NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação

Master Program in Statistics and Information Management

Modelling the Airbnb listings' price in Lisbon using local spatial regressions

Ivanildo Semedo Correia Fernandes

Dissertation presented as partial requirement for obtaining the Master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management Sc
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

MODELLING THE AIRBNB LISTINGS' PRICE IN LISBON USING LOCAL SPATIAL REGRESSIONS

by

Ivanildo Fernandes

Dissertation presented as partial requirement for obtaining the Master's degree in Statistics and Information Management, with a specialization in Information Analysis and Management

Advisor: Ana Cristina Costa

November 2018

ABSTRACT

Sharing economy market, such as Uber and Airbnb, have been growing rapidly in the last few years, providing extra income to agents from the supply side, and low costs to those in demand side. Although its adoption provided benefits for stakeholders and to the global economy of the areas in which they are inserted, several authors and politicians have been referencing the negative externalities brought with it, such as an increase in rents and real estate prices and a decrease in hotels' revenue. However, most of the externalities pointed out, were not based on any empirical analysis.

The aim of this study is to analyze Airbnb market within Lisbon municipality, focusing mainly the modelling spatial variation of Airbnb listings' price. For this purpose, it was employed an ordinary least square (OLS) model and a geographical weighted regression (GWR) model to identify the main factors affecting the Airbnb listings' price. The results showed that the GWR model performs better than the OLS model, and it allows assessing the local impact of the explanatory variables on the Airbnb listings' price. In conclusion, it was found that the price of the two types of Airbnb listings (entire home/apartments and private/shared rooms) are not affected by the same factors and that statistically significant differences varied across parishes within Lisbon municipality. Perhaps, there is a need to test if it is plausible to apply a unique regulatory policy considering Airbnb and Lisbon market as an aggregated concept or by Airbnb listing type and Lisbon parishes.

KEYWORDS

Airbnb; Listing; Accommodation; Lisbon; Price

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AIC	Akaike's Information Criterion
AICc	Corrected Akaike's Information Criterion
EDA	Exploratory Data Analysis
GWR	Geographically Weighted Regression
OLS	Ordinary Least Squares
RNAL	National Registration of Local Accommodation
RNT	National Tourism Registration

1 INTRODUCTION

In recent years, the so-called sharing economy has emerged as a new socioeconomic system that have changed the way goods and services are created, produced, distributed and consumed among individuals. These platforms enable individuals to easily share underutilized inventories with one another, e.g., houses and cars, through fee-based sharing (Alsudais, 2017). Sharing economy has gained a significant place in the transportation and hospitality industry, thus providing propitious opportunities for individuals and corporates to create small, medium and large businesses with growth at a rapid pace (Tussyadiah & Pesonen, 2015). Technological innovation and supply-side flexibility, were pointed as being the main factors that enabled the rapid growth of peer-to-peer platforms and the appearance of successful companies providing a range of different services, such as local accommodation (Airbnb), transportation (Uber and Lyft) and social network (TaskRabbit). Nowadays, Airbnb and Uber are considered the most successful sharing economy platforms (Quattrone, Proserpio, Quercia, Capra, & Musolesi, 2016).

In this study all the attention will be focused on the local accommodation market, more specifically on Airbnb. Airbnb defines itself as a global community that provides its guests with a unique experience from the beginning to the end of their journey, which includes accommodation, activities and the possibility to "live like a local" (https://www.airbnb.com). It acts as an intermediary between individuals with extra spaces and those that are looking for accommodation.

Founded in 2008, Airbnb already has millions of listings registered in more than 191 countries, in which it offers range from a simple apartment to castles, boats and houses built into the trees. Although Airbnb has emerged in recent years, the concept of local accommodation is not new (Gallagher, 2017). There are many stories of people who stayed in some type of "sharing economy" even before there was internet. In fact, there is a Wikipedia page regarding the emergence of the "*Homeshare*" topic, in which Airbnb is not even mentioned[1]. However, Airbnb benefited from some factors that have made the company the most successful local accommodation services provider nowadays. For instance, the ability to advertise properties for free on a simple, efficient and attractive platform, the establishment of a reliable social network and the promotion of different experiences, were given as examples.

According to Cansoy & Schor (2016), individuals are likely to participate in the sharing economy market only if the marginal utility (profit) withdrawn from it is greater than its costs. In this sense, it is expected for low-income individuals to be predominant in this market, since the obtained benefits (access to goods and services) are superior, or more significant when compared with high-income individuals. Nevertheless, the total travel expenses incurred by Airbnb customers are not necessarily lower than those who use other services, e.g., hotels, since some travelers have a certain budget plan, which may imply spending less in accommodation, by using Airbnb, to be possible to save more money for other trip related expenses (Forgacs & Dimanche, 2016). Due to its variety of listing and scale prices, it can be said that Airbnb agents is composed by individuals with different needs and from different social status.

Since 2010, period that Airbnb was adopted in almost every European countries' capitals, its growth has been exponential, by offering travelers and owners, a service that goes according to their needs.

[1] *https://en.wikipedia.org/wiki/Homeshare*

However, this adoption is controversial (Coyle & Yeung, 2016). According to Gallagher (2017), Airbnb came at a time in which there is a disconnection in the society, with a record number of people living alone and spending more time isolated, others that lost their jobs or are simply wondering around. The nature of the local accommodation also made Airbnb dealt with all the unintended consequences of putting strangers together, including attacks and lapse. Consequently, in recent years, Airbnb had to confront another issue on its platform, the presence of racial and other types of discrimination. In recent years, several regulatory battles took place in different cities around the world, due to the rapid growth of the sharing economy. On the one hand, some argue that Airbnb adoption can bring several benefits, including extra income for agents providing this type of service, better allocation and use of resources and new economic activities for cities and their respective parishes. On the other hand, some critics pointed to the negative externalities arising from its adoption, such as alleged disruption in local residential structures, increase in rents and decrease in the hotel revenues.

Researchers have developed several investigations regarding how to regulate the Sharing Economy market. However, most of them did not have as main concern the understanding of the phenomenon's features, how it has been adopted throughout the years and who benefit from it. That said, there are not enough reliable content so that the competent authorities can based on it when defining regulatory policies. It's important to understand in which location there is a strong/week presence of local accommodation, if this type of sharing economy must be taken as being an aggregated concept, with no significant variations within its modalities, or if there is a need to split the market in groups due to its significant differences and impacts.

Price is also another issue attached to Airbnb market. Until now, Airbnb properties (listings), registered on its platform, are not restricted to any type of pricing policy. According to Airbnb itself, hosts are completely free to assign the desired price to be charged for sharing a property. Under this condition, it is difficult for the regulatory authorities to assess the market value based on the existing listings and their features.

In this study, the main purpose is, instead of providing a list of regulatory policies to be implemented, to give a set of indicators to be taken as reference when regulating the local accommodation market. More specifically, the main objective of this research is to identify factors that influence the Airbnb listings' price within Lisbon municipality. The methodological approach will be based on linear regression models.

From that general objective, a set of secondary objectives emerge:

- To gather reliable data for potential explanatory variables;
- To investigate patterns of spatial correlation in the data;
- To evaluate whether Airbnb listings should be analyzed as whole or separated by listings' type;
- To estimate a somewhat reliable linear regression model using Ordinary Least Squares (OLS);
- To estimate a Geographically Weighted Regression (GWR) model;
- To compare the models' performance.

2 LITERATURE REVIEW

Airbnb diversity is one of the key factors which allowed the company to fulfill different user needs. The listings registered on its platform are categorized as entire apartment, when neither host nor other guests will be present in the home during one's stays, meaning that the guest has unique and exclusive access to the entire home/apartment, as private room when the exclusive access is applied only to one room, and as shared room when the guest must share the entire apartment, including the room, with the host and/or other guest(s).

Despite the different type of listings offered by Airbnb, there is a noticeable preference for a specific type of listing due to its high number of demand and supply. Considering a sample of more 2 million Airbnb listings, collected from 193 countries worldwide, Ke (2017) found that 68.5% of the listings are classified as entire home/apartments, 29.8% as private rooms and the remaining 1.7% as shared rooms. Listings classified as entire/home apartments are predominant in almost every city where Airbnb operates, e.g., Barcelona (Gutierrez, Garcia-Palomares, Romanillos & Salas-Olmedo, 2016), Lisbon (Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa, 2016) and London (Quattrone et al., 2016). The previous conclusion raised some doubts regarding Airbnb as being related with the term “sharing”.

Although the global idea associated with the sharing economy market, especially that local accommodation is about sharing, social interaction and integration, according with previous remarks, some market players prefer to spend more money in exchange to more space and privacy. This situation occurs when guests choose entire home/apartment rather than private/shared room, where there is a little or even no direct contact with the host or other guests (Dogru & Pekin, 2015). Furthermore, not all the Airbnb listings are functioning as a pure local accommodation business and some of them are not provided by singular individuals. Gyódi (2017), concluded that, in Warsaw, Poland, local accommodation market is not only provided by individuals seeking for extra income, but mostly by professional agencies (Business-to-Consumer). In addition, they also mentioned Airbnb as being functioning as an alternative to long-term house rental, rather than short-term for the majority of the registered listings. Likewise, Cardoso (2015) mentioned that although Airbnb listings are used mainly for short-term rent, they are often used as a permanent leasing option. On the other hand, in 2015, 82% of the Airbnb hosts in Boston, United States, had only one listing registered on Airbnb platform, which means that hosts are composed mostly by singular individuals looking for extra income (Lee, 2016). However, commercial operators are very active in this type of market since the remaining 18% of the hosts hold 46% of the registered Airbnb listings. Besides the global idea regarding Airbnb, a platform that enable individuals to share extra space with individuals who are looking for accommodation, in exchange of money, is true, this type of market is constantly changing, and it can present different patterns depending on the area where it is inserted.

Since the main goal of local accommodation market is to provide guests with a cultural experience, it becomes more attractive to foreign tourist rather than local tourist (Neeser, 2015). According to Airbnb summer travel report 2015[2], Airbnb guests traveling during the summer came from all over the world, traveling from more than 57,000 different cities. As Airbnb guests are mainly foreign tourists from different countries and speaking different languages and considering the fact that the

[2] Report: *https://blog.atairbnb.com/wp-content/uploads/2015/09/Airbnb-Summer-Travel-Report-1.pdf*

communication using Airbnb platform is stablished exclusively between host and guest, there is a need to choose a language that can be understandable in both sides. Based on Airbnb listings reviews worldwide, Ke (2017) mentioned that, in terms of communication between host and guest, the predominant languages found in the comments were English, with a weight of 72.8%, followed by French (10,3%) and Spanish (3.8%). Portuguese language was found with a weight of 1%.

Local accommodation market is constantly growing, conquering new cities and clients worldwide. One of the frequently asked questions is "*Where are Airbnb listings located?".* At the continental level, Airbnb listings are concentrated mostly in Western Europe, North America, East and South Asia and Pacific Asia. At the country level, the United States leads the ranking of countries with the highest number of listings, followed by France, Italy, Spain and the United Kingdom (Ke, 2017). Within each city, Airbnb listings are more concentrated in the city center (Gyódi, 2017).

As previously mentioned, Airbnb guests are more likely to choose entire home/apartments listings over private/shared rooms. However, guests are not concerned only with the type of listings. When choosing a listing, Dogru & Pekin (2015) found that the space, cleanliness, number of photos, accessibility, family atmosphere, free breakfast, location and unique experiences are the most valued factors by Airbnb guests. Given that within the local accommodation market there are preferences on the demand side, in terms of listings characteristics, Airbnb hosts also apply different prices scale for different listings. In this sense, is there a criterion used for hosts to define prices according to each listing type and features? As described on the Airbnb platform, hosts are completely free to assign the desired price to be charged for sharing the property. However, there is a pricing tool which provides hosts with competitive rate recommendations. Despite this, it is not possible to accurately assess local accommodation market prices, which would allow hosts to be able to set fair and competitive prices based on listings attributes. Indeed, there is lack of strategic information linked with market segmentation and functioning, competitors and related geographic area, which turn it difficult for investors to evaluate their own property against the rest and apply the best pricing policies in order to be well succeeded within the Airbnb market.

Whenever an income is spent on a product/service, it becomes the income of another individual or enterprise, generating a cyclical income flow. For this reason, the economic impact of a new service/product is not always limited to the areas where it is inserted. Through an analysis on how Airbnb guests' expenses in one area may impact other areas, within New Orleans city, Levendis & Dicle (2016) concluded that, for instance, expenses on food and housing generate profits to landlords who in turn spend part of their income locally and the remaining in other regions, namely on clothing, food, accommodation and transport, generating income to other market players outside the areas covered by Airbnb market. Considering that Airbnb guests, in addition to the accommodation costs, also spend money on touristic activities, e.g., entertainment, transportation and souvenirs (Zhang, Chen, Han, & Yang, 2017), it also has an impact on the local employment rate.

Despite the benefits provided by Airbnb to those who operate in the local accommodation market, its rapid growth is changing the tourist accommodation model in a way that it has generated several conflicts worldwide, especially in those cities with mass tourism (Gutierrez et al., 2016). In fact, studies conducted by many researchers in different cities worldwide, found evidences that local accommodation can affect rents. For instance, in the case of Barcelona, since more than half of the Airbnb listings consists of entire home/apartment, the rental flats were being removed from the market due to the Airbnb expansion and consequently increasing rents and driving out local population (Gutierrez et al., 2016). Furthermore, it was noticed a transformation in the business structure of these areas justified by an increase in shops and restaurants geared to tourists.

Since Airbnb emergency, the city of Los Angeles, California, United States, has been visited by over than forty-five million tourist each year, whom stay concentrated in its seven most expensive and density neighborhoods (Lee, 2016). During 2014, this increase had an impact specially on these seven neighborhoods, making rents 20% higher and increasing 33% faster than in the remaining city's neighborhoods. According to Merante & Horn (2016), if the 24% growth registered on the Airbnb listings during 2015 in Boston continues for the next three years, *ceteris paribus*, rents will be $178/month higher. From the economic point of view, one of the main reasons that can explain such increases is the fact that there is not enough supply to keep up with increases in demand for local accommodation, generating a pressure for increases in rents and real estate prices. According to Lee (2016), each property (entire house, apartment or single room) that was previously occupied by a resident of a given city, and now is used to provide local accommodation services, is a property that has been removed from the rental market. Similarly, Merante & Horn (2016) support the idea that local accommodation market is driving up rents by decreasing the supply of residential properties that used to be available for locals.

It has been proved that the rise and rapid expansion of local accommodation market has affected rents and real estate prices in different cities around the world. However, the negative impacts of the local accommodation market are not reflected only on rents and real estate prices. Services provided by this type of sharing economy are comparable to those provided by hotels, which make them direct competitors (Lehr, 2015). Zervas, Proserpio, & Byers (2013), estimated that Airbnb entry into the Texas state market in the United States, had a negative impact on revenue generated by hotel rooms. On average, each 10% increase in Airbnb market resulted in a decrease of 0.39% in revenues generated by hotel rooms. They also mentioned that hotels located in areas with a strong presence of Airbnb listings, were forced to lower their prices to face the competition.

By comparing hotels' revenue before and after the Airbnb introduction to Oslo, Norway, Ytreberg (2016), estimated that a 10% increase in the supply of Airbnb listings is associated with a 0.307% decrease in hotel revenues. Nevertheless, hotels operating in the low and medium price segment are most affected and the Airbnb impact may be partial. According to Gyódi (2017), unlike the Airbnb listings classified as entire home/apartment, the private/shared rooms are the ones considered to directly compete with traditional hostels. Equivalent to the previous researches conclusions regarding Texas, in Oslo, hotels were also forced to strategically lower their prices in response to Airbnb massive grow rather than increasing their occupancy rates. Another important remark is that hotels in Oslo can face the effects brought by Airbnb emergency by increasing their supply since hotels' revenues are more impacted by increases in hotel supply rather than increases in Airbnb supply.

Although there are studies proving the Airbnb impact on hotels' revenues, this scenario only suits for areas in which these studies were developed. There is no statistical evidence supporting the idea that hotels' revenues worldwide are affected by Airbnb market. For instance, in Korea, despite the increases in the number of tourists, it was found that Airbnb listings do not have impacts on hotels' revenue (Choi, Jung, Ryu, Kim, & Yoon, 2015). In the United States, since the launch of Airbnb, 2008, until the end of 2014, the hotels' occupancy rates were not affected in every city (Goree, 2016). While in San Francisco, California, the Airbnb introduction did not significantly impact the hotels' occupancy rate, in Chicago, Illinois, this introduction had a marginally significant negative impact on the occupancy rate. Neeser (2015), found that in some European countries, such as Norway, Finland and Sweden, besides the Airbnb contribution to the reduction of the average price of hotel room, it does not have a significant effect on hotels' revenue per available room. Most of hotels' customers are individuals travelling for work whom did not replaced hotel services by Airbnb services (Goree, 2016), which can make hotels' revenue per available room almost unchanged in some markets.

