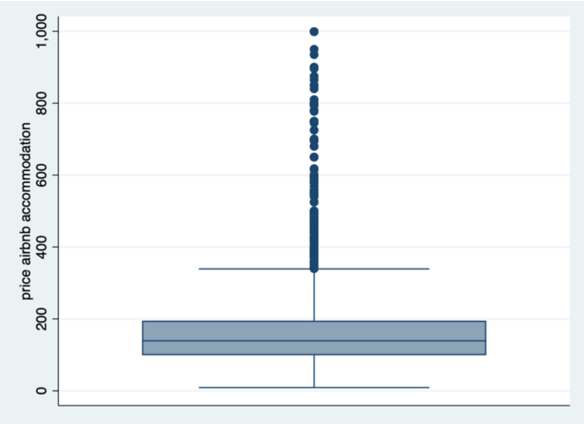

Figure 1: Boxplot rental price

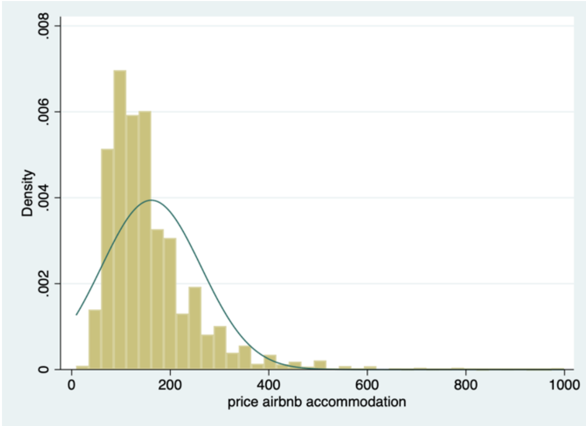

Figure 2: Histogram rental price

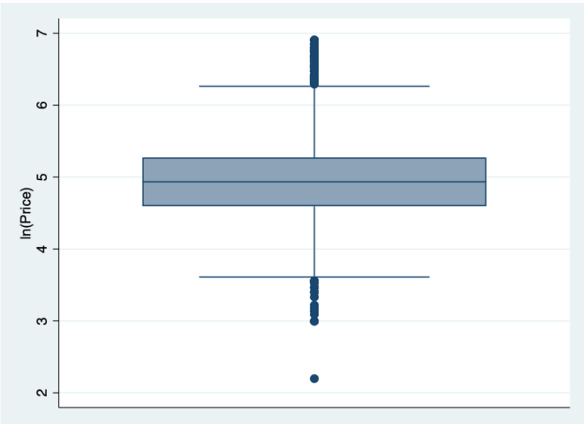

Figure 3: Boxplot log rental price

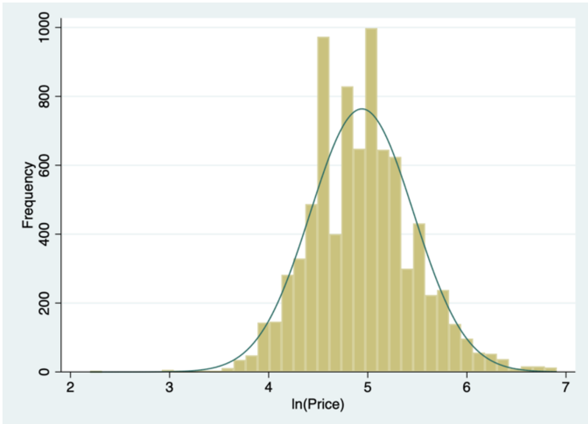

Figure 4: Histogram log rental price

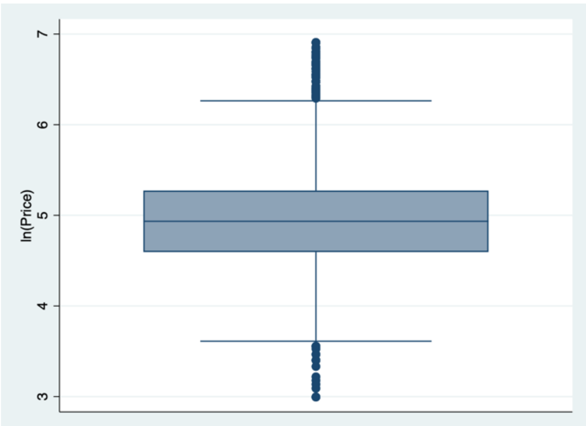

Figure 5: Boxplot log rental price removed outlier

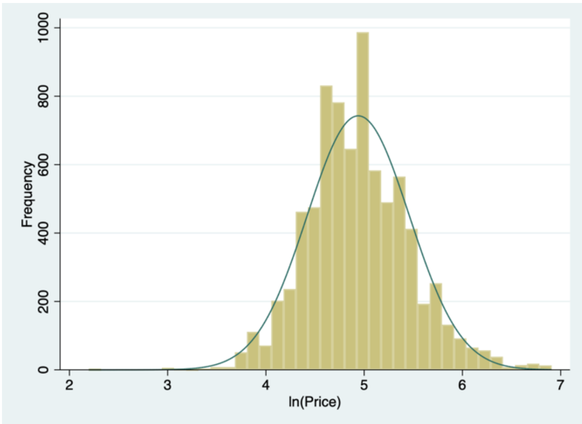

Figure 6: Histogram log rental price removed outlier

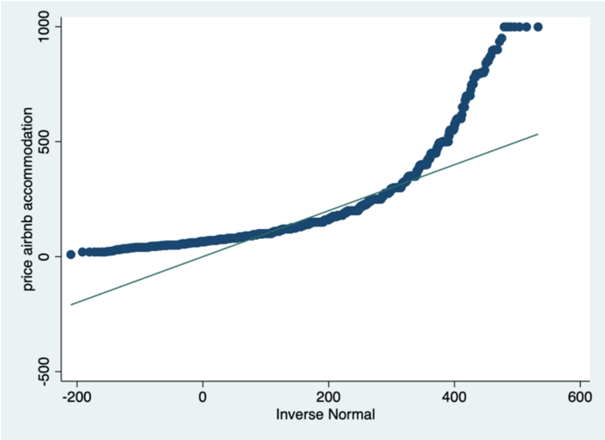

Figure 7: QQ-plot rental price

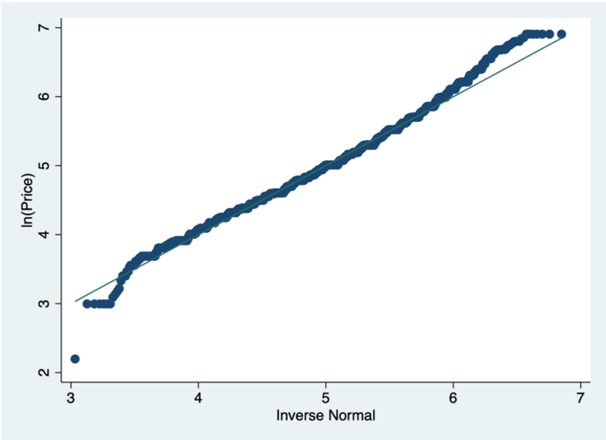

Figure 8: QQ-plot log rental price

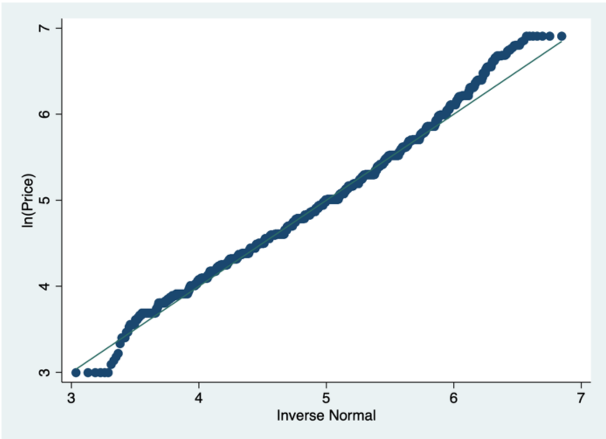

Figure 9: QQ-plot log rental price removed outlier

Appendix 5: Operationalization independent variable

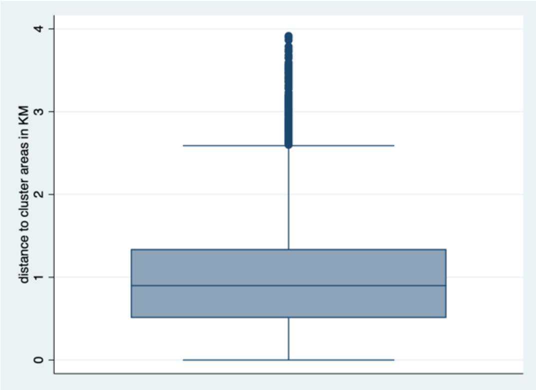

Figure 10: Boxplot distance attractive urban green space

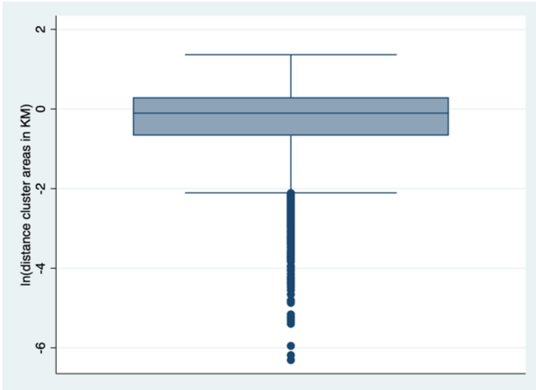

Figure 11: Boxplot log distance attractive urban green space

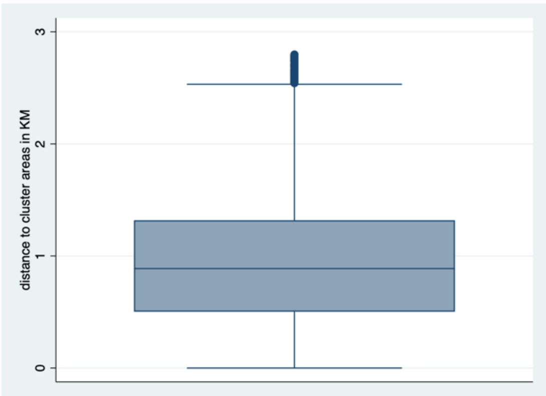

Figure 12: Boxplot log distance attractive urban green space removed outlier

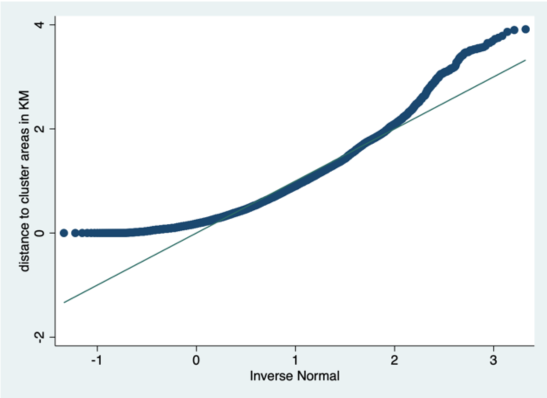

Figure 13: QQ-plot distance attractive urban green space

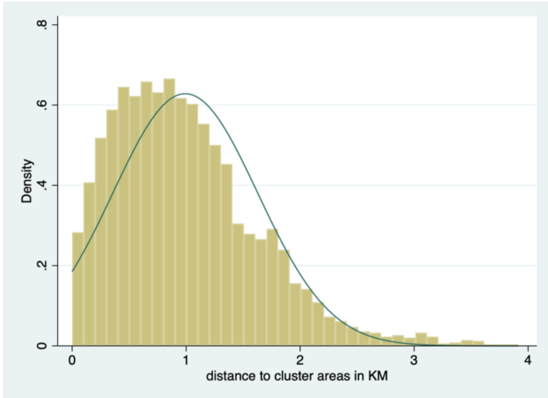

Figure 14: Histogram distance attractive urban green space

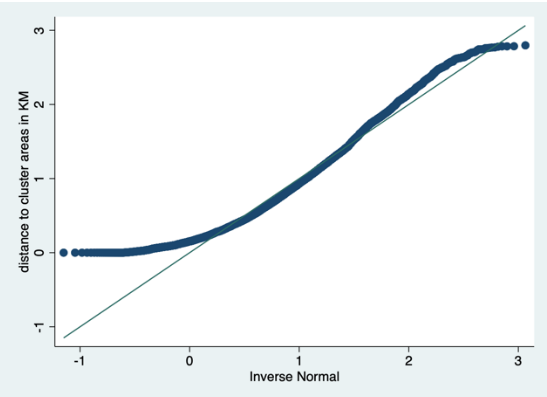

Figure 15: QQ-plot distance attractive urban green space removed outlier

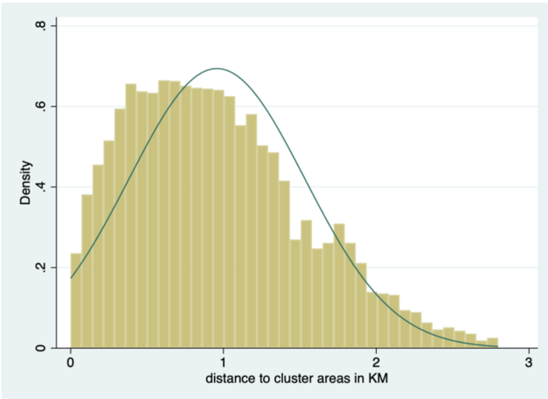

Figure 16: Histogram distance attractive urban green space removed outlier