Although hotels and Airbnb are considered direct competitors, they work under different business model, which can benefit one more than the other. As mentioned before, Airbnb flexibility to give response to demand in different seasonal periods has significantly limited hotel pricing in periods of high demand. Furthermore, unlike what happen for hotels, there is no cost related with increases in Airbnb listings supply, since its platform is free, facilitating the emergence of new listings.

Researchers went further trying to estimate Airbnb impact in different field. Aside from hotels, real state and rents, there are studies linked with the estimation of local accommodation market impacts on social environment field. Xu, Kim, & Pennington-Gray (2017) conducted a study based on data from 67 counties in the state of Florida, United States, and proved that there is a significant spatial relationship between Airbnb and crimes, i.e., Airbnb is positively related with property crime (robbery and motor vehicle theft) and negatively related with violent crime (murder and rape). Although Airbnb has a significant positive relationship with crimes, it was found that this relationship varies according to the listing type. While private and entire home/apartment has negative correlations, shared rooms are positively related with crime, especially in less touristic intense areas.

It is worth mentioning that the large number of researches related with local accommodation impacts on other existing markets have captured Airbnb's attention, which led it to conduct its own research. Contrary to previous statements, Airbnb survey on the economic impacts of space sharing in cities around the world has founded that 74% of Airbnb listings are properties outside of major hotel areas[3]. It was also mentioned the record growth in the daily occupancy rate of hotels simultaneously with local accommodation market evolution, making Airbnb a complementary element for the existing tourism industry in Paris. Airbnb also highlighted the fact that 81% of its hosts worldwide share the house where they live in. The previous conclusion is controversy with those brought by Lee (2016) and Merante & Horn (2016), stating that local accommodation market is driving up rents by decreasing the supply of residential properties.

The rapid growth of Airbnb market in touristic cities creates a regulatory challenge since some researchers found empirical evidence pointing to the significant impact of local accommodation on the traditional hotel industry services, housing market and social environment (Gyódi, 2017).

According to Gutierrez et al. (2016), researchers are not giving enough attention regarding the emergence and expansion of local accommodation platforms in touristic cities, especially in terms of the location and impacts in the city. Although, some studies have proposed several regulations criteria for adjusting local accommodation market with the aim of reducing its negative impacts, most of the suggested regulatory policies were not based on empirical evidences, e.g., the

[3] Airbnb research: *https://www.airbnb.pt/economic-impact*

understanding of different types of local accommodation, its adoption and impacts in each neighborhood, city and country (Quattrone et al., 2016). In practice, instead of regulating the emergence of the local accommodation based on empirical evidences, some cities decided to embrace this type of market without imposing any regulatory policy, while others opted to ban this practice completely from the market. In addition, several municipalities were forced to regulate local accommodation market based on old regulatory policies that were previously created to be applied in a market with different features. According to Jefferson-Jones (2015), laws were designed to regulate relationships in a competitive economy, not in a collaborative one.

Since most listings registered on the Airbnb platform are entire home/apartment, critics are using this statement to justify the increases in rents and real estate prices, claiming that the properties that are being used for local accommodation are those that were previously used by locals (Ke, 2017). Among other arguments are the alleged disruption in local residential structures and the failure to pay fees, especially when the owner holds multiple listings.

In Los Angeles, California, despite the benefits brought by the Airbnb adoption to the local economy, some criticism emerged stating that this type of market harms neighborhoods, distorts the housing market, undermines labor unions and aggravates housing crisis (Lee, 2016). Complementary to the previous criticisms regarding Airbnb impact on other markets, Edelman & Luca (2013) found strong evidence pointing to online discrimination between Airbnb host and guest. In order to bring relevant evidences to debates related to sharing economy, Cansoy & Schor (2016) analyzed Airbnb listings information from the cities and neighborhoods of the United States of America whose population is less than 500,000 inhabitants, and concluded that despite of Airbnb's claims concerning its positive impact in the middle class and income deviation for more diversified and low-income areas are true, there is also evidence that high-income areas are likely to be more profitable and have better rating than low-income areas.

Due to numerous criticisms, local accommodation market expansion caught government's attention from different countries worldwide, which are trying to find regulatory policies that fits this type of market. Along with governments, many researchers are conducting studies on how to regulate local accommodation market. According to Quattrone et al. (2016) theory, the best way to regulate the local accommodation market is to take into account how, when and where to regulate. Among the mentioned suggestions, it was highlighted the regulation through the transfer of rights, i.e., each owner should have the right to integrate into the local accommodation market for a certain period, with the option to effectively transfer its right to another owner. In terms of "what to regulate", it must be considered that the local accommodation market is not uniform, i.e., renting a private/shared room may not have the same impacts as renting an entire home/apartment. On the other hand, Cansoy & Schor (2016) referred to education as being the key piece for the understanding of local accommodation market features, since the areas inhabited by individuals with a higher education level stablish higher prices for Airbnb listings, receive more reviews, better ranking and consequently generate more income. On the other hand, for Lee (2016), the precise data is the key to better regulate the Airbnb market. The large number of Airbnb listings whose generate profits from illegal rentals is pointed as being the main cause of rent increases, reduction in housing supply and segregation.

With the aim of collecting more taxes over Airbnb market and reduce its advantage over the traditional accommodation in Barcelona, the government is trying to control the expansion of this type of accommodation by imposing periodical inspections to ensure that taxes are paid and that all the properties are functioning legally (1.8). In addition, fines up to 90,000€ were imposed. Similarly, Lee (2016) recommended the implementation of an occupancy tax around 14% to any property listed on the Airbnb platform in Los Angeles, California, for more than 75 days to prevent Airbnb host to gain unfair competitive advantage over hotels. Together with the occupancy tax, it was also suggested that Los Angeles should set a maximum number of listings owned by each host and the listings lifetime (maximum number of days that a listing is allowed to be available to the public).

From Jefferson-Jones (2015) point of view, local accommodation shall be regulated based on 5 restrictions:

- **full prohibitions** for some localities;
- **quantitative restriction** by imposing a limit on the number of units available in the market;
- **proximity restriction** by forbid the registration of new properties due to its proximity to existent properties or a vulnerable area;
- **operational restriction** which is related to how each property should operate in the market, e.g., imposing a maximum overnight occupancy based on the capacity of each property and the maximum number of times that each property is allowed to operate in the market; and
- **licensing restriction** in which each individual or enterprise who wants to operate in local accommodation market must submit the property through and inspection in order to obtain a license to do so.

In fact, there are a lot of suggestion on how to regulate local accommodation market. However, according to the previous reviews, Airbnb can present different features depending on the area in which it is inserted. Airbnb impacts in Lisbon may not be the same as in London meaning that the regulatory policies applied in a specific city may not fit perfectly if applied in a different city. Regulatory policies must be designed taking into consideration a given neighbourhood, city or a country, and be based on empirical evidences.

Tourism sector in Portugal reached its highest value ever on May 2016, and Lisbon played a major role in this development, since it is an electoral destination whose occupancy rate (72.5%) has exceeded those from destinations such as Rome, Madrid and Paris[4]. The diversity of the tourist offers, the security, the mild climate throughout the year and the good air and sea accessibilities are pointed out as the main factors that justify this reality.

Portugal's tourism market plays an important role in the national and regional economic development. In 2013, the tourism sector accounted for almost 13% of the number of companies, 5% of turnover and 10% of the number of people employed by the total non-financial corporations in the country (Vieira & Moreira dos Santos, 2016). "Accommodation and catering" were the predominant economic activity concerning to the number of persons employed (75%) and the number of companies (71%). In terms of geographical location, around 56% of the companies, 79% of the turnover and 66% of the number of people working in the sector, were associated with companies located in Lisbon, Oporto and Faro districts. From 2005 to 2014, the number of overnight stays

[4] Document: *https://www.visitlisboa.com/sites/default/files/2017-06/RTL_0617_Jun.pdf*

increased by 30%, with the majority being foreign tourists coming mainly from countries, such as United Kingdom, Germany, Spain, France and the Netherlands. In this context, these five countries accounted for 64% of foreign overnight stays in Portugal in 2014. Outside the European continent, tourists from Brazil and the United States are the most frequent.

Comparable to what happened in other massive touristic cities around the world, Lisbon did not escape to "Airbnb fever". In 2015, there were approximately 44,808 listings registered on *Airbnb.pt*. Lisbon, besides being the capital of Portugal, it is also the city with the largest number of Airbnb listings in the country. In the same year, Lisbon accounted with more than 9,200 listings (Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa, 2016). Regarding the relative weight, the Airbnb Lisbon market report showed that the proportion of real estate dedicated to local accommodation in Lisbon is more than 10%, being more concentrated in *Santa Maria Maior* (22%), *Misericordia* (18.5%) and *Santo António* (11.1%) parishes.

According to Cruz (2016), the main factors related to the large number of local accommodations offer in Lisbon are the lack of hotels in the historical city center, the high number of vacant buildings, the economic crisis in Portugal and legislative changes. On the other hand, from the Airbnb point of view, Lisbon's low cost of living, culture and the regeneration of old buildings were pointed as the main reasons for the success of local accommodation market in the city.

Although local accommodation figure was introduced into Portugal's legal system in 2008, it was in 2014 when its own legal regimes was created by considering the similarities and differences to the existing tourist offers (Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa, 2016). Within the Portugal's territory, local accommodation is defined as being apartments, dwellings or lodging establishments that provide temporary accommodation services to tourists for less than 30 days in exchange for a fee and that do not meet the requirements to be considered tourist enterprises. The activity of providing local accommodation services may be operated by a natural or legal person through advertising means, such as travel and tourism agencies or internet sites and, in addition to overnight stays, other services may be provided, namely cleaning or reception.

Related to what happened in other cities, some researchers pointed to the negative impacts brought by Airbnb to the city of Lisbon. Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa (2016), concluded, by comparing the evolution of home prices in the parishes with high and low presence of local accommodation, that after the legislative amendment which took place in 2014, the effect of the local accommodation market in Portugal has caused an increase of 13.2% (1.48€/m^2) in rents and 30.5% (651€/m^2) in real estate price. These increases are more reflected in the parishes of Lisbon and Oporto cities.

Similarly, Cruz (2016) also mentioned some Airbnb negative externalities, such as the rising cost of properties in historic centers and also the increase in rents caused by demand pressure for buildings and fractions to convert into local accommodation, which made housing less accessible for residents in some Lisbon neighborhoods.

Besides the previous remarks, it was also found some positive benefits brought by Airbnb adoption in Lisbon city. Contrary to what happened in some countries, Airbnb did not have a negative impact on the hotel market, since the number of foreign tourists visiting Lisbon is constantly increasing in the past few years allowing both, local accommodation and the traditional hotel market to be benefited from it. Local government was also benefited from it. Tourist tax on overnight stays applied by the Lisbon city council totaled a revenue of 16.7 million euros between January 2016 and March 2017, with 1.1 million euros being delivered by Airbnb. Local accommodation also has an important and positive contribution to urban rehabilitation in Lisbon, directly and indirectly increasing the number of interventions in the buildings and fractions across the historic center of the city. In addition, local accommodation not only increased the tourism revenues, but also a more decentralized and equal distribution of these gains.

According to Cruz, (2016), approximately 25.5% of the local housing units in Lisbon are in tourist-only buildings, 63.7% are inserted in residential buildings and the remaining 42.5% within commercial buildings or offices. Despite the various studies on the hotel industry, especially on hotels and resorts, there is still a reduced number of researches on this type of short-term tourist accommodation known as local accommodation.

3 DATA AND METHODS

3.1 DATA COLLECTION

Local accommodation is offered mostly through digital platforms, with Airbnb and *HomeAway* being the most relevant. According to a study developed by the Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa (2016), in 2015, 44,808 properties in Portugal were registered on *Airbnb.com* (13,478 more than in RNAL), and 24,662 properties registered on *Homeaway.com* (6,668 less than in RNAL). These values are distributed mainly among the parishes of the Lisbon municipalities, Oporto and Algarve. Based on this exercise, it was concluded that the Airbnb platform provides a better coverage of the local accommodation market in Portugal, when compared to data collected by local sources or other local accommodation platforms. These differences occur due to the large number of properties rented without any formal agreement.

The chosen area to perform the local accommodation market analysis was Lisbon municipality. Even though it is the capital of Portugal, Lisbon is composed by 24 parishes and it has the largest number of local accommodations registered on the Airbnb platform in Portugal. As mentioned in the previous chapter, the tourism sector in Portugal has been significantly increasing in recent years and Lisbon played a major role on its evolution. The diversity of the touristic offers, security, mild climate throughout the year and the good air and sea accessibilities, made Lisbon occupancy rate exceed those from destinations, such as Rome, Madrid and Paris.

For the time being, there is not an official source retained by Airbnb providing raw data from the listings registered on its platform to the general public. However, to address this issue, many researchers used data that have been extracted directly from the Airbnb website (Sheppard & Udell, 2016). Sources such as *Inside Airbnb*[5] and *Tom Slee* [6] made available a large amount of Airbnb data regarding different cities worldwide, with different periodicity and no charge associated. For this study, data collected from *Tom Slee* platform were used instead of *Inside Airbnb*, since the second source do not provide data for Portugal market. *Tom Slee* source code is in Python 3, a widely used high-level programming language, and it scrapes data from the Airbnb website for a city (labelled as search area) and stores the result in a database. Each collection of a single city is called a survey. A single database holds many separate surveys, including some of the same city.

Once the main goal of this research is to geographically analyses the spatial behavior of Airbnb listings' price, one important step to achieve it, was to collect spatial data from the study area. *Geodados*[7] is a platform created by Lisbon city hall for the provision of geographical data regarding Lisbon municipality, in the education, accommodation, entertainment, sports, environment, culture and other fields.

To be combined with Airbnb data, it was collected from *Geodados* the following information:

- Spatial limits of Lisbon parishes (including river);
- Spatial limits of Lisbon municipality;

[5] Website*:* *http://insideAirbnb.com/index.html*
[6] Website*:* *http://tomslee.net/*
[7] Website: http://geodados.cm-lisboa.pt/

- Spatial coordinates of Lisbon municipality metro stations;
- Spatial coordinates of Lisbon municipality museums;
- Spatial coordinates of Lisbon municipality national monuments.
- Spatial coordinates of Lisbon municipality *tuk tuk* [8] stations.

Airbnb listings data collected from the *Tom Slee* website correspond to 13 variables (quantitative and qualitative) and 13,232 observations. These variables and the respective description are listed in the table below.

Variable	Description
Host_id	A unique number identifying an Airbnb host
Room_id	A unique number identifying an Airbnb listing
Neighborhood	A sub region of the city or search area for which the survey is carried out
Minstay	The minimum stay for a visit, as posted by the host
Price	Price (in €) for a night stay
Room_type	Listing type (entire home/apartment, private room or shared room)
Reviews	Total number of reviews that a listing has received
Satisf	Average rating (out of five) that the listing has received from those visitors who left a review
Capacity	Maximum number of guests a listing can accommodate.
Latitude	Latitude of the listing (decimal)
Longitude	Longitude of the listing (decimal)
Bathrooms	Number of bathrooms offered by a listing
Bedrooms	Number of bedrooms offered by a listing

Table 1 - Original variables

The variables listed in the previous table contain useful information that can enable a better understanding of the Airbnb market in Lisbon. However, this information may not be enough to accurately analyze the market, meaning that there is a need for additional information. Zhang et al. (2017) pointed out some variable which can have an impact on the Airbnb listings' price, such as the distance to the convention city center, highways and local attractions. On the other hand, the number of foursquare check-ins per km^2, score for accessibility to public transportation, number of attractions and entertainment places and other cities attractiveness, were also explored to evaluate their impact on Airbnb listings' price variation (Quattrone et al., 2016). Considering the previous remarks, new variables were created based on the existing ones as shown in the Table 2.