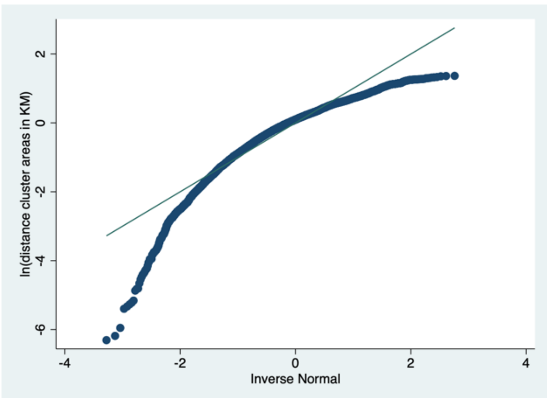

Figure 17: QQ-plot log distance attractive urban green space

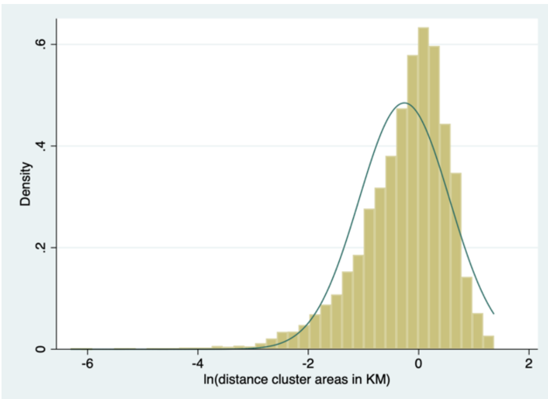

Figure 18: Histogram log distance attractive urban green space

Table 2: Skewness and kurtosis test distance from attractive urban green space

Skewness and kurtosis tests for normality					
				Joint test	
Variable	Obs	Pr(skewness)	Pr(kurtosis)	Adj chi2 (2)	Prob>chi2
Distance attractive urban green space	8,244	0.0000	0.0000	940.08	.
Distance attractive urban green space after deleting extreme values	8,113	0.0000	0.0006	352.47	0.0000
Logarithm Distance attractive urban green space	8,244	0.0000	0.0000	940.08	0.0000

Appendix 6: Two-way scatterplot

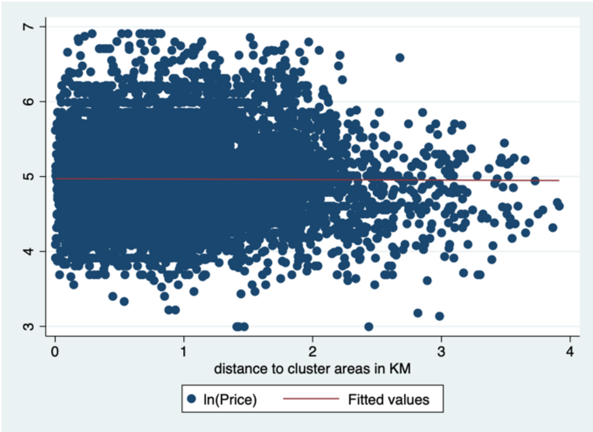

Figure 19: Two-way scatterplot distance attractive urban green space

Appendix 7: Correlation matrix

Table 3 gives an overview of the correlation matrix. Relatively low positive correlation can be seen between the variables room type and property type (0.332), the number of bathrooms and the number of accommodates (0.387), number of reviews and room type (0.378) and the number of reviews and host is superhost (0.300).

Table 3: Matrix of correlations

Variables	(1)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
(1) distance attractive green space	1.000														
(3) category Attractive green space	-0.169	1.000													
(4) Property type	0.143	-0.061	1.000												
(5) Room type	0.062	-0.002	0.332	1.000											