[8] *Tuk tuk* is a different type of vehicles directed to tourist transportation held by a company named *tuk tuk Lisbon*. Website: http://tuk-tuk-lisboa.pt/

Variable	Description
n_list	Total number of Airbnb listings held by each host
CENTER	Distance, in kilometers, from the city center (São Jorge Castle)
nr_metro	Total number of metro station in a ray of 500 meters
nr_mon	Total number of national monuments in a ray of 1 km
nr_tuk	Total number of *tuk tuk* station in a ray of 1 km
nr_mus	Total number of museums in a ray of 1 km

Table 2 – Created variables

Zhang et al. (2017) identified as single listing host, all the hosts that have only one listing registered on Airbnb platform, and multiple listing host as the remaining ones who hold more than one listing. This partition can help identify those individuals who are using the platform only as an alternative to gain extra income (single listing host) and those individuals and professional agencies that are investing in this type of market (multiple listing host). The variable that contains this information was named *n_list* and it was created by counting all *room_id* which have the same *host_id*. The total number of *tuktuk* station, museums, monuments (in a ray of 1 kilometer) and metro station (in a ray of 500 meters) resulted from a joint between the Airbnb listings coordinates (latitude and longitude) and the shapefiles collected from the source *Geodados* (Lisbon city hall). A ray of 1 kilometer and 500 meters were the chosen metrics to create the previous variables since after considering several scenarios, they presented the highest correlation with the dependent variable.

São Jorge castle is a national monument which occupies a privileged area of the old medieval *alcáçova* (citadel) and consists of the castle, ruins of the former royal palace and part of the elite neighborhood[9]. The castle was chosen to be Lisbon city center under this research due the fact that it is located near to the parish with the highest number of listings registered on the Airbnb platform, the *Santa Maria Maior* parish, with 2,798 listings registered in June 2017, and because it is the national monument most visited in Portugal. During 2016, São Jorge Castle was visited by almost 1.8 million people[10]. Additional scenarios were tested, e.g., roundabout of *Marques de Pombal* and *Luís de Camões* square, but end up being discarded due to their low correlation with Airbnb listings' price.

According to data published by the National Statistics Institute (INE Portugal)[11], the highest values for the monthly occupancy rate registered during the year of 2016, took place on July, August and September, with values of 72,6%, 81,5% and 71,5% (per bed), and 84,1%, 89,0% and 89,4% (per bedroom) consecutively. The mentioned months corresponds to the summer season in Portugal, which is one of the main reasons that explains the high numbers of occupancy rate. Due to demand variations, some listings registered on the Airbnb platform are only available during a specific period and not throughout the entire year (Zhang et al., 2017). However, the chosen period for the data collection was based not only on the available data, but also with the aim of capturing a reasonable number of listings to have more representativeness of the population. Since Airbnb listings work based on orders/reserves, and considering July, August and September as being the months in which tourism peaks its highest numbers (occupancy rate), the reference month chosen for the data collection was June 2017.

[9] http://castelodesaojorge.pt/en/history/
[10] https://viagens.sapo.pt/viajar/viajar-portugal/artigos/os-monumentos-mais-visitados-em-portugal
[11] Statistics Portugal website: https://www.ine.pt

Based on Airbnb listings' geographical coordinates combined with Lisbon parishes' spatial coordinates, Figure 1 shows the spatial distribution of listings throughout Lisbon municipality.

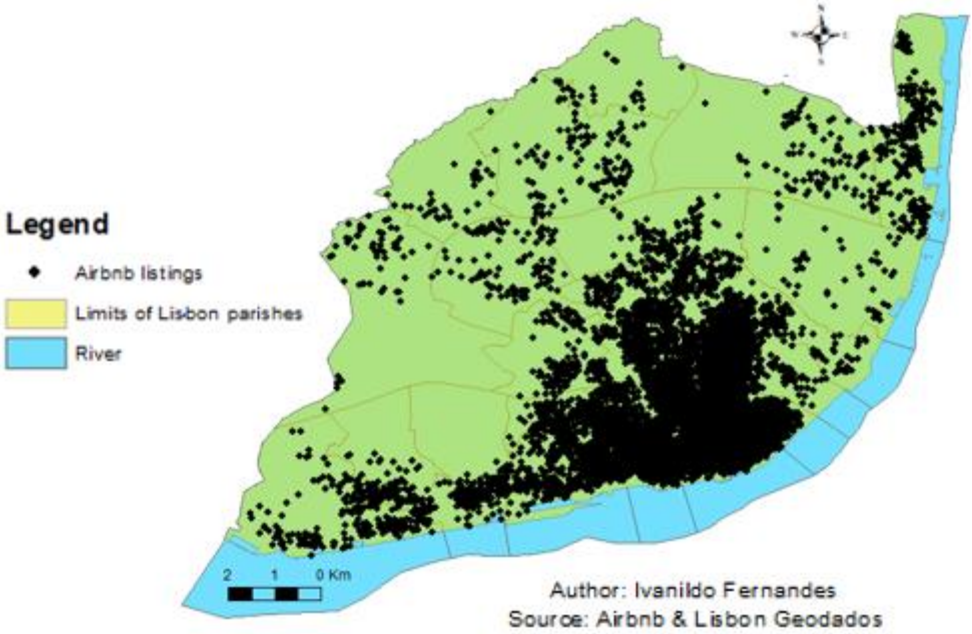

Figure 1 - Spatial distribution of Airbnb listings

Airbnb listings can be found throughout all the area covered by Lisbon municipality, being more concentrated in some parishes than others.

3.2 Data Quality Control

Since Airbnb rarely releases raw data, the accuracy of the data collected from non-official source can only be assessed by comparison made with other studies or with occasional Airbnb public statements. The quality of such data was safeguarded using indicators from *Airdna*[12], *Geodados* and other researches as reference. *Airdna* provides some useful indicators at the city level and its sub regions such as, average prices, occupancy rate and total number of listings for a period of 25 months. The comparison between those indicators and the raw data was crucial to assess its accuracy and representativeness.

The total number of Airbnb listings in November 2017 published on the *Airdna.com* was approximately 12,435 listings, less than 797 when compared to the data from *Tom Slee* website (13,232 listings on June 2017). The difference can be explained by the disparity between the period of data collection. As mentioned before, it is expected a higher number of active listings on *Airbnb.com* during the summer season (July, August and September).

In 2015, 9,273 listings were registered on *Aibnb.com*, with *Santa Maria Maior*, *Misericordia, Santo António*, *Arroios* and *São Vicente* being the top 5 parishes with the highest number of listings in Lisbon municipality (Nova School of Business and Economics & Faculdade de Direito da Universidade

[12] *Airdna* offers analytical tools held by *Airbnb* that enable short-term rental managers, investors and others to have access to data and indicators on Airbnb market of several cities.
Website: https://www.airdna.co/market-data

de Lisboa, 2016). Nonetheless, when looking to data from the source used in this study, the results are similar, i.e., the top 5 remains the same. To be able to compare the total number of listings between the results from the previous research (9,273 listings) and the *Tom Slee* source, data collected from the same reference date are required, since the data being analyzed in this research were collected on June 2017. *Tom slee* website also have available data collected in 2015, which corresponds to 8,969 Airbnb *listings*. The difference between the two sources is not significant (approximately 300 listings) which can be justified by differences in the data collection methods and periodicity.

When combining the Airbnb listings coordinates with the spatial limits of Lisbon municipality, using ArcGIS software, there is a synchronization, i.e., every listing from the data source under study lies inside the area representing Lisbon municipality (Figure 1). Similarly, it was also included the spatial limits of Lisbon parishes, to check whether a listing identified as belonging to a given parish (*Tom Slee* source), lies inside the limits representing the same parish after including the spatial limits of Lisbon parishes. In fact, the data collected from a non-official source was proved to be reliable to be used in the current study, since it does not differ much from the observed reality.

3.3 MODEL ESTIMATION

One of the trends in recent years among the published studies is the increased use of advanced spatial methods to analyze different subjects (L. Krause & Bitter, 2012). According to Dardala & Constantin (2015), location is the key concept to analyze the tourism sector, since it depends mainly on the nature, built and cultural and social characteristics of a given territory.

Several researches made used of geographically weighted regression (GWR) to study different subjects. For instance, Bitter, Mulligan, & Dall'erba (2006) applied GWR regression to examine spatial heterogeneity in housing attribute prices in Pima County, Arizona, United States. It was also proved that GWR methods can result in an improvement, in terms of explanatory power and predictive accuracy, over the global regression models. Yang & Wong (2013) applied spatial data analysis to study the distribution of inbound and domestic tourist flows and growth rates in China cities. Similarly, Martinho (2013) and Vieira & Moreira dos Santos (2016) resorted to spatial analysis to study the tourism sector in Portugal.

Spatial regression is also popular among researchers who seek to study local accommodation. Based on GWR model, Quattrone et al. (2016) concluded that, in London, the most attractive areas, with easy access to public transport, in which its residents are young, employed and born outside the UK, tend to have a higher concentration of properties intended for local accommodation. Combined with socio-economic variables, it was also found that Airbnb listings offering is strongly correlated with the distance to the city center. Based on geolocated big data and geolocated photographs sources, Gutierrez et al. (2016) concluded, by analyzing bivariate spatial correlation between the distribution of Airbnb listings and hotel rooms in Barcelona, that there is a close spatial relationship between Airbnb listings and hotels rooms, i.e., while city's main hotel axis is predominated by Airbnb listings which tend to be concentrated in the city center and also in some residential districts, hotels rooms predominates in some peripherical areas. It was also found that Airbnb can take more advantage of the proximity to the main tourist attractions of the city than the hotel sector. With a different aim, Xu et al. (2017) proved, based on localized regression models (GWR), that there is a positive spatial relationship between the geographical locations of Airbnb listings and incidents of criminal activities.

According to Zhang et al. (2017), regression methods such as OLS and quantile have been frequently used to investigate factors affecting Airbnb listings' prices. However, these methods take no account of location in its analysis, since it explores only the relationships between the dependent variable and the explanatory variables, camouflaging spatial heterogeneity in the relations. Farber & Yeates (2006) also found that when dealing with spatial variation in house prices, geographical weighed regression models produce better results. Space plays no role in the modeling process when using the previous regression methods, i.e., the relationships being measured are assumed to be stationary over space (Brunsdon, Fotheringham, & Charlton, 1998). In cases when these relationships are not stationary, the use of OLS regression method can lead to problems in the interpretation of parameter estimates. GWR model recognize the existence of spatial variations in relationships and provides a way in which they can be measured, allowing the estimated parameters to vary over across regions to accommodate potential spatial dependencies.

Even though it was proved that the GWR regression methods produces better results when analyzing data with spatial patterns, a common approach is to first identify the very best OLS model possible, and then use the same set of explanatory variables in GWR model estimation. This strategy attempts to solve multicollinearity issues among predictors.

3.3.1 Ordinary Least Squares (OLS) regression

The OLS model specification is expressed as:

$$y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + e_i$$

Where

- y_i **-** denotes the *i*th observation of dependent variable;
- $\mathbf{x_i}$ **-** represents the explanatory variables;
- **α** (intercept) - is the predicted value of *y* when all the explanatory variables are equal to 0;
- **β = (β$_1$, β$_2$, . . ., β$_p$) -** are the coefficients of the predictors estimated in the model;
- e_i - is the random error.

Usually real data are not generated by an ideal experiment which can make it difficult to meet the ideal conditions for OLS to provide a good estimate, unbiased and efficient (*Gauss Markov* assumptions), i.e., OLS estimator's alpha and beta are BLUE (best linear unbiased estimator). The ideal conditions for an OLS estimator's to be BLUE are the following:

- Linearity in parameters (alpha and beta), meaning that the model is correctly specified;
- The expected value of the error term is 0;
- Homoskedasticity: the variance of the error term is constant, meaning that the model uncertainty (error variance) is identical across observations;
- The error term is independently distributed and not correlated;
- X_i is deterministic: x is uncorrelated with the error. There is not a relation of collinearity among the X_i (i.e. the explanatory variables should not be correlated with each other).

The violation of the previous assumptions can make the parameters estimates biased and not reliable, e.g., non-linear relationships leads to an inappropriate model, spatial non-stationarity leads to heteroscedastic residuals, spatial autocorrelation of the residuals leads to an unreliable model.

3.3.2 Geographically Weighted Regression (GWR)

The GWR is an extension of the traditional regression equation, e.g., OLS model, where the dependent variable is predicted by a series of explanatory variables in which the estimated parameters can vary over space.

For each location, the GWR model uses a spatial weights matrix to fit a local regression using neighboring observations. The weights matrix is obtained through an adaptive kernel that uses a "near-Gaussian weighting function". The software used (ArcGIS) uses the AICc (Akaike Information Criterion) to determine the optimal number of neighbors in each location, and therefore the bandwidth parameter of that function is variable over space.

The basic GWR model is expressed as:

$$y_i = \alpha(u_i,v_i) + \beta_1(u_i,v_i)x_{i1} + \beta_2(u_i,v_i)x_{i2} + \ldots + \beta_p(u_i,v_i)x_{ip} + e_i$$

where

- y_i – denotes the dependent variable at location *i*;
- x_{ip} – denotes the value of the *Pth* explanatory variable at location *i*;
- **α** (intercept) - is the predicted value of *y* in the location *i*, when all the explanatory variables are equal to 0
- **β = (β1, β2, . . ., β_p) -** are the coefficients of the predictors estimated in the model;
- **(u_i,v_i)** – denotes de coordinates of the location *i;*
- e_i - is the random error at location *i*.

For further details see Fotheringham, Brunsdon, & Charlton (2002).

3.3.3 Model selection

To compare the performance of different OLS models, as well as between the best OLS model and GWR models, the Adjusted R2 or the Corrected Akaike Information Criterion (AICc) were used.

4 RESULTS AND DISCUSSION

4.1 EXPLORATORY DATA ANALYSIS

The main goal of this chapter is to get an insight into the data being used. This approach is the first step to summarize the main characteristics of the data in order to find out what it can tell beyond the formal modeling or hypothesis testing task. Exploratory Data Analysis (EDA) is composed by several techniques that enable the univariate and bivariate data analysis to:

- maximize insight into a data set;
- uncover underlying structure;
- extract important variables;
- detect outliers and anomalies;
- test underlying assumptions;
- develop parsimonious models; and
- determine optimal factor settings.

After performing the EDA, it is expected to have a better understanding of each variable's feature, the correlation between the dependent variable (listings price) and the explanatory variables and among the explanatory variables themselves. In addition, it's also important to identify whether a variable transformation is needed. Regarding the variables' correlation, it's desired to have high correlation values between the dependent and the explanatory variables and low correlation values between the explanatory variables themselves. High correlation values between the explanatory variables can lead to some issues during the model estimation, such as multicollinearity, meaning that the partial regression coefficients may not be estimated precisely, or their signs and magnitudes may be changed. In the presence of multicollinear, a solution can be the removal of one of the explanatory variables highly correlated.

4.1.1 Airbnb Listings' Price (dependent variable)

Airbnb listings' price for a night stay is the variable under study. Initially, the dependent variable unit of measure, after the data collection from *Tom Slee* website, was dollar per unit. However, the first transformation applied was the conversion of the price from dollar ($) to euro (€), since euro is the main currency used for transactions in the study area (Lisbon).

The average listings' price for a night stay in Lisbon is approximately 74€ (Table A1 - appendix). However, there is a large price variability which can go from a minimum of 9€ to a maximum of 6,378€. When looking to each listing type, separately, Airbnb listings classified as entire home/apartment are the most frequent (9,764 listings), meaning that this type of listing is the one that contributes the most for the large price variability, since its maximum and minimum price values are the same as when analysing the listings as whole. On the other hand, private and shared room represent 25% and 1,3%, consecutively, of the overall listings registered on Airbnb platform within Lisbon market. Due to their similarities, the two types of listings (private and shared room) were considered as one listing type, private/shared rooms.

To assess the listings' price distribution, two histograms were computed based on the 13,232 observations (Figure A3 – appendix). The first histogram (Price in euros) shows a strong skewed (non-symmetric) distribution to the left and some outliers, indicating that transformations may be needed to avoid bias in the estimated models' results. One approach that can help mitigate this problem is to calculate the logarithm of the variable price. After transforming the dependent variable (*l_price*), the results became better, as shown in the second histogram (Logarithm of Price). The price distribution is no longer skewed to the left, becoming close to the normal distribution. Still, the boxplot under the histogram shows that outliers remained unchanged.

As mentioned in previous chapters, Airbnb's hosts are free to set the desired price to be charged in exchange of sharing a property, which can lead to a huge price variation among its listings and, consequently, the appearance of outliers. After considering a few scenarios, it was decided to stablish a maximum listings' price of 500€ by removing all the observations with a price greater than 500€, which lead to the removal of 79 observations (less than 1% of the total). When looking to the removed observations, separately, the mean and the standard deviation are approximately 1,113€ and 821€ consecutively, very different from those when considering the entire dataset (74€ and 115€). In fact, after removing all the observations considered as being outliers and 10 additional observations, whose spatial coordinates (latitude and longitude) lies inside the river[13], the remaining observations correspond to a total of 13,143 listings with mean equal to 67,4€ and a standard deviation of 53€ (Table A2 - appendix). There was a significant improvement after transforming the dependent variable and removing potential outliers by considering only listings with a price lower or equal than 500€.