(6) Accommodates	0.051	-0.069	0.129	-0.241	1.000										
(7) Bathrooms	-0.001	-0.014	0.091	-0.076	0.387	1.000									
(8) Host response time	0.031	0.005	0.124	0.240	-0.041	-0.002	1.000								
(9) Host response rate	-0.008	-0.006	-0.005	-0.030	-0.005	0.009	0.283	1.000							
(10) Host is superhost	0.040	-0.004	0.160	0.232	-0.034	0.003	0.181	0.141	1.000						
(11) Host listings count	-0.008	0.008	-0.009	-0.016	0.038	0.046	0.086	0.014	-0.041	1.000					
(12) Host identity verified	-0.007	-0.007	-0.028	-0.049	0.053	0.035	-0.085	0.001	0.058	-0.061	1.000				
(13) Number of reviews	0.031	0.004	0.169	0.377	-0.083	-0.067	0.168	0.011	0.300	-0.036	0.124	1.000			
(14) Review score value	0.009	0.009	0.007	-0.031	-0.065	0.034	-0.051	0.097	0.151	-0.128	0.066	0.031	1.000		
(15) Cancellation policy	-0.036	0.021	0.012	-0.033	0.131	0.081	0.042	0.024	0.014	0.054	0.047	0.001	-0.044	1.000	
(16) Distance city center	-0.034	-0.164	0.052	0.089	0.050	0.035	-0.006	0.038	-0.012	-0.030	0.010	-0.064	0.010	-0.061	1.000

Appendix 8: VIF values

Table 4: Variance inflation factor

	VIF	1/VIF
Distance attractive urban green space	1.04	.963
Property type		
2	1.15	.870
3	1.23	.810
Room type	1.50	.665
Accommodates	1.33	.750
Bathrooms	1.21	.823
Host response time		
2	1.63	.614
3	1.87	.534
Host response rate	1.14	.880
Host is superhost	1.19	.839
Host listings count	1.04	.965
Host identity is verified	1.05	.950
Number of reviews	1.29	.776
Review score value	1.07	.931
Cancellation policy	1.03	.966
Distance city center	1.07	.932
Mean VIF	1.24	.