Location is a commonly accepted factor proven to effect Airbnb listing prices (Zhang et al., 2017). Spatial autocorrelation is an important factor to have in mind when estimating OLS models, since if there is evidence regarding spatial correlation among the data, the observations are not considered to be independent from one another. A commonly used statistic that describes spatial autocorrelation is *Moran's I*. It evaluates whether the pattern expressed is clustered, dispersed, or random. By looking to the *z-score* and *p-value* from the *Moran's I* summary results regarding the dependent variable (Table A7 – appendix), the indication is to reject the null hypothesis (Airbnb listing prices are randomly distributed across Lisbon area), i.e., there is spatial autocorrelation among the dependent variable. Another important conclusion taken from the *Moran's* summary results is the positive *Moran's I* index value, which indicates tendency toward clustering.

Based on the previous remarks, there is possibility of having clusters of neighborhoods composed by Airbnb listings with high and/or low prices. Hot spot analysis[14] is commonly used to help identifying where features (Airbnb listings) with either high or low prices cluster spatially. Airbnb listings with high price values surrounded by other Airbnb listings with high price values are considered statistically significant (hot spot). On the other, the cold spot is also statistically significant, and it represents Airbnb listings with low price values surrounded by other Airbnb listings with low price values.

[13] Boats used as local accommodation, which is an Airbnb listing type not considered in this study due to its features and total number of observations.

[14] Hot Spot Analysis interpretation: *http://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm*

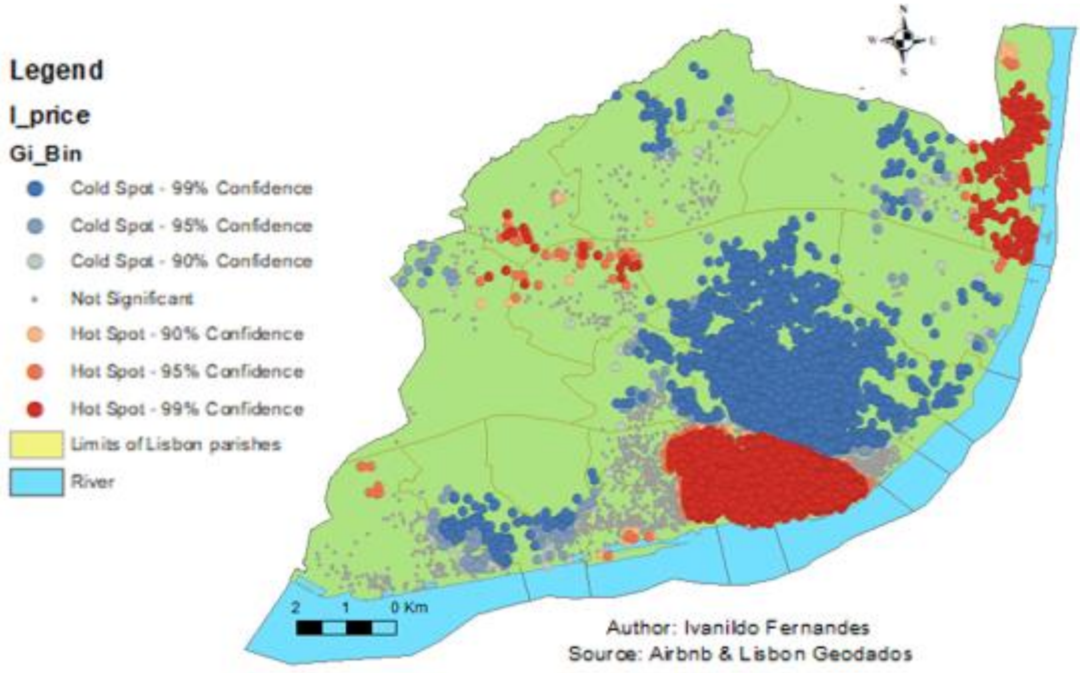

Figure 2 - Hot Spot Analysis (l_price variable)

Considering the hot spot analysis results (Figure 2), parishes, such as *Parque das Nações*, *Santa Maria Maior*, *Misericórdia* and *Santo António*, are statistically significant hot spot, i.e., composed mainly by Airbnb listings with high price values. In addition, statistically significant cold spot is composed by *Avenidas Novas*, *Areeiro*, *Arroios*, *Penha de França*, *Alvalade* and *Campolide* parishes. Spatial non-stationarity arises from those local differences, thus a GWR model is more appropriate than a global model, such as the OLS model, as expected.

4.1.2 Explanatory Variables

Besides the fact that the Airbnb listings are present throughout all the parishes of Lisbon municipality, there are parishes with a strong listing's concentration, such as *Arroios*, *Misericordia*, *Santa Maria Maior*, *Santo António* and *São Vicente*, which have more than 1000 listings (Figure 3). However, the total number of listings in parishes like *Benfica*, *Beato*, *Carnide*, *Marvila* and *Santa Clara* do not exceed 100 listing.

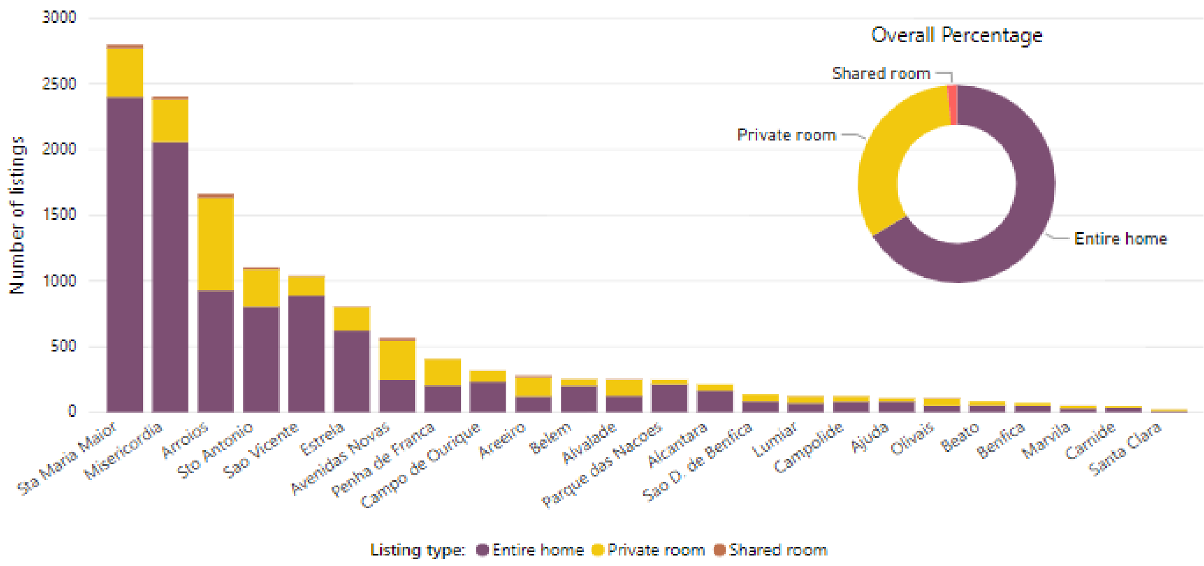

Figure 3 - Airbnb listings in each parish grouped by listing type

According to Ke (2017), 68.5% of the listings in more than 190 countries are classified as entire home/apartments, 29.8% as private rooms and the remaining 1.7% are shared rooms. Regarding Figure 3, it seems that Airbnb market, considering only Lisbon's scenario, reflects a similar pattern in which most of the listings are classified as entire home/apartment (73.7%), followed by private room (24.9%) and shared room (1.3%). Although, parishes such as *Areeiro*, *Avenidas Novas* and *Olivais* are the only ones where the majority of Airbnb listings are classified as private room instead of entire home/apartment. Airbnb listings classified as shared room is exclusive of some parishes, since this type of listing is not present in every Lisbon parishes as shown in Figure 3. Based on Airbnb listings' geographical coordinates combined with Lisbon parishes' spatial coordinates, a map was previously created to better understand the spatial distribution of listings throughout Lisbon municipality (Figure 1). Airbnb listings are most concentrated in the south of Lisbon, which corresponds to *Santa Maria Maior*, *Misericórdia*, *Arroios* and *Santo António* parishes. There are two huge areas represented on the map without any listing inside. These areas correspond to Lisbon airport, on the north, and Monsanto (diverse forest) on the south-west.

Figure 4 contains the spatial distribution of Airbnb listings classified as entire home/apartment (*Home*) and private shared room (*Room*). Despite the difference between the two types of listings regarding the number of observations and price, it seems that they follow the same spatial patterns. Both type of listings can be found throughout all Lisbon parishes and are concentrated mostly in *Misericórdia*, *Santa Maria Maior*, *Arroios*, *Santo António* e *São Vicente* parishes, as shown in the Figure 4.

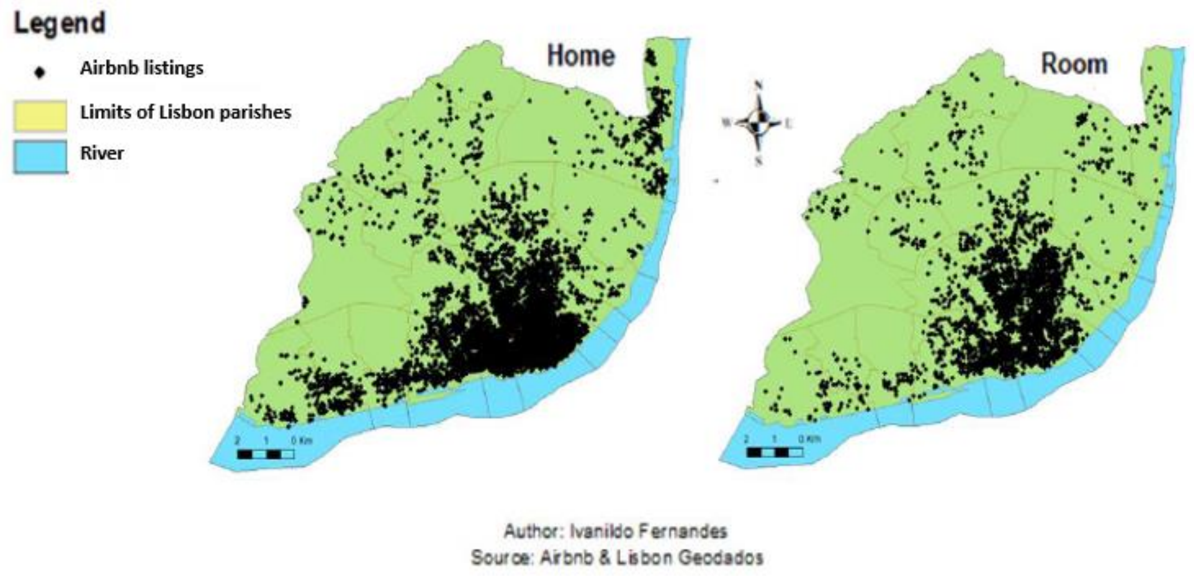

Figure 4 - Airbnb listings spatial distribution per type

Another important insight about the data, is related to the number of listings held by each host (*n_list* variable). In this sense, hosts were classified as single-listing host (hosts that hold only 1 listing) and multi-listing host (hosts that hold more than 1 listing), which varies between a minimum of 1 and a maximum of 171 listings per host. On average, each host has 10 listings registered on Airbnb platform (Table A2 - appendix). These values led to the conclusion that Airbnb market in Lisbon is not explored only by the individuals who are looking for an extra income, but also by specialized agencies. However, the previous results are different from those when analyzing data from Portugal as a whole. The research made by Nova School of Business and Economics & Faculdade de Direito da Universidade de Lisboa (2016), quantified that 81.2% of Airbnb listings in Portugal are held by single listing host.

Airbnb is not the only source providing local accommodation data regarding Lisbon market. *Tourism Portugal* platform makes available centralized information concerning the tourism enterprises and companies operating in Portugal, including RNAL data[15]. Up to June 2017, there was 7,537 local accommodation records classified as apartments registered on RNT, 5,695 records less than Airbnb listings' data collected from *Tom Slee* website. Besides the large difference between the number of records of the two sources, they follow a similar pattern as shown in the Figure 5. As mentioned in previous chapters, the reason why RNAL has lower local accommodation's data is due to the large number of properties rented without any formal agreement.

[15] RNAL data: https://rnt.turismodeportugal.pt/RNAL/ConsultaRegisto.aspx?Origem=CP&FiltroVisivel=True

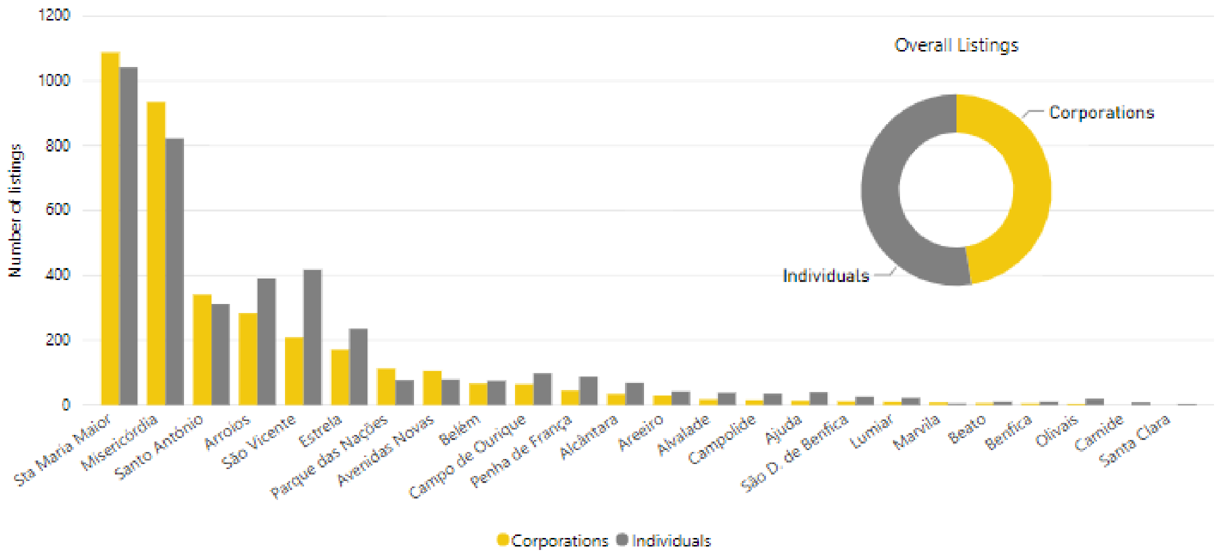

Figure 5 - RNAL – Airbnb listings in each parish grouped by host type

By comparing the information represented in the previous chart (Figure 3) and the chart built based on RNAL's data (Figure 5), it's remarkable that, in both charts, the raking of parishes by the number of listings are very similar. RNAL data come to support the idea that local accommodation market is also being dominated by specialized agencies. In fact, considering only Lisbon municipality, 47,4% of RNAL records are held by corporations. From the Figure 5, it can be noticed that the number of local accommodations held by individuals and corporations are very similar in almost every parish, except for *Arroios*, *São Vicente* and *Estrela* where the individuals' dominance is notable.

Local accommodation market attracts both national and foreign investors. RNAL data also revealed that only 3% of the overall records are held by foreign investors. However, these 3% represents individuals/corporations prevenient from more than 40 different countries. France leads the ranking with 74 records (Figure A4 - appendix), followed by Brazil (21 records) and Switzerland (16 records). Since more than half of Lisbon's local accommodations properties are not included in RNAL's data, it is difficult to measure the exact number of foreign investors operating in the market.

The histograms represented in the scatterplot's matrix (Figure A1 and A2 - appendix), enable the assess to the variables' individual distribution. Analogous to the conclusions drawn from the analysis of the last histogram (Figure A3 – appendix), the dependent variable continues to follow a normal distribution. On the other hand, the explanatory variables *n_list* and *reviews* show a strong skewed (non-symmetric) distribution to the left. It's worth to mention that the variable price had the same structure as these two explanatory variables, which was solved by applying some transformations (logarithm of price). That said, computing the logarithm of some of these explanatory variables can be a solution to be tested and evaluate whether transforming a variable can help improving models' predictive capacity.

Due to the high number of missing values, the variables *minstay* and *bathrooms* were excluded from the analysis.