Table 5: Variance inflation factor

	VIF	1/VIF
attractive green space 0-250 m	1.361	.735
attractive green space 250-500m	1.548	.646
attractive green space 500-750m	1.555	.643
attractive green space 750-1000m	1.566	.639
attractive green space 1000-1250m	1.510	.662
Attractive green space 1250-1500m	1.403	.713
Property type		
2	1.141	.876
3	1.141	.876
Room type	1.507	.664
Accommodates	1.336	.749
Bathrooms	1.216	.822
Host response time		
2	1.629	.614
3	1.872	.534
Host response rate	1.137	.880
Host is superhost	1.193	.838
Host listings count	1.038	.964
Host identity is verified	1.053	.949
Number of reviews	1.288	.776
Review score value	1.074	.931
Cancellation policy	1.037	.965
Distance city center	1.115	.897
Mean VIF	1.325	.

Appendix 9: Testing OLS assumptions

OLS assumptions	
Average value of the errors is zero	$E(\mu t) = 0$
The variance of the errors is constant and finite over all values	$Var(\mu t) = \sigma^2 < \infty$
The errors are statistically independent of one another	$Cov(\mu i, \mu j) = 0$
X-variables are not correlated with the error term of the estimated equation	$Cov (yt, xt) = 0$
The errors are normally distributed	$\mu t \sim N(0, \sigma^2)$

Figure 20: OLS assumptions (Brooks and Tscolacos, 2015)

When the OLS assumptions are met, the OLS estimators have desirable properties and statistical inferences can be made (Berry, 1993). Brooks and Tsolacos (2015) define this type of estimator as best linear unbiased estimator (BLUE). If the OLS assumptions are met, the properties of the OLS estimator are consistent, unbiased and efficient. The first assumption is linearity meaning that the average value of the errors is zero. If a constant term is included in the regression equation, this assumption cannot be violated (Brooks and Tscolacos, 2015). Assumption 2 refers to homoskedasticity and can be explained as the situation in which the variance of the errors is constant. There is a possibility that the errors are not constant and are said to be heteroskedastic. As result, is the OLS estimator still unbiased and consistent but no longer BLUE (Brooks and Tscolacos, 2015). This assumption is tested using both graphical and statistical tests. The rvplot shows a residual-versus-fitted plot in which the residuals are graphed against the fitted values. In a well-fitted model, there is no pattern present (StataCorp, 2019)

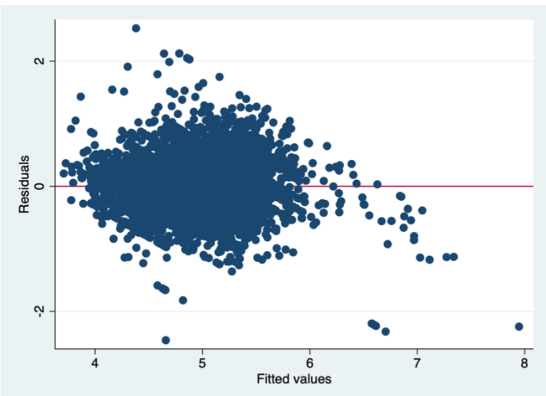

Figure 21: Residuals-versus-fitted plot

In order to extent the analysis into homoscedasticity two additional statistical tests are conducted. These are the Breusch-Pagan test and White's test. The H0 of the Breusch-Pagan test is the variance of the errors is constant. This means that the errors are homoskedastic. H0 is rejected when the p-value < 0.05. The results show a p-value of 0.0000 and this is below 0.05. Therefore, is heteroskedasticity assumed and H0 rejected.

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
Ho: Constant variance
Variables: fitted values of log price
chi2(1) = 280.41
Prob > chi2 = 0.0000

Figure 22: Breusch-Pagan test

The White's test for homoskedasticity shows the following results. The p-value is 0.000. The H0 of the White's test, assuming homoskedasticity, can therefore be rejected because the p-value < 0.05. The results of both the graphical and statistical tests for homoskedasticity show that the second OLS assumption cannot be met. Therefore, is the robust option included in the regression analysis.

White's test for Ho: homoskedasticity
against Ha: unrestricted heteroskedasticity

chi2(198) = 1941.99

Prob > chi2 = 0.0000

Cameron & Trivedi's decomposition of IM-test

Source	chi2	df	p
Heteroskedasticity	1941.990	198	0.000
Skewness	584.290	19	0.000
Kurtosis	30.880	1	0.000
Total	2557.160	218	0.000

Figure 23: White's test for homoskedasticity

Assumption 3 concerns autocorrelation and within this particular study it concerns spatial autocorrelation. As spatial autocorrelation could occur in house price data because when the price of a property is set, selling prices of nearby properties are considered (Daams, et al., 2019). This could similarly be considered with Airbnb data on rental prices of Airbnb accommodations. Therefore, are clustered standard errors at street-level, six-digit zip codes, included in the regression analysis. Assumption 4 considers independence of the error term. This means that the error term is not correlated with the independent variables. Berry (1993) defines the error term as the differences between the predicted and observed values of the dependent variable. Based on the OLS coefficient estimates. When the situation occurs where the independent variables are correlated with the error term, the error term will take on a high value and *y*, the dependent variable, will similarly be high. As result, biased and inconsistent parameter estimates appear (Brooks and Tscolacos, 2015).

Assumption 5 refers to the normality of the errors. To check whether the errors are normally distributed a Kernel density plot, standardized normal probability plot and a histogram of the residuals are graphed.