Following the same approach used when analyzing the dependent variable, host spot analysis for the explanatory variables was also computed (Figure A5 to A7 – appendix). From the hot spot analysis outputs, a summary table containing the important results was created (Table 3). As shown in the table below, for each explanatory variable, it was identified those parishes with large concentration of statistically significant hot/cold spot features.

	Lisbon Parishes	
Variables	**Hot Spot**	**Cold Spot**
bedrooms	Arroios; Misericórdia; Santo António	Estrela; São Vicente; Santa Maria Maior
capacity	Misericórdia; Santa Maria Maior; Santo António	Alvalade; Avenidas Novas; Areeiro; Arroios; Penha de França
n_list	Avenidas Novas; Estrela; Misericórdia; Santa Maria Maior	Alcântara; Penha de França; São Vicente; Arroios
reviews	Misericórdia; Santa Maria Maior; São Vicente; Estrela	All the remaining parishes
satisf	Estrela; Misericórdia; Santa Maria Maior; São Vicente; Santo António	Alvalade; Arroios; Areeiro; Avenidas Novas; Campolide; Campo de Ourique; Carnide; Lumiar; Parque das Nações; São Domingos de Benfica; Benfica
type	Alcântara; Estrela; Misericórdia; Parque das Nações; Santa Maria Maior; São Vicente; Santo António	Areeiro; Arroios; Avenidas Novas; Campolide; Lumiar; Olivais; Penha de França; São Domingos de Benfica

Table 3 - Hot Spot Analysis (Summary)

Misericórdia is the only Lisbon parish which has a largest concentration of statistically significant hot spot analysis among the explanatory variables. In fact, Airbnb listings located within *Misericórdia* parish are mostly entire home/apartment held by multi-listing hosts, with high number of bedrooms, which allocate more guests, receive more reviews and have better ratings.

Dogru & Pekin (2015), considered the number of reviews as being the number of times a given Airbnb listing was rented, since only guests who stayed in a property are allowed to provide reviews. In this sense, hosts' revenue can be estimated by multiplying the number of reviews by the price. Considering the variable *reviews* as being a proxy of the number of times a listing was rented, the hot spot analysis results presented in Table 3, shows that there are only 4 parishes receiving a high number of visitors in the largest part of their comprehensive area (*Misericórdia*, *Santa Maria Maior*, *São Vicente* and *Santo António*). In fact, these 4 parishes were found to have the highest concentration of Airbnb listings (more than 1000). All the remaining parishes presented a statistically significant cold spot, meaning that most of the listings within those parishes receive lower number of reviews (guest). In addition, Table A1 (appendix) shows that the average number of reviews received by listings classified as entire home/apartment is approximately 32.9 reviews, while private and shared rooms received on average 19 and 9.7 reviews respectively. Following the previous thoughts regarding the meaning of the variable *reviews*, it can be said that entire home/apartment is the most demanded listings type from Airbnb platform in Lisbon.

Based on the results for the *type* variable, it was possible to identify parishes dominated by listings classified as entire home/apartment (hot spot) and those by listings classified as private/shared room (cold spot). Another important remark regarding the hot spot analysis results, is the group of parishes dominated by multi-listing hosts (hot spot) and single-listing hosts (cold spot) that can be seen in the row corresponding to the *n_list* variable.

4.1.3 Bivariate Description

This section is reserved for the analysis of the relationship between the dependent variable and the explanatory variables, and between the explanatory variables themselves.

The scatter plot matrix (Figure A1 - appendix) shows some pattern when looking to the plots that represent the relationship between the dependent variable and a few explanatory variables, such as *bedrooms* and *capacity*. It seems that Airbnb listings' price increase with an increase in the number of *bedrooms* and with the overall capacity presented on the platform. On the other hand, there are also strong patterns in some plots among the explanatory variables themselves (Figure A2 - appendix). The variables *nr_mon*, *nr_mus* and *nr_tuk* seems to be a perfect combination of one another, one increase with an increase in the other one. These variables cannot be all included in the same model to be estimated, so that multicollinearity issues can be avoided.

In addition to the previous findings, the *Pearson Correlation Matrix* was created (Table A3 - appendix). High correlation values were found between the dependent variable (*l_price*) and some explanatory variables, such as *bedroom* and *capacity*. Regarding the explanatory variables' correlation among themselves, the variables *center*, *nr_mon*, *nr_mus* and *nr_tuk* are highly correlated with one another as well as *bedroom* and *capacity*. Both scatterplot and the correlation metrics led to similar conclusions. Besides the high correlation value between the variable *bedrooms* and *capacity*, which was expected since the number of bedrooms is one of the main determinants of a listing global capacity, it was decided to include both in the model estimation. Apparently, when using *Airbnb.com* it was noticed that some listings with the same number of bedrooms can have different capacity level. This situation occurs in cases when, for instance, hosts offered couches as additional beds to increase the overall capacity of a listing.

4.2 OLS MODELS

With the aim of finding whether the candidate explanatory variables yield any properly specified OLS model, exploratory regressions were performed, which evaluates all possible combinations of explanatory variables, by looking for the OLS models that best explain the Airbnb listings' price. The variable *Room_type* was used to create a new dummy variable (*type*), which takes the value 1 if the Airbnb listing is classified as entire home/apartment, or 0 if it's a private/shared room.

Considering all observations (*All data*), the generated report suggested three similar models with 5 explanatory variables each (Table A4 – appendix). The only difference between the three-suggested model is regarding the 5th explanatory variable, that can be *nr_mus*[16] (first model), *nr_tuk*[17] (second model) or *nr_mon*[18] (third model). As mentioned in the previous chapter, there is a high correlation between the variables *nr_mus*, *nr_tuk* and *nr_mus*, which can explain the fact that they were included in separated models to avoid multicollinearity problems.

Based on measure of goodness of fit (AIC – Akaike information criterion), it was concluded that the best model among the three options is the first model, since it has the lowest AIC's value. This model is composed by 5 explanatory variables (c*apacity*, *bedrooms*, *satisfy*, *type* and *nr_mus*), explaining

---

[16] Total number of museums in a ray of 1 km
[17] Total number of tuk tuk station in a ray of 1 km
[18] Total number of museums in a ray of 1 km

52% of the Airbnb listings' price variation within Lisbon market. Even though the previous model (initial model) predicted more than 50% of the dependent variable's variability, some related problems were found, e.g., residual non-normality.

Considering the previous model (lowest AIC value) as being the initial model, further actions were taken intended to find the OLS model that better explains Airbnb listings' price variations. In this sense, additional explanatory variables were included in the model and some transformations were applied, as shown in the first 8 rows of the Table A5 (appendix), where the column *Base model* is named as "All data". Each row contains important information regarding the tested models.

After an initial analysis, it seems that it is not worth to include the variables *center*, *nr_metro* and *n_list* in the final OLS model, since the percentage of variation explained by the 5 explanatory variables (R^2 and adjusted R^2) from the initial model remains unchanged after adding into the model each one of the previous variables separately. Furthermore, the AIC's values from these tested models are almost equal or even greater than the AIC's values from the initial model. Among all possible scenario (tested models), the best OLS model, with the highest adjusted R^2 and the lowest AIC's value, is the model in which the variable *satisf* was transformed (*l_satisf*) and two more explanatory variables were added to the initial model (*center* and the logarithm of *reviews*). Although it was decided to not include the variable *center* in the model, it was noticed that after adding the variable *l_reviews*, the adjusted R^2 also increased with the inclusion of the variable *center*.

The final model specification (model 1) together with the results of the estimated model are presented below.

Model 1 (*All data*): $l\_price = \alpha + \beta_1 capacity + \beta_2 bedrooms + \beta_3 type + \beta_4 center + \beta_5 nr\_mus + \beta_6 l\_reviews + \beta_7 l\_satisf$

Variable	*Intercept*	*Capacity*	*Bedrooms*	*Type*	*Center*	*Nr_mus*	*l_Reviews*	*l_Satisf*
Coefficient	3.1900	0.0627	0.1304	0.5544	0.0169	0.0147	-0.0668	-0.0592
StdError	0.0144	0.0027	0.0057	0.0096	0.0027	0.0006	0.0038	0.0089
t-value	220.46	22.43	22.82	57.44	6.09	21.9	-17.46	-6.60
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 4 – Final OLS model (All data)

All the coefficient from the OLS model (Table 4) are statistically significant ($p < 0.001$) and the R^2 is approximately 54% (Figure A8 - appendix), 2 percentage points more when compared to the initial model.

In the OLS models, the dependent variable is in its log-transformed state, and the explanatory variables are in their original metric. Hence, to interpret the percent of change in the price, first it was necessary to exponentiate the explanatory variable's coefficient, subtract one from this number, and then multiply by 100.

On average, Airbnb listings classified as entire home/apartment are 74% more expensive than those classified as private/shared room, *ceteris paribus*. When looking to the coefficients of the explanatory variables, such as *capacity* and *bedrooms*, which were found to be high correlated, they both have a positive impact on the price variation but with different weights. By looking to the coefficient signs throughout the whole model, only the variables *l_reviews* and *l_satisf* have a negative sign. Airbnb listings with high number of reviews and better ratings are likely to have lower prices.

Despite all the effort done within the scope of finding the best OLS model, which comprised inclusion/exclusion of explanatory variables as well as some related transformations, the problems associated with the estimated models remained unchanged. OLS model diagnostics results (Figure A8 – appendix) shows that the *Koenker* (BP) Statistic test is statistically significant (*), meaning that the relationships modelled are not consistent (either due to non-stationarity or heteroskedasticity). In addition, as previously mentioned, the initial model presented problems concerned with residuals normality, which is still an issue in the final estimated model (*Jarque-Bera* Statistic test). When the residuals are not normally distributed, it is an alert informing that the model predictions are biased. However, the overall model is statistically significant (*Joint F-Statistic*).

According to Zhang et al. (2017), including the variable *type* in the model does not bring any additional knowledge, since its logical that prices for Airbnb listings classified as entire home/apartment are always higher than private/shared room due to its higher capacity to accommodate more guests. However, there is no prove that a unique model can explain in the same way the price variability for both types of listings. One of the main purposes of this research is to evaluate whether Airbnb listings can be analysed as a whole or separated by listings' type. For this reason, the data was separated by the type of listing, entire home/apartment (9,679 listings) and private shared room (3,464 listings). Then, three models were estimated based on the same explanatory variables (Table A6 - appendix). First, using the entire dataset (*All data*), second using data regarding listings classified as entire home/apartment (*Home*) and the third one using the remaining data (*Room*). For these new models, the variable *type* was excluded, because the data was already separated by listing type.

The R^2 for the first model (*All data*) is 42%, 12 percentage points less than when considering variable *type* in the model. In fact, the listing type has a huge impact on the Airbnb listings' price variation. As expected, the results from the second model (*Home*) are very similar to those from the first model, since entire home/apartment represents approximately 74% of the total observations. On the other hand, the explanatory variables used in the initial model did not performed well when using private/shared room data, as shown in third model's results (*Room*). The model explains only 12% of the price variations for Airbnb listings classified as private/shared room. Another important remark is that the variable *bedrooms* is not statistically significant in the third model. An explanation could be the fact that the number of bedrooms does not bring additional benefit for a guest who is interested in a private/shared room. As suspected, there are evidences that Airbnb listings' price for different categories (*Home* and *Room*) are not influenced by the same factors. The explanatory variables used in the model 1 (*All data*) explains better the price variability of entire home/apartments than private/shared room. It is worth to mention that the *Home* model presented better results not only when compared to *Room* model, but also in comparison with the first model (*All data*).

Once again, exploratory regressions were computed with the aim of finding whether the candidate explanatory variables yield any properly specified OLS model for the two types of listing. While the suggested OLS model for entire home/apartment is composed by the same explanatory variables as those from the initial model (Table A4 – appendix), there were some explanatory variables exclusion and inclusion in the OLS model for the private/shared room. For instance, the number of listings held by each host was found to have a significant impact on the private/shared room price variation (*n_list*), and the variable *bedrooms* was excluded.

In terms of R^2, there were no significant changes when compared to previous estimations, in which the initial model was applied to the two datasets. For the entire home/apartment observations, the OLS model (*Home*) suggested by the explanatory regression explains 42% of the price variations, while with the *Room* observations, it is only 12%.

It has been seen that when building a model for the two datasets (*Home* and *Room*) some explanatory variables that were left out may be significant to explain the different type of listings' price, or those that are already in the model may have to be excluded and additional transformations may be required. In this sense, similarly to what was done to find the initial model, several potential models were estimated in which variables were tested, as well as some transformations, to find the model that best fits each type of data (Table A5 - appendix). For the Airbnb listings classified as entire home/apartment, the best model is very similar to the initial model, with the exception that the total number of metro station in a ray of 500 meters (*nr_metro*) were added to the model. The model is identified in the Table A5 (appendix), where the column *Base model* is identified as *Home (#1)* and the column *ID* is equal to 8. The model specification and results are the following:

Model 2 (*Home*): $l\_price = \alpha + \beta_1 capcity + \beta_2 bedrooms + \beta_3 center + \beta_4 nr\_metro + \beta_5 nr\_mus + \beta_6 l\_reviews + \beta_7 satisf$

Variable	*Intercept*	*Capacity*	*Bedrooms*	*Center*	*Nr_metro*	*Nr_mus*	*l_Reviews*	*l_Satisf*
Coefficient	3.7108	0.0693	0.1252	0.0144	0.026	0.01257	-0.0759	-0.0254
StdError	0.0169	0.0029	0.0058	0.0029	0.0041	0.0007	0.0038	0.0093
t-value	218.57	23.45	21.39	4.89	6.34	17.47	-19.79	-2.71
p-value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0067

Table 5 – Final OLS model (Home)

In fact, the results are close to those obtained in the first estimated model (*All data*), where the only difference is regarding the inclusion of the variable *nr_metro*. According to results presented in the Table 5, having a metro station in a ray less or equal than 500 meters, can increase the average price for listings classified as entire/home apartments in 2.6%, *ceteris paribus*.

For the Airbnb listings classified as private/shared room, the final model is identified in the Table A5 (appendix), where the column *Base model* is identified as *Room (#1)* and the *ID* is equal to 5. The model specification and additional details are the following:

Model 3 (*Room*) – $l\_price = \alpha + \beta_1 capacity + \beta_2 n\_list + \beta_3 center + \beta_4 nr\_mus + \beta_5 l\_reviews + \beta_6 l\_satisf$

Variable	*Intercept*	*Capacity*	*N_list*	*Center*	*Nr_mus*	*l_Reviews*	*l_Satisf*
Coefficient	3.3910	0.0321	-0.0012	0.0263	0.0187	-0.0364	-0.1517
StdError	0.0321	0.0064	0.0005	0.0063	0.0015	0.0102	0.0218
t-value	105.40	4.46	-2.11	4.16	12	-3.54	-6.95
p-value	0.0000	0.0044	0.0012	0.0003	0.0000	0.0000	0.0000

Table 6 – Final OLS model (Room)

Besides the fact that the percentage of variation explained by the final *Room* model is higher than the one from the model suggested in the exploratory regression, the R^2 value is still very small compared to the other models, stating at approximately 13% (Table A5 – appendix). In terms of explanatory variables, the number of bedrooms was excluded from the model due to non-significance, while the number of listings held by each host (*n_list*) was found to have an impact on the listings' price variation. Contrary to what was expected, the variable *n_list* has a negative impact on the listings price, meaning that multi-listing hosts set a lower price than single listing-hosts. As mentioned in other chapters, Airbnb market in Lisbon is also strongly controlled by corporates (specialized agencies), that are more aware in what concerns the competitive prices. On the other hand, hosts who held only one listing are usually those individuals looking for an extra income who may not be interested on applying competitive prices if the listing is being profitable. When looking to the common explanatory variables between *Home* and *Room* models, there is no significant changes in terms of coefficients' weight or sign.

For each model, a residual plot versus the predicted dependent variable was created as shown in the Figure A9 (appendix). The plot for the first model (*All data*) look as if there is a structure, suggesting the existence of dependency between the residuals and the fitted values. However, in the second model (*Home*), in which all the Airbnb listings classified as private/shared room were removed from the database, the *residuals vs fit* plot seems to have less structure than the previous plot. The last plot (*Room*) appears to be part of the existent structure in the first plot (*All data*). As mentioned before, the available explanatory variables used to estimate the OLS model, explains better the entire home/apartment price variability rather than private/shared room, which can be the reason why there is still a huge portion of variability not explained by the model as shown in the third plot (*Room*).