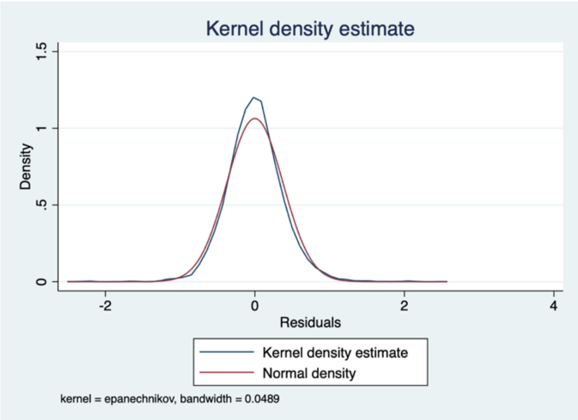

Figure 24: Kernel density plot

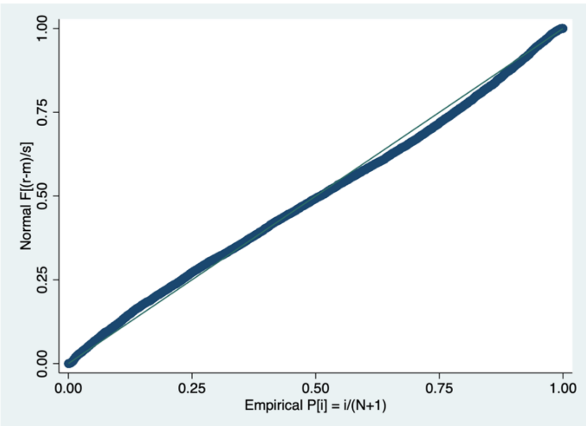

Figure 25: pnorm residuals

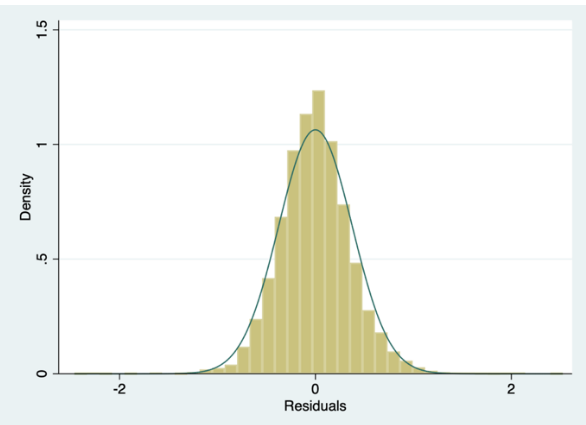

Figure 26: Histogram residuals

Appendix 10: Chow-test

Table 6: Chow-test

Chow-test analysis in Stata:
test _b[distance_cluster_areas_KM_d]=0, notest (1) distance_cluster_areas_KM_d = 0 test _b[d]=0, accum (1) distance_cluster_areas_KM_d = 0 (2) d = 0 F (2, 3796) = 0.56 Prob > F = 0.5721

Appendix 11: DO-FILE

```
*Install asdoc
ssc install asdoc, replace

*DATA PREPARATION

*Drop unneccessary variables
. drop medium_url
. drop picture_url
. drop xl_picture_url
. drop host_id
. drop host_url
. drop host_name
. drop host_since
. drop host_location
. drop host_about
. drop host_acceptance_rate
. drop host_thumbnail_url
. drop host_picture_url
. drop host_neighbourhood
. drop host_has_profile_pic
. drop host_total_listings_count
. drop host_verifications
. drop street
. drop state
. drop country_code
. drop country
. drop city
. drop market
. drop smart_location
. drop X
. drop latitude
. drop Y
. drop longitude
. drop amenities
. drop square_feet
. drop weekly_price
. drop monthly_price
. drop security_deposit
. drop cleaning_fee
. drop minimum_minimum_nights
. drop maximum_minimum_nights
. drop minimum_maximum_nights
. drop maximum_maximum_nights
. drop minimum_nights_avg_ntm
. drop maximum_nights_avg_ntm
. drop calendar_updated
. drop has_availability
. drop availability_30
. drop availability_60
. drop availability_90
. drop calendar_last_scraped
. drop number_of_reviews_ltm
. drop first_review
```

```
. drop last_review
. drop requires_license
. drop license
. drop jurisdiction_names
. drop instant_bookable
. drop is_business_travel_ready
. drop calculated_host_listings_count_e
. drop calculated_host_listings_count_p
. drop calculated_host_listings_count_s
. drop reviews_per_month
. drop availability_365
. drop require_guest_phone_verification
. drop require_guest_profile_picture
. drop review_scores_accuracy
. drop extra_people
. drop distance_local_markers
. drop is_location_exact
. drop review_scores_rating
. drop bed_type
. drop minimum_nights
. drop maximum_nights
. drop guests_included
. drop review_scores_cleanliness
. drop review_scores_checkin
. drop review_scores_communication
. drop review_scores_location
. drop neighbourhood_group_cleansed
. drop beds
. drop bedrooms

*checking data: destring/encode variables
. tab price
. gen PRICE = real(price)
. tab host_response_rate
. gen HOST_RESPONSE_RATE = real(host_response_rate)
. tab property_type
. encode property_type, gen(PROPERTY_TYPE)
. tab room_type
. encode room_type, gen(ROOM_TYPE)
. tab cancellation_policy
. encode cancellation_policy, gen(CANCELLATION_POLICY)
. tab host_is_superhost
. encode host_is_superhost, gen(HOST_IS_SUPERHOST)
. tab host_identity_verified
. encode host_identity_verified, gen(HOST_IDENTITY_VERIFIED)
. tab zipcode
. encode zipcode, gen(zip_code)
. tab host_response_time
. encode host_response_time, gen(HOST_RESPONSE_TIME)
. tab neighbourhood_cleansed
. encode neighbourhood_cleansed, gen(NEIGHBOURHOOD_CLEANSED)

*Drop missing values
. drop if PRICE == 0
. drop if PRICE == .
```

```
. drop if HOST_RESPONSE_RATE == 0
. drop if HOST_IS_SUPERHOST == .
. drop if host_listings_count == 0
. drop if review_scores_value == .
. drop if HOST_RESPONSE_TIME == 1
. drop if HOST_RESPONSE_TIME == .
. drop if bathrooms == 0
. drop if bathrooms == .
. drop if distance_cluster_areas == .
. drop if distance_cluster_areas == 0
. drop if zip_code == 0
. drop if zip_code == 3820
. drop if zip_code == 3821
. drop if zip_code == 3822
. drop if zip_code == .
. drop if NEIGHBOURHOOD_CLEANSED == .
. drop if NEIGHBOURHOOD_CLEANSED == 0

*Change distance variables into km
. generate distance_cluster_areas_KM = distance_cluster_areas/1000
. sum distance_cluster_areas distance_cluster_areas_KM
. generate distance_city_center_KM = distance_city_center/1000
. sum distance_city_center distance_city_center_KM

*Label
. label variable PRICE "price airbnb accommodation"
. label variable HOST_RESPONSE_RATE "host response rate"
. label variable HOST_RESPONSE_TIME "host response time"
. label variable PROPERTY_TYPE "property type"
. label variable ROOM_TYPE "room type"
. label variable CANCELLATION_POLICY "cancellation policy"
. label variable host_listings_count "host listings count"
. label variable accommodates "number of accommodates"
. label variable bathrooms "number of bathrooms"
. label variable number_of_reviews "number of reviews"
. label variable review_scores_value "review score value"
. label variable calculated_host_listings_count "host listings count"
. label variable distance_cluster_areas "distance cluster areas"
. label variable distance_city_center "distance city center"
. label variable distance_cluster_areas_KM "distance to cluster areas in KM"
. label variable distance_city_center_KM "distance to city center in KM"
. label variable zip_code "six-digit zip codes"
. label variable NEIGHBOURHOOD_CLEANSED "neighbourhood"

*Remove labels
. label define HOST_IS_SUPERHOST 2 "", modify
. label define HOST_IDENTITY_VERIFIED 2 "", modify

*Analyze dependent variable
. sum PRICE
. hist PRICE, normal
. graph export PRICE_hist.png, replace
. graph box PRICE
. graph export PRICE.png, replace
. qnorm PRICE
```

```
. graph export PRICE_qnorm.png, replace
*-->log transformation
. gen log_price = log(PRICE)
. label variable log_price "ln(Price)"
. hist log_price, frequency normal
. graph export log_price_hist.png, replace
. graph box log_price
. graph export log_price_box.png, replace
. qnorm log_price
. graph export log_price_qnorm.png, replace
*remove outliers
. drop if log_price < 2.8
. graph box log_price
. graph export log_price_box_removedoutlier.png, replace
. qnorm log_price
. graph export log_price_qnorm_removedoutlier.png
. sum PRICE log_price

*independent variables
**Distance to cluster areas (attractive urban green space)
. hist distance_cluster_areas_KM, normal
. graph export distance_cluster_areas_KM_hist.png, replace
. graph box distance_cluster_areas_KM
. graph export distance_cluster_areas_KM_box.png, replace
. qnorm distance_cluster_areas_KM
. graph export distance_cluster_areas_KM_qnorm.png, replace
. sktest distance_cluster_areas_KM
*--> log transformation
. gen log_distance_cluster_areas_KM = log(distance_cluster_areas_KM)
. label variable log_distance_cluster_areas_KM "ln(distance cluster areas in KM)
. hist log_distance_cluster_areas_KM, normal
. graph export log_distance_cluster_areas_KM_hist2.png, replace
. graph box log_distance_cluster_areas_KM
. graph export log_distance_cluster_areas_KM_box2.png, replace
. qnorm log_distance_cluster_areas_KM
. graph export log_distance_cluster_areas_KM_qnorm2.png, replace
. sktest distance_cluster_areas_KM

**distance attractive urban green space in categories
 generate distance_clusterareas_categories = distance_cluster_areas
*0-250m, 250-500m, 500-750m, 750-1000m, >1km
. recode distance_clusterareas_categories 1000/max=0 0/250= 1 250/500= 2 500/750 =3 750/1000=4 ,
generate(attractive_greenspace_250_1000)
. label define attractive_greenspace_250_1000 0 ">1000m" 1 "0-250m" 2 "250-500" 3 "500-750m"
4"750-1000"
. label values attractive_greenspace_250_1000 attractive_greenspace_250_1000
. tab attractive_greenspace_250_1000

*0-250m, 250-500m, 500-750m, 750-1000m, 1000-1250m, >1,25km
. recode distance_clusterareas_categories 1250/max=0 0/250= 1 250/500= 2 500/750 =3 750/1000=4
1000/1250=5 , generate(attractive_greenspace_250_1250)
. label define attractive_greenspace_250_1250 0 ">1250m" 1 "0-250m" 2 "250-500" 3 "500-750m"
4"750-1000" 5"1000-1250."
. label values attractive_greenspace_250_1250 attractive_greenspace_250_1250
. tab attractive_greenspace_250_1250
```

```
*0-250m, 250-500m, 500-750m, 750-1000m, 1000-1250m,1250-1500, >1.5km
. recode distance_clusterareas_categories 1500/max=0 0/250= 1 250/500= 2 500/750 =3 750/1000=4
1000/1250=5 1250/1500=6 , generate(attractive_greenspace_250_1500)
. label define attractive_greenspace_250_1500 0 ">1500m" 1 "0-250m" 2 "250-500" 3 "500-750m"
4"750-1000" 5"1000-1250" 6"1250-1500"
. label values attractive_greenspace_250_1500 attractive_greenspace_250_1500
. tab attractive_greenspace_250_1500

*Analyze control variables
**property characteristics
***PROPERTY_TYPE
. tab PROPERTY_TYPE

*-->delete categories with low amount of observations
. drop if PROPERTY_TYPE == 3
. drop if PROPERTY_TYPE == 7
. drop if PROPERTY_TYPE == 9
. drop if PROPERTY_TYPE == 10
. drop if PROPERTY_TYPE == 11
. drop if PROPERTY_TYPE == 12
. drop if PROPERTY_TYPE == 15