4.3 GWR MODEL

As previously mentioned, OLS models showed some problems regarding the model assumptions. Since non-spatial statistical methods were used to analyze spatial data, it was expected that some problems might arise, e.g., non-stationarity/heteroskedasticity, residuals normality and so on. The goodness-of-fit (indicative of how well the estimated values correspond to those observed) of the models using GWR are better than those using GOLS (global regression models), such as ordinary least squares (Dogru & Pekin, 2015). One advantage of GWR regression model is that the spatial patterns inherent in the parameter estimates can be easily mapped and visualized (Bitter et al., 2006). In conclusion, GWR models may be regarded as the one which accounts best for the spatial variation in listings' prices. However, the estimated OLS models are also important to select the potential explanatory variables to be used in the GWR model, as well as to obtain insights of their impacts on the dependent variables. In addition, OLS model can also be used as reference model to be compared with the GWR model, which will enable to measure improvements.

The initial idea was to use the same explanatory variables from the final OLS models (*All data*, *Home* and *Room*) in the estimation of the GWR models. However, due to multicollinearity issues the variable *nr_mus* had to be excluded from the models. With the aim of replacing the excluded variable (*nr_mus*), variables, such as *nr_mon* and *nr_tuk*, which were found to be high correlated with *nr_mus,* were used to test new GWR models. However, neither *nr_mon* nor *nr_tuk* were able to pass through the GWR model estimation without showing multicollinearity with the remaining explanatory variables. The decision was to proceed with the GWR models' estimation, without including the explanatory variables *nr_mus, nr_mon* and *nr_tuk*.

Although the initial idea was to be based on OLS models' explanatory variables to estimate the GWR models, some changes had to be done in order to find the best GWR, considering the available explanatory variables. Based on the same approach used in OLS models' estimation, several GWR candidate models were tested and the final models are presented in the table below.

Data	Explanatory variables	Model	R2	AdjR2	AICc
Room	*Capacity; Center; l_Reviews; l_Satisf; nr_metro; n_list*	GWR	0,15	0,14	5150,11
		OLS	0,09	0,09	5353,05
Home	*Capacity; Bedrooms; l_Reviews; l_Satisf; n_list; l_Center*	GWR	0,46	0,46	7683,42
		OLS	0,41	0,41	8355,44
All Data	*Capacity; Bedrooms; l_Reviews; l_Satisf; n_list; l_Center*	GWR	0,45	0,45	16483,6
		OLS	0,39	0,39	17674,3

Table 7 - GWR vs OLS models' results

Since the GWR models do not have the same explanatory variables used to estimate the OLS models, new OLS models were estimated based on GWR models' explanatory variables, so that comparisons can be made. As expected, all the GWR models presented better results when compared to the OLS models (Table 7). The percentage of room (private/shared) price variation explained by the model continued to the very low (R2=15%). On the other hand, despite the exclusion of the variable *nr_mus,* which was found to have a huge impact on the listings' price variation, when estimating OLS models, the GWR models still presented better results than OLS models. It's also worth to mention that in both regression methods (GWR and OLS), the available explanatory variables used, explains better the price variation of the Airbnb listings classified as entire/home apartment than private/shared room, or even when both type of listings are included in the same database (*All data*).

When using GWR models, the interest is focus mainly in examining how the relationships between the Airbnb listings' price and each explanatory variable change across Lisbon municipality. However, answers to the previous question cannot be given only by looking to the global models' results (Table 7). It's important to spatially analyze some models' indicators, e.g., how well the local regression fits the observed dependent values (local R2), evaluates local collinearity (condition number), the reliability of estimated coefficients' values by looking to their respective standard errors and so on.
As concluded before, in both regression methodologies, the estimated models for *Home* and *All data* always showed similar results, which make sense since more than 70% of the observations correspond to entire home/apartment. For this reason, it seems to be more efficient to compare only the GWR models' results for the two types of listings, private/shared room and entire home/apartment.

First, the local R^2 maps were analyzed to evaluate in which parishes the GWR models are poorly or well performing. Figure 6 contains the spatial distribution of the local R^2 results for the entire home/apartment (*Home*) and private shared room (*Room*) GWR models.

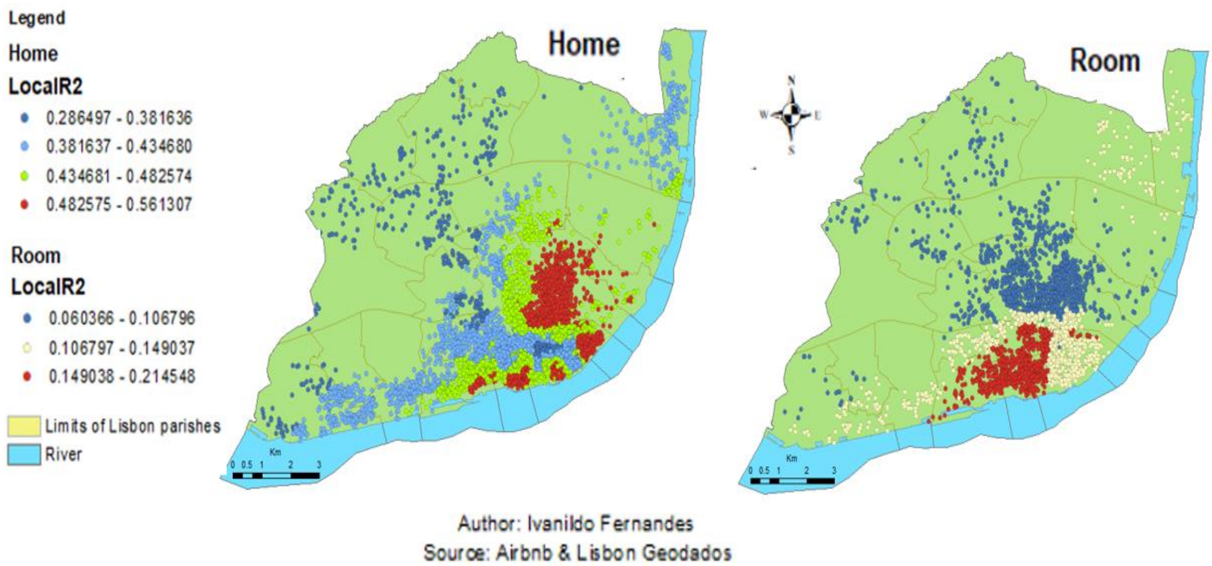

Figure 6 – Spatial distribution of local R^2 for Home and Room models

Along with the conclusions from the global R^2 results (Table 7), the local R^2 for the two types of listings continued to be very different, varying between 6% and 21% in *Room* model, and between 26% and 56% in *Home* model. The differences are not concerned only with the coefficients' weight, but also in terms of the ranking. While *Arroios*, *Areeiro*, *Penha de França*, *Beato* and *Avenidas Novas* parishes have the highest local R^2 values in the *Home* model (43% - 56%), when looking into the *Room* model's results, those parishes show the lowest values (6% - 14%). On the other hand, *Estrela*, *Misericórdia*, *Santo António* and *Santa Maria Maior* parishes appeared in both GWR models with high R^2 values.

Second, maps for the coefficient estimates were created to enable assess to a more detailed information regarding their signs and magnitude of impact (weights) on the Airbnb listings' price in each parish across Lisbon municipality. By this stage, it's clear that the available explanatory variables used to estimate both, GWR and OLS models, do not explain in the same way the price variability of the two types of Airbnb listings. For OLS models, all the common explanatory variables used in the two estimated models (*Home* and *Room*) showed the same coefficient sign. Similar to the conclusion drawn from the local R^2 analysis, there are evidences that the coefficients' estimates can also present different values and signs for the same explanatory variable in the two GWR models, depending on the location.

Figure 7 contains the visual a representation of the spatial distribution of the coefficients' estimates (*coeff*), along with the respective standard errors (*StdError*) for the variable *capacity* in the *Home* model.

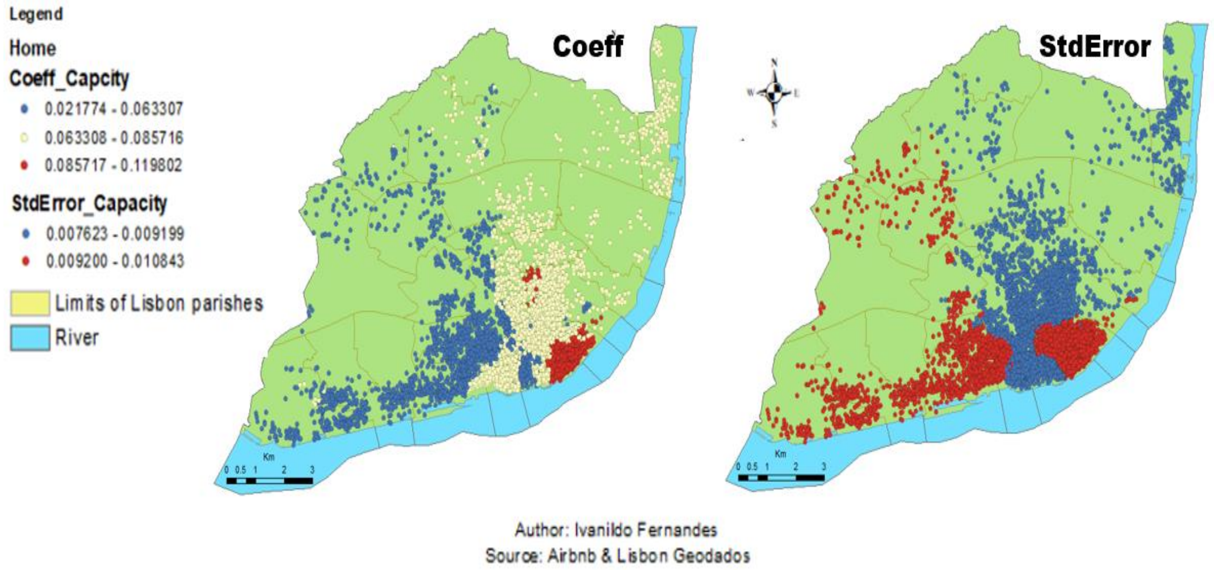

Figure 7 – Spatial distribution of coefficients and Standard Error (capacity variable) of the Home model

It seems that the listings' capacity has more impact on the price variation for Airbnb listings located in parishes such as *São Vicente*, *Santa Maria Maior*, *Arroios*, *Areeiro* and *Penha de França*. However, the same explanatory variable has less impact on listings' price variation in *Carnide*, *Benfica*, *Alcântara*, *Estrela*, *Ajuda* and *São Domingos de Benfica* parishes. With the aim of measuring the reliability of the those estimates, map for the standard errors (*StdError*) was also created as shown in the right side of the Figure 7. All the standard errors are near zero and below the coefficients' estimates, meaning that those estimates from the variable *capacity* are reliable.

When looking into the results produced by the *Room* model (Figure A14 – appendix), the conclusion is identical, the coefficients' estimates are also reliable. Even though the coefficients' estimates were found to be reliable in both models, there are some others notable differences. For instance, in the *Home* model, the variable *capacity* was found to have positive coefficients' sign in every Lisbon parishes (Figure 7), while for the *Room* model, the same variable can show different sign (positive or negative), depending on the parish (Figure A14 – appendix). For the *home* model, the coefficients' values for the variable *capacity* vary between 0.02 and 0.12, although for the *room* model they vary between -0.03 and 0.12.

In fact, the listings' capacity can play different role depending on the listings' type. For entire home/apartment, the more the listings' capacity increases, the more the guest get benefited. On the other hand, an increase in the listings' capacity of a shared/private room, only means that guests will have to share the property with more guests, which can explain the negative sigh for the variable *capacity* in the *Room* model for some parishes. In addition, variables such as *l_reviews* (Figure A12 and A16 - appendix) and *l_satisf* (Figure A13 and A17 - appendix) also showed different coefficients' sign in each GWR models. Only the explanatory variable *n_list* (Figure A20 and A18 - appendix) and *l_center*/*center* (Figure A10 and *A15* - appendix) showed the same coefficient signs in both GWR models.

The coefficients' estimates related to each explanatory variable were also found to be different throughout Lisbon municipality, having more impact the price variability of listings located in specific parishes. There are also differences when looking into the common variables' coefficients estimates for *Home* and *Room* models. The variables *capacity* and *bedrooms*, which were found to have the most impact on the entire home/apartment price variability, among the remaining explanatory variables in the OLS model, continue to have the same position in the GWR model. While in parishes such as *Estrela*, *Campo de Ourique* and *Santo António* for each additional bedroom, the average listings' price can increase between 16.4% and 23.8%, *ceteris paribus*, in *Parque das Nações*, *Olivais*, *Lumiar* and *Santa Clara* parishes the percentage of increase is between 6.3% and 12% (Figure A11 - appendix). The same happens with the variable *capacity*, in which the listings' price can increase between 8.9% and 12.7% (*São Vicente* and *Santa Maria Maior*) or between 2.2% and 6.5% (*Carnide*, *Benfica*, *Campolide*, *Alcantara*, *São Domingos de Benfica*, *Alcantara* and *Campo de Ourique*) with an increase of one unit in the listings' capacity, as shown in Figure 7.

Despite the differences between the GWR models, it was found that the number of listing held by each host (*n_list*) is the one that has less impact on the price variability regarding all Airbnb listings types. The coefficients' estimates for this variable vary between -0.002 and 0.001 (Figure A20 – appendix) for GWR model using entire home/apartment data (*home*), and between -0.005 and 0.007 (Figure A18 – appendix) for the GWR model using private/shared room data (*room*)

To enforce the conclusion regarding the multicollinearity, spatial distributions of the condition numbers for each GWR model were computed as shown in the figure below.

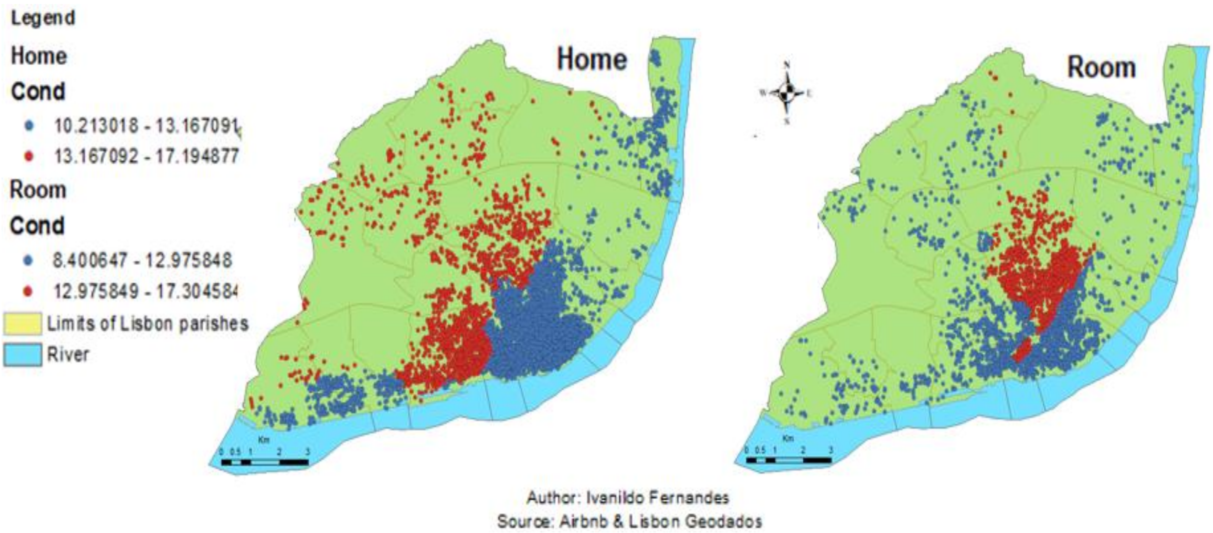

Figure 8 – Spatial distribution of condition numbers for Home and Room models

Align with the previous results, the condition numbers do not exceed 18 (<30), which means that there is no local collinearity. In fact, multicollinearity has nothing to do with standard errors, which is the one that indicates if the coefficients are reliable. In this sense, from Figure A10 to A20 (appendix) it can be noticed that all the standard errors computed for each explanatory variable (*Home* and *Room* models) are very small (near zero) when compared to the related coefficients' estimates, thus the coefficients' estimates produced by the GWR models are reliable.

GWR model results were combined with some information's from previous chapters, such as the number of listings per parish and the hot spot analysis results, to enable a more in-depth analysis of the obtained results. Those information's were combined with the local R^2 of the two GWR models.

For this analysis, only information's regarding the top parishes with highest/lowest number of Airbnb listings were collected as shown in the table below.

Parishes	Nº of listing	Hot spot analysis (l_price)	Local R2 (Home)	Local R2 (Room)
Sta Maria Maior	> 1000	Hot spot	38% - 56%	10% - 21%
Misericórdia	> 1000	Hot spot	38% - 56%	14% - 21%
Arroios	> 1000	Hot/cold spot	48% - 56%	6% - 21%
Santo António	> 1000	Hot spot	38% - 48%	10% - 21%
São Vicente	> 1000	Hot spot/Not significant	48% - 56%	10% - 21%
Santa Clara	< 100	Cold spot	28% - 38%	6% - 10%
Carnide	< 100	Hot spot/Not significant	28% - 38%	6% - 10%
Marvila	< 100	Cold spot	38% - 48%	6% - 14%
Benfica	< 100	Cold spot/Not significant	28% - 38%	6% - 10%
Beato	< 100	Cold spot	43% - 56%	6% - 10%

Table 8 – Additional analysis

The first five parishes correspond to those in which the number of Airbnb listings is higher than 1000 listings, and the last five parishes are those with lower concentration of Airbnb listings, less than 100. While Airbnb listings with high price values surrounded by other Airbnb listings with high price values (hot spot) are likely to be located within parishes with high number of listings, Airbnb listings with low price values surrounded by other Airbnb listings with low price values (cold spot) are located in parishes with low number of listings (Table 8). After combined the information regarding the number of listings, hot spot analysis and local R^2, it was concluded that the percentage of price variability explained by the two GWR models (*home* and *room*) are higher in parishes with high listings number than those with low listings number.

5 CONCLUSIONS

Airbnb is the leading provider of local accommodation services in Lisbon. The reference month chosen for the data collection was June 2017. In this period, parishes such as *Arroios*, *Misericordia*, *Santa Maria Maior*, *Santo António* and *São Vicente* were found to have the highest concentration of Airbnb listings (more than 1000 listings). In contrast, *Benfica*, *Beato*, *Carnide*, *Marvila* and *Santa Clara* parishes had less than 100 listings. *Misericórdia* was the only parish found to have, at the same time, the largest number of Airbnb listings classified as entire/home apartment, with a high number of bedrooms and capacity to allocate more guests, receive more reviews, have better ratings and are held mainly by multi-listing host. Entire home/apartment was found to be the most predominant listing type not only in *Misericórdia*, but in almost every Lisbon parish, representing approximately 74% of the total number of listings.

RNAL's data confirmed the existence of a strong presence of specialized agency participating in the local accommodation market within Lisbon municipality, since more than 47% of its records are held by corporations. Similar conclusion was retrieved from the data used in this research, which showed that, on average, each host held 10 Airbnb listings. The number of Airbnb listings was also used to identify hosts that are looking for extra income (single-listing host) and the specialized agencies (multi-listing host). In fact, 4,471 hosts were found to have only one listing registered on *Airbnb.pt* (single-listing host) and the remaining 8,761 held more than two (multi-listing host).

The average listings' price, considering all types of listings, is approximately 74€, with a minimum of 9€ and a maximum of 6,378€. The large price variability is explained by the fact that Airbnb hosts are free to set up the desired price to be charged for sharing a property, without any restriction. *Parque das Nações*, *Santa Maria Maior*, *Misericórdia* and *Santo António* parishes were found to be composed mainly by Airbnb listings with high price values, while Airbnb listings with low price values are concentrated mostly in *Avenidas Novas*, *Areeiro*, *Arroios*, *Penha de França*, *Alvalade* and *Campolide* parishes.

As expected, GWR models produced better results when compared to OLS models. For instance, when considering all observations (*All data*), the GWR model was able to explain 45% of the overall listings' prices variability, while the OLS model explained only 39%. These two models were estimated based on the same explanatory variables (*capacity*, *bedrooms*, *l_reviews*, *l_satisf*, *n_list* and *l_center*). After estimating models based on the type of listings, entire home/apartment and private/shared room, GWR model also produced better results. However, it was concluded that the price of the two types of Airbnb listings are not strongly impacted by the same factors. The available explanatory variables used in the model estimation (GWR and OLS) were able to better explain the price variability of the entire home/apartment rather than private/shared room. For entire home/apartment, the GWR model explained between 43% and 56% of the price variability in parishes such as *Arroios*, *Areeiro*, *Penha de França*, *Beato* and *Avenidas Novas*, while for private/shared room, the price variability explained in those parishes are between 6% and 14%. In addition, it was also found that the same explanatory variable can have positive or negative impact, with different weights across Lisbon parishes and also in the same parish.

Local accommodation modalities have different features, which can be impacted by different factors. When imposing regulatory policies, it's important to take into account the fact that entire home/apartment and private/shared room have different features and impacts on hosts' revenue and location in which they are inserted. One of the main criticisms regarding local accommodation was that rental flats are being removed from the market and, consequently, increasing rents and driving out local population (García-palomares & Gutiérrez, 2016). This is mostly applicable for listings classified as entire home/apartment, since in the case of private/shared room, it does not necessarily mean that a flat is being removed from rental market to be used exclusively as local accommodation. Usually, in these cases, most of the hosts share their main house with guests.

Location is another important term to have in consideration. Parishes, such as *Santa Maria Maior*, *Misericórdia*, *Arroios*, *Santo António*, *Santa Clara* and *São Vicente,* besides having the highest number of Airbnb listings, also are the most expensive ones.

Even though the number of listings registered in the Airbnb platform is constant growing, more than half of these properties are being rented without any formal agreement, which makes it difficult to monitor.

5.1 LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The main limitation faced when performing this research was regarding the data. As mentioned before, instead of providing data to the public, Airbnb just make available some insights analysis concerning its listings at the city/neighborhood level, such as revenue growth, rental analysis, top properties, pricing and many other indicators. Assess to this information required a payment and it is mostly directed to investors rather than researchers. For this reason, non-official source (*Tom Slee*) was used in this research, which did not have most of the information needed. *Inside Airbnb* is another non-official source which provides more detailed information regarding Airbnb listings, when compared to *Tom Slee*. However, this second source was not providing information for the Lisbon market until 15/09/2018. The lack of information was the main reason why the GWR model estimated based on the available explanatory variables was poorly explaining the price variation of the Airbnb listings classified as private/shared rooms.

For future studies, it could be interesting to use data provided by the *Inside Airbnb*, so that new explanatory variables can be used to test if they are statistically significant to explain the price variation of each type of listing. For instance, the number of bathrooms, smoke permission (allowed or not allowed), number of reservations received, cleaning tax and others, can be used as explanatory variables for future models. In addition, socio-economic data concerning each Lisbon parish can also be tested.

References

Alsudais, A. (2017). Quantifying the Offline Interactions between Hosts and Guests of Airbnb.

Bitter, C., Mulligan, G. F., & Dall'erba, S. (2006). Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*. https://doi.org/10.1007/s10109-006-0028-7

Brunsdon, C., Fotheringham, S., & Charlton, M. (1998). Geographically weighted regression-Modelling spatial non-stationarity. *Journal of Royal Statistical Society*. https://doi.org/10.1111/1467-9884.00145

Cansoy, M., & Schor, J. (2016). Who Gets to Share in the "Sharing Economy": Understanding the Patterns of Participation and Exchange in Airbnb. *Unpublished Paper, Boston College*. Retrieved from http://www.bc.edu/content/dam/files/schools/cas_sites/sociology/pdf/SharingEconomy.pdf

Cardoso, D. (2015). To buy or to rent? That is the question! (Portuguese Real Estate market on Trial).

Choi, K. H., Jung, J., Ryu, S., Kim, S. Do, & Yoon, S. M. (2015). The relationship between Airbnb and the hotel revenue: In the case of Korea. *Indian Journal of Science and Technology*. https://doi.org/10.17485/ijst/2015/v8i26/81013

Coyle, D., & Yeung, T. Y. C. (2016). Understanding Airbnb in Fourteen European Cities.

Cruz, T. P. C.-R. (2016). O impacto do alojamento local na reabilitação urbana em Lisboa.

Dardala, A. E., & Constantin, D.-L. (2015). A Spatial Analysis of Tourism Activity in Romania.

Dogru, T., & Pekin, O. (2015). Pricing Airbnb accommodations: A hedonic pricing approach. *Researchgate.Net*.

Edelman, B., & Luca, M. (2013). Digital Discrimination: The case of Airbnb.com. https://doi.org/10.1038/eye.2013.33

Farber, S., & Yeates, M. (2006). A Comparison of Localized Regression Models in a Hedonic House Priee Context. *Canadian Journal of Regional Science*.

Forgacs, G., & Dimanche, F. (2016). Revenue challenges for hotels in the sharing economy: Facing the Airbnb menace. *Journal of Revenue and Pricing Management*. https://doi.org/10.1057/s41272-016-0071-z

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

Gallagher, L. (2017). *The Airbnb story: How to Disrupt an Industry, Make Billions of Dollars... and Plenty of Enemies*.

García-palomares, J. C., & Gutiérrez, J. (2016). Airbnb in tourist cities : comparing spatial patterns of hotels and peer-to-peer accommodation. *ArXiv Preprint*.

Goree, K. (2016). Battle of the Beds : The Economic Impact of Airbnb on the Hotel Industry in Chicago and San Francisco. https://doi.org/10.1111/1745-9133.12174

Gutierrez, J., Garcia-Palomares, J. C., Romanillos, G., & Salas-Olmedo, M. H. (2016). Airbnb in touristic cities: comparing spatial patterns of hotels and peer-to-peer accommodations. *ArXiv Preprint*. https://doi.org/10.1016/j.tourman.2017.05.003

Gyódi, K. Airbnb and Booking.com: Sharing Economy Competing Against Traditional Firms? (2017). Retrieved from delab.uw.edu.pl

Jefferson-Jones, J. (2015). Airbnb and the Housing Segment of the Modern “Sharing Economy”: Are Short-Term Rental Restrictions an Unconstitutional Taking ?

Ke, Q. (2017). Sharing Means Renting?: An Entire-Marketplace Analysis of Airbnb. https://doi.org/10.2139/ssrn.2902840

L. Krause, A., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*. https://doi.org/10.1016/j.cities.2012.06.006

Lee, D. (2016). How Airbnb short-term rentals exacerbate Los Angeles’s affordable housing crisis: Analysis and policy recommendations. *Harvard Law & Policy Review*. https://doi.org/10.1525/sp.2007.54.1.23.

Lehr, D. (2015). An Analysis of the Changing Competitive Landscape in the Hotel Industry Regarding Airbnb. https://doi.org/10.1108/17554250910948712

Levendis, J. D., & Dicle, M. F. (2016). The Economic Impact of Airbnb on New Orleans. *Ssrn*. https://doi.org/10.2139/ssrn.2856771

Martinho, V. J. P. D. (2013). Spatial analysis of the tourism supply, in Portugal, at NUTs III level.

Merante, M., & Horn, K. M. (2016). Is Home Sharing Driving Up Rents ? Evidence from Airbnb in Boston.

Neeser, D. (2015). Does Airbnb Hurt Hotel Business: Evidence from the Nordic Countries. https://doi.org/10.13140/RG.2.1.4939.8248

Nova School of Business and Economics, & Faculdade de Direito da Universidade de Lisboa. (2016). O Alojamento Local em Portugal - qual o fenómeno?

Quattrone, G., Proserpio, D., Quercia, D., Capra, L., & Musolesi, M. (2016). Who Benefits from the Sharing Economy of Airbnb? International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/2872427.2874815

Sheppard, S., & Udell, A. (2016). Do Airbnb properties affect house prices?

Tussyadiah, Iis P., & Pesonen, J. (2015). Impacts of Peer-to-Peer Accommodation Use on Travel Patterns. *Journal of Travel Research*. https://doi.org/10.1177/0047287515608505

Vieira, A. C. M., & Moreira dos Santos, L. D. P. (2016). O impacto económico do turismo – uma abordagem espacial.

Xu, Y., Kim, J., & Pennington-Gray, L. (2017). Explore the Spatial Relationship between Airbnb Rental and Crime. *Tourism Travel and Research Association: Advancing Tourism Research Globally*.

Yang, Y., & Wong, K. K. F. (2013). Spatial Distribution of Tourist Flows to China’s Cities. *Tourism Geographies*. https://doi.org/10.1080/14616688.2012.675511

Ytreberg, N. S. (2016). Competitive effects of Airbnb on the Norwegian hotel market.

Zervas, G., Proserpio, D., & Byers, J. W. (2013). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. https://doi.org/10.2139/ssrn.2366898

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

APPENDIX

Variables	Listing Type	Nº of observations	Minimum	Maximum	Mean	Standard Deviation
price	Entire home	9,764	9 €	6,378 €	86.18 €	129.61 €
	Private room	3,297	9 €	392 €	38.86 €	38.54 €
	Shared room	171	9 €	391 €	29.27 €	45.84 €
	Total	**13,232**	**9 €**	**6,378 €**	**73.60 €**	**115.04 €**
n_list	Entire home	9,764	1	171	10.84	25.87
	Private room	3,297	1	86	8.54	15.37
	Shared room	171	1	17	6.56	4.4
	Total	**13,232**	**1**	**171**	**10.21**	**23.54**
reveiws	Entire home	9,764	0	438	32.9	44.7
	Private room	3,297	0	344	19	35.6
	Shared room	171	0	96	9.7	14.4
	Total	**13,232**	**0**	**438**	**29.13**	**42.8**
satisf	Entire home	9,764	0	5	3.5	2
	Private room	3,297	0	5	2.6	2.2
	Shared room	171	0	5	2.4	2.2
	Total	**13,232**	**0**	**5**	**3.3**	**2.1**
bedrooms	Entire home	9,764	0	10	1.7	1.2
	Private room	3,297	0	10	1.05	0.4
	Shared room	171	1	1	1	0
	Total	**13,232**	**0**	**10**	**1.5**	**1.06**
capacity	Entire home	9,764	1	16	4.5	2.2
	Private room	3,297	1	16	2	1
	Shared room	171	1	16	4.6	3.6
	Total	**13,232**	**1**	**16**	**4**	**2.3**

Table A1 – Summary statistics

Variables	Nº of observations	Minimum	Maximum	Mean	Standard Deviation
price	13143	9 €	491 €	67.4€	52.6€
n_list	13143	1	171	10.25	23.6
reveiws	13143	0	138	29.3	42.8
satisf	13143	0	5	3.3	2.1
bedrooms	13143	0	10	1.5	1
capacity	13143	1	16	4	2.3
CENTER	13143	0	9.85	2.2	1.7
nr_metro	13143	0	4	1	0.9
nr_mon	13143	0	32	12.6	10
nr_mus	13143	0	23	9.9	7.5
nr_tuk	13143	0	16	5.7	5.2

Table A2 – Summary statistics after removing outliers

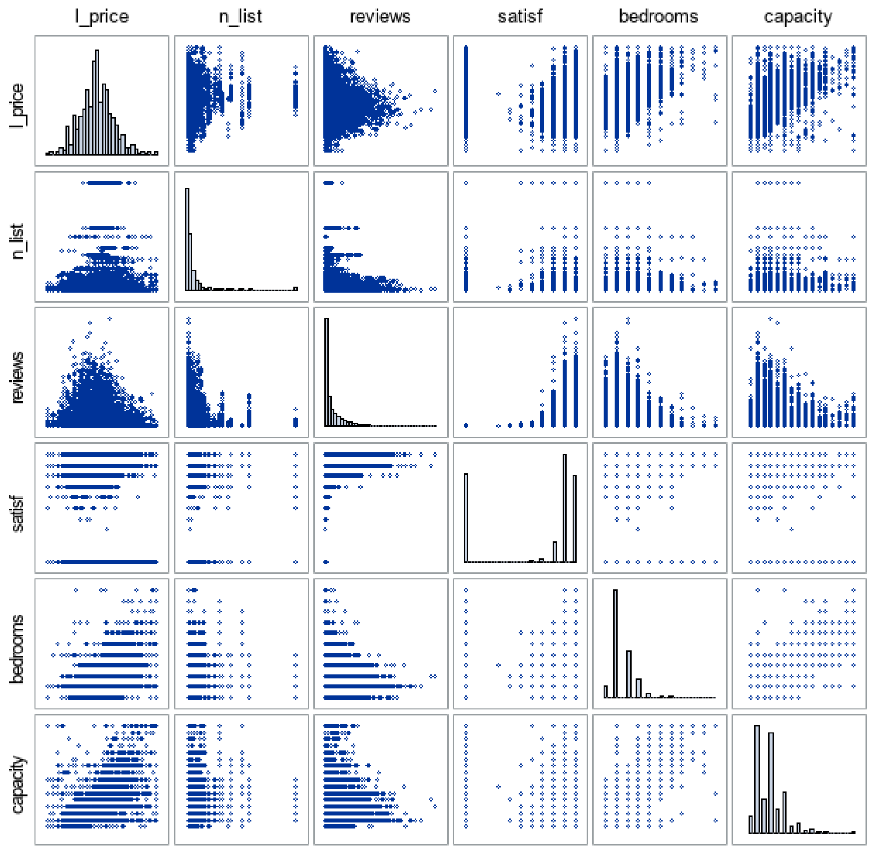

Figure A1 - Scatter Plot Matrix (original variables)