. drop if PROPERTY_TYPE == 16
. drop if PROPERTY_TYPE == 17
. drop if PROPERTY_TYPE == 18
. drop if PROPERTY_TYPE == 30
. drop if PROPERTY_TYPE == 31
. drop if PROPERTY_TYPE == 33

. tab PROPERTY_TYPE
. recode PROPERTY_TYPE 2 14 27 29=1 1 13 23 32 =2 4 6 19 20 21=3 5 28 24 22 =4,
generate(property_type_category)
. label define property_type_category 1"multi-family building" 2"single-family building" 3"room"
4"other"
. label values property_type_category property_type_category
. tab property_type_category

*-->try 2 types of categories. second one is in regression
. recode PROPERTY_TYPE 2 =1 23 32 27 =2 1 3 4 5 6 13 14 19 20 21 22 24 28 29=3,
generate(property_type_category2)
. label define property_type_category2 1"apartment" 2"house" 3"other"
. label values property_type_category2 property_type_category2
. tab property_type_category2

***ROOM_TYPE
. tab ROOM_TYPE
. replace ROOM_TYPE = 0 if (ROOM_TYPE == 1)
. replace ROOM_TYPE = 1 if (ROOM_TYPE >1)
. label define ROOM_TYPE 0 "entire home/apartment" 1 "private room" , modify
. tab ROOM_TYPE

***accommodates
. hist accommodates
. graph export accommodates_HIST.png, replace
. graph box accommodates
```

```
. graph export accomoddates_BOX.png, replace

***bathrooms
. hist bathrooms
. graph export bathrooms_HIST.png, replace
. graph box bathrooms
. graph export bathrooms_BOX.png, replace

**host characteristics
. tab HOST_RESPONSE_TIME
*->delete category with few observations
. drop if HOST_RESPONSE_TIME == 2
. recode HOST_RESPONSE_TIME  5 =0 2 3 4 =1, generate(host_respone_time_category)
. label define host_respone_time_category 0"within an hour" 1"otherwise"
. label values host_respone_time_category host_respone_time_category
. tab host_respone_time_category

. recode HOST_RESPONSE_TIME 3=1 4=2 5=3, HOST_RESPONSE_TIME_1
. label define HOST_RESPONSE_TIME_1 1"within a day" 2"within a few hours" 3"with an hour"
. label values HOST_RESPONSE_TIME_1 HOST_RESPONSE_TIME_1
. tab HOST_RESPONSE_TIME_1

***HOST_RESPONSE_RATE
. hist HOST_RESPONSE_RATE
. graph export HOST_RESPONSE_RATE.png, replace
. graph box HOST_RESPONSE_RATE
. graph export HOST_RESPONSE_RATE_BOX.png, replace

*->not normally distributed, therefore transoformation into dummy variable
. replace HOST_RESPONSE_RATE = 0 if (HOST_RESPONSE_RATE < 1)
. replace HOST_RESPONSE_RATE = 1 if (HOST_RESPONSE_RATE == 1)
. label define HOST_RESPONSE_RATE 0 "less than 100%" 1 "100%", modify
. tabulate HOST_RESPONSE_RATE
. label list HOST_RESPONSE_RATE

***HOST_IS_SUPERHOST
. replace HOST_IS_SUPERHOST = 0 if (HOST_IS_SUPERHOST == 1)
. replace HOST_IS_SUPERHOST = 1 if (HOST_IS_SUPERHOST == 2)
. label define HOST_IS_SUPERHOST 0 "NO" 1 "YES" , modify
. tabulate HOST_IS_SUPERHOST
. label list HOST_IS_SUPERHOST

***Host_listing_count
. hist host_listings_count
. graph export host_listings_count.png, replace
. graph box host_listings_count
. graph export host_listings_count.png, replace

****HOST_IDENTITY_VERIFIED
. replace HOST_IDENTITY_VERIFIED = 0 if (HOST_IDENTITY_VERIFIED == 1)
. replace HOST_IDENTITY_VERIFIED = 1 if (HOST_IDENTITY_VERIFIED == 2)
. label define HOST_IDENTITY_VERIFIED 0 "NO" 1 "YES" , modify
. tabulate HOST_IDENTITY_VERIFIED
. label list HOST_IDENTITY_VERIFIED
```

```
***number_of_reviews
. hist number_of_reviews
. graph export number_of_reviews.png, replace
. graph box number_of_reviews
. graph export number_of_reviews.png, replace

***review_score_value
. hist review_scores_value
. graph export review_scores_value.png, replace
. graph box review_scores_value
. graph export review_scores_value.png, replace

***CANCELLATION_POLICY
. tab CANCELLATION_POLICY
. recode CANCELLATION_POLICY 1=1 2=2 3/max=3, generate(cancelation_policy_category)
. label define cancelation_policy_category 1"flexible" 2"moderate" 3"strict"
. label values cancelation_policy_category cancelation_policy_category
. tab cancelation_policy_category
*-->niet significant, categorien herindelen
. generate cancellation_policy_category2 = cancelation_policy_category
. tab cancellation_policy_category2
. replace cancellation_policy_category2 = 0 if (cancellation_policy_category2 < 3)
. replace cancellation_policy_category2 = 1 if (cancellation_policy_category2 >= 3)
. label define cancellation_policy_category2 0 "less strict" 1 "strict" , modify
. tab cancellation_policy_category2

*Distance to city center (site specific characteristics)
. hist distance_city_center_KM, normal
. graph export distance_city_center_KM.png, replace
. graph box distance_city_center_KM
. graph export distance_city_center_KM_BOX.png, replace
. qnorm distance_city_center_KM
. graph export distance_city_center_KM_qnorm.png, replace
. sktest distance_city_center_KM

*--> delete extreme values
. drop if distance_city_center_KM > 5.5
. hist distance_city_center_KM, normal
. graph export distance_city_center_KM_hist2.png, replace
. graph box distance_city_center_KM
. graph export distance_city_center_KM_box2.png, replace
. qnorm distance_city_center_KM
. graph export distance_city_center_KM_qnorm2.png, replace
. sktest distance_city_center_KM

*-->Log transformation
. gen log_distance_city_center_KM = log(distance_city_center_KM)
. label variable log_distance_city_center_KM "ln(distance_city_center_KM)"
. hist log_distance_city_center_KM, normal
. graph export log_distance_city_center_KM_hist.png, replace
. graph box log_distance_city_center_KM
. graph export log_distance_city_center_KM_box.png, replace
. qnorm log_distance_city_center_KM
. graph export log_distance_city_center_KM_qnorm.png, replace
*delete outliers
```

```
. drop if log_distance_city_center_KM < -1.8
. hist log_distance_city_center_KM, normal
. graph export log_distance_city_center_KM_hist2.png, replace
. graph box log_distance_city_center_KM
. graph export log_distance_city_center_KM_box2.png, replace
. qnorm log_distance_city_center_KM
. graph export log_distance_city_center_KM2_qnorm.png, replace

*City districts
. recode NEIGHBOURHOOD_CLEANSED 5 6=1 3 8 21 14=2 7 11 19 16=3 4 9 22=4 20 18 15 13
12=5 17=6 1 2 10=7, generate (stadsdelen)
. label define stadsdelen 1 "Centrum" 2 "West" 3 "Nieuw-West" 4 "Zuid" 5 "Oost" 6 "Noord" 7
"Zuidoost"
. label values  stadsdelen stadsdelen
. tab stadsdelen
*Merge 2 datasets (airbnb data and BBGA data) for controlling
. generate STADSDELEN = stadsdelen
. merge m:1 STADSDELEN using stadsdelen, gen(mer1)

**First regressions**
. regress log_price distance_cluster_areas_KM i.property_type_category i.ROOM_TYPE
accommodates bathrooms i.host_respone_time_category i.HOST_RESPONSE_RATE
i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value i.cancelation_policy_category distance_city_center_KM

*regression with appropriate categories
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE
accommodates bathrooms i.HOST_RESPONSE_TIME i.HOST_RESPONSE_RATE
i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value i.cancellation_policy_category2 distance_city_center_KM

*ANALYSIS
**correlation matrix
. corr distance_cluster_areas_KM attractive_greenspace_250_1500 property_type_category2
ROOM_TYPE accommodates bathrooms  HOST_RESPONSE_TIME_1 HOST_RESPONSE_RATE
HOST_IS_SUPERHOST host_listings_count HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value cancellation_policy_category2 distance_city_center_KM

**export correlation matrix to word
. asdoc corr distance_cluster_areas_KM attractive_greenspace_250_1500 property_type_category2
ROOM_TYPE accommodates bathrooms  HOST_RESPONSE_TIME_1 HOST_RESPONSE_RATE
HOST_IS_SUPERHOST host_listings_count HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value cancellation_policy_category2 distance_city_center_KM

**VIF
. regress log_price distance_cluster_areas_KM i.attractive_greenspace_250_1500
i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i. HOST_RESPONSE_TIME_1
i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count
i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value
i.cancellation_policy_category2 distance_city_center_KM
. estat vif
*VIF without categories in distance
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE
accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE
```

i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM
. estat vif
. asdoc vif
*VIF categories in distance
. regress log_price i.attractive_greenspace_250_1500 i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM
. estat vif
. regress log_price i.attractive_greenspace_250_1500 i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM
. estat vif
. asdoc vif

*Summarize results - descriptive statistics
. sum log_price distance_cluster_areas_KM attractive_greenspace_250_1500 property_type_category2 ROOM_TYPE accommodates bathrooms HOST_RESPONSE_TIME_1 HOST_RESPONSE_RATE HOST_IS_SUPERHOST host_listings_count HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value cancellation_policy_category2 distance_city_center_KM

**Export descriptive statistics to word
. asdoc sum log_price distance_cluster_areas_KM attractive_greenspace_250_1500 property_type_category2 ROOM_TYPE accommodates bathrooms HOST_RESPONSE_TIME_1 HOST_RESPONSE_RATE HOST_IS_SUPERHOST host_listings_count HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value cancellation_policy_category2 distance_city_center_KM

*lineaire relatie x & y
. twoway (scatter log_price distance_cluster_areas_KM) lfit log_price distance_cluster_areas_KM
. graph export twoway_scatter_clusterareas.png

*TESTING ASSUMPTIONS
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM

**assumption 2 Homoscedasticity
. rvfplot, yline(0)
. graph export rvplot.png, replace
. estat hettest
. asdoc estat hettest
. estat imtest, white
. asdoc imtest , white

*robust standard errors
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM, vce(robust)

**assumption 3 No autocorrelation
*clustered standard errors
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM, vce(cluster zip_code)

**assumption 4 Independence
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM
. estat vif
. asdoc vif

**assumption 5 Normality of residuals
. predict r, resid
. hist r, normal
. kdensity r, normal
. pnorm r
. qnorm r

. graph export hist_residual_normal.png, replace
. graph export kdensity_residual_normal.png, replace
. graph export pnorm_residual.png, replace
. graph export qnorm_residual.png, replace

*export models
**model 1 base specification
. regress log_price i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM i.stadsdelen, vce(cluster zip_code)
. outreg2 using myreg.doc, append ctitle(model 1) addtext(spatial controls, Neighborhood) label
**model 2 main specification
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM i.stadsdelen, vce(cluster zip_code)
. outreg2 using myreg.doc, append ctitle(model 2) addtext(spatial controls, Neighborhood) label
**model 3 including categories in distance
. regress log_price i.attractive_greenspace_250_1500 i.property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value i.cancellation_policy_category2 distance_city_center_KM i.stadsdelen, vce(cluster zip_code)
. outreg2 using myreg.doc, append ctitle(model 3) addtext(spatial controls, Neighborhood) label

*chow test
. recode distance_clusterareas_categories 100/max=0 0/100= 1, generate(attractive_greenspace_view)
. label define attractive_greenspace_view 0 "no view" 1 "view"
. label values attractive_greenspace_view attractive_greenspace_view
. tab attractive_greenspace_view

```
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE
accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE
i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value i.cancellation_policy_category2 distance_city_center_KM, robust
cluster(zip_code)
. ereturn list
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE
accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE
i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value i.cancellation_policy_category2 distance_city_center_KM if
attractive_greenspace_view == 0, robust cluster(zip_code)
. ereturn list
. regress log_price distance_cluster_areas_KM i.property_type_category2 i.ROOM_TYPE
accommodates bathrooms i.HOST_RESPONSE_TIME_1 i.HOST_RESPONSE_RATE
i.HOST_IS_SUPERHOST host_listings_count i.HOST_IDENTITY_VERIFIED number_of_reviews
review_scores_value i.cancellation_policy_category2 distance_city_center_KM if
attractive_greenspace_view == 1, robust cluster(zip_code)
. ereturn list
. gen d = 0
. replace d = 1 if attractive_greenspace_view == 1
. gen distance_cluster_areas_KM_d = distance_cluster_areas_KM * d
. regress log_price distance_cluster_areas_KM distance_cluster_areas_KM_d d i.
property_type_category2 i.ROOM_TYPE accommodates bathrooms i.HOST_RESPONSE_TIME_1
i.HOST_RESPONSE_RATE i.HOST_IS_SUPERHOST host_listings_count
i.HOST_IDENTITY_VERIFIED number_of_reviews review_scores_value
i.cancellation_policy_category2 distance_city_center_KM, robust cluster(zip_code)
. test _b[distance_cluster_areas_KM_d]=0, notest
. test _b[d]=0, accum
```