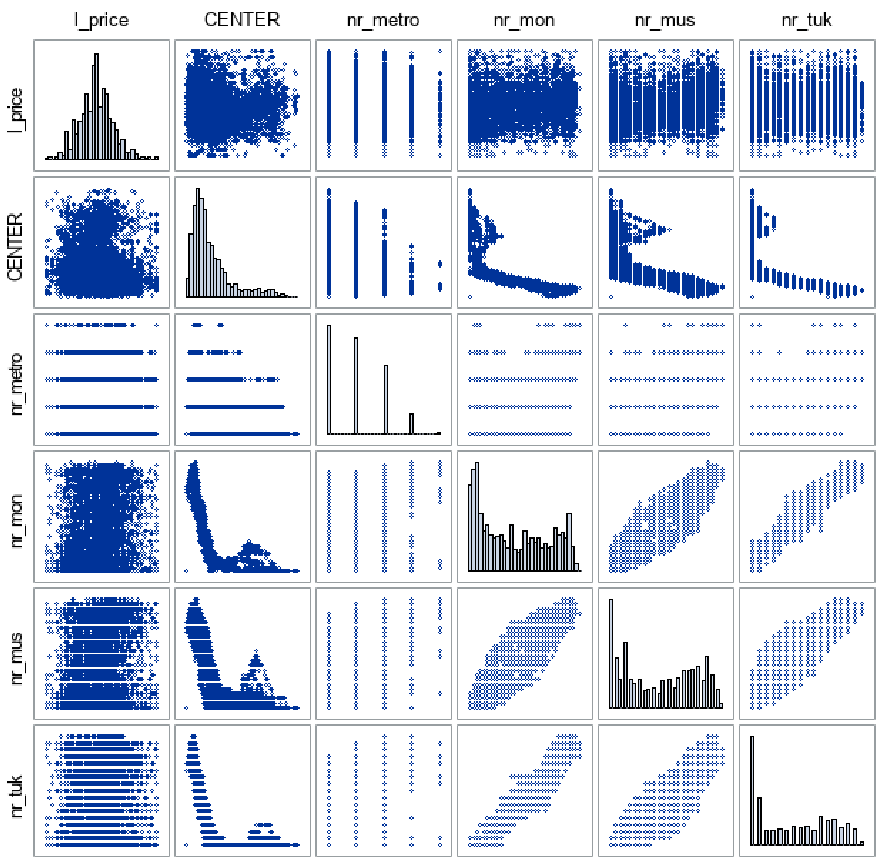

Figure A2 - Scatter Plot Matrix (new variables)

	Pearson Correlation Coefficients, N = 13143									
	Prob > |r| under H0: Rho=0									
	l_price	*n_list*	*reviews*	*satisf*	*CENTER*	*bedrooms*	*capacity*	*nr_metro*	*nr_mon*	*nr_mus*	*nr_tuk*
l_price	1.00000										
n_list	0.06534	1.00000									
reviews	-0.07402	-0.10451	1.00000								
satisf	-0.07713	-0.10480	0.42969	1.00000							
CENTER	-0.09528	-0.05170	-0.23167	-0.19153	1.00000						
bedrooms	0.51697	0.06221	-0.03930	0.02836	-0.00151	1.00000					
capacity	0.58670	0.06639	0.01156	0.08601	-0.06941	0.78536	1.00000				
nr_metro	0.03877	0.05934	0.05042	0.02495	-0.27042	0.05149	0.04053	1.00000			
nr_mon	0.18766	0.04234	0.26258	0.20226	-0.73225	0.00004	0.09341	0.35010	1.00000		
nr_mus	0.21705	0.05931	0.25721	0.20499	-0.67004	-0.00175	0.10426	0.23783	0.91587	1.00000	
nr_tuk	0.19459	0.04567	0.26528	0.19986	-0.70947	-0.00637	0.09179	0.34598	0.97866	0.92706	1.00000

Table A3 - Pearson correlation matrix

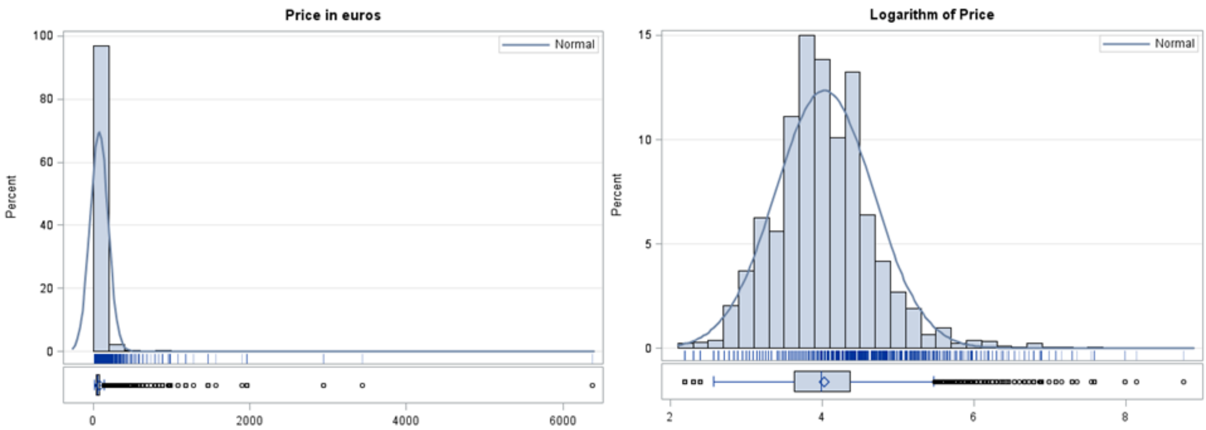

Figure A 3 – Price distribution

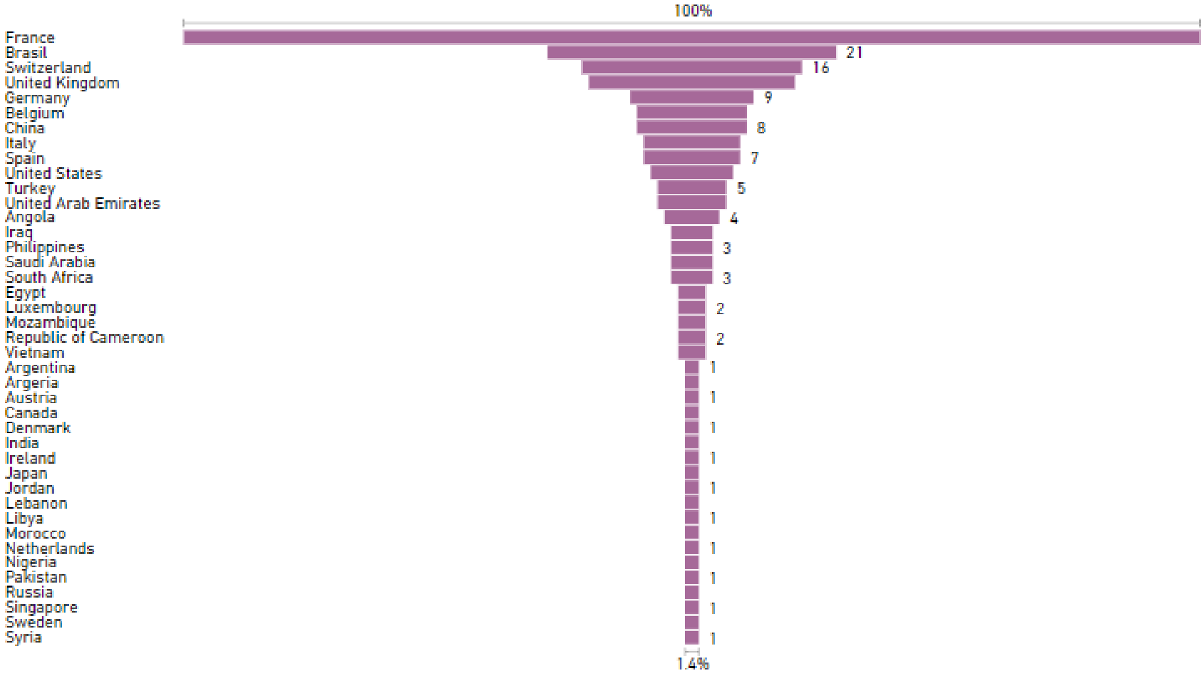

Figure A4 - RNAL – Total number of listings by host's residence country (excluding Portugal)

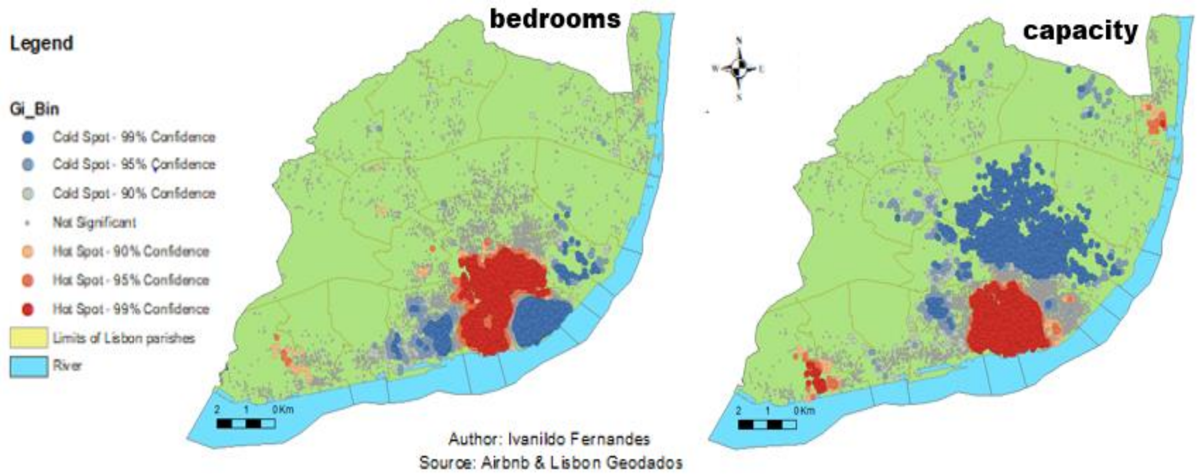

Figure A5 - Hot Spot Analysis (bedrooms and *capacity variable)*

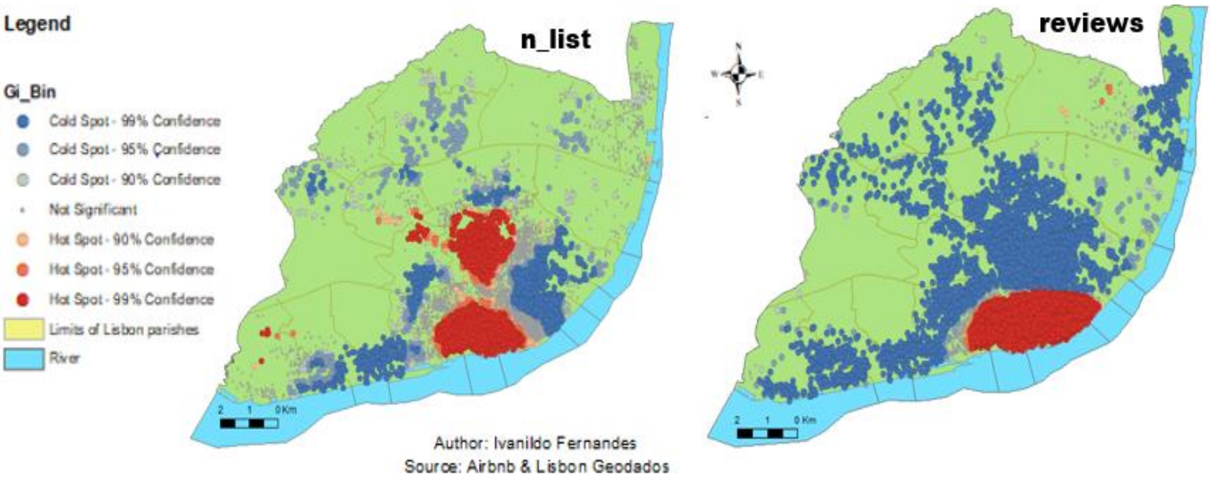

Figure A6 - Hot Spot Analysis (n_list and *reviews variable)*

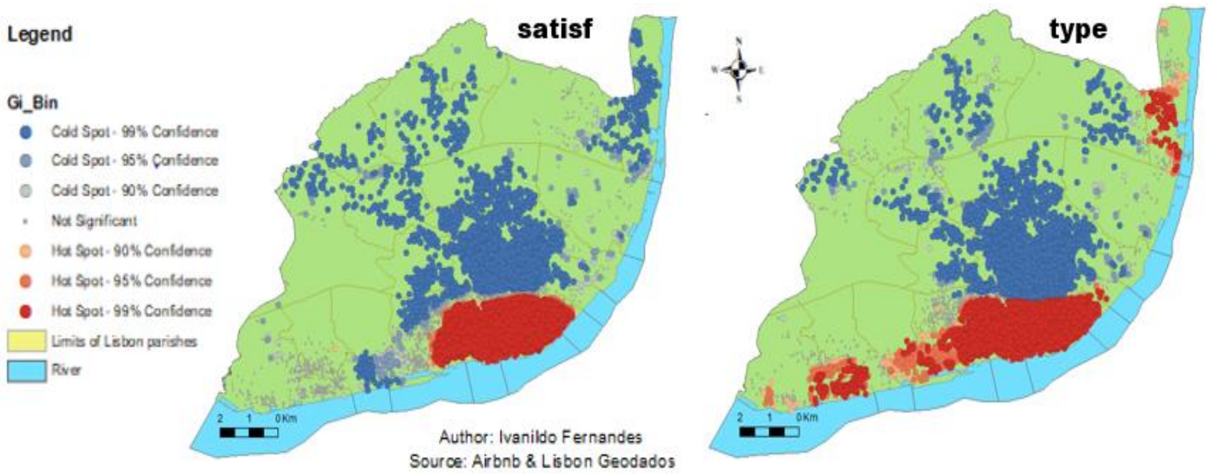

Figure A7 - Hot Spot Analysis (satisf and *type variable)*

		Models	AdjR2	AICc
All data	#1	*- satisf*** + capacity*** + bedrooms*** + type*** + nr_mus****	0.52	14553.57
	#2	*- satisf*** + capacity*** + bedrooms*** + type*** + nr_tuk****	0.52	14629.62
	#3	*- satisf*** + capacity*** + bedrooms*** + type*** + nr_mon****	0.52	14670.32
Home	#1	*- Reviews*** - satisf*** + capacity*** + bedrooms*** + nr_mus****	0.42	8372.48
	#2	*- Reviews*** - satisf*** + capacity*** + bedrooms*** + nr_mon****	0.41	8477.78
	#3	*- Center*** - satisf*** + capacity*** + bedrooms*** + nr_mus****	0.40	8575.33
Room	#1	*- satisf*** + capacity*** - n_list*** + Center*** + nr_mus****	0.12	5230.04
	#2	*- Reviews* - satisf*** + capacity*** + Center*** + nr_mus****	0.12	5232.97
	#3	*- satisf*** + capacity*** + bedrooms + Center*** + nr_mus****	0.12	5233.26
		Model Variable significance (* = 0.10; ** = 0.05; *** = 0.01)		

Table A4 – Exploratory regression results

Base model	ID	Additional variable(s)/transformations	R2	AdjR2	AICc
All data (#1)	1	*Center*	0.522	0.520	14505.76
	2	*Reviews*	0.526	0.526	14388.23

	3	*Nr_metro*	0.520	0.520	14548.31
	4	*N_list*	0.520	0.520	14557.84
	5	*Reviews; Center*	0.528	0.527	14344.62
	6	*l_Center; l_Reviews*	0.533	0.533	14191.14
	7	*Center; l_Reviews*	0.534	0.534	14157.99
	8	*Center; l_Reviews; l_Satisf*	0.535	0.535	14140.80
Home (#1)	1	*Center*	0.417	0.416	8350.21
	2	*Nr_metro*	0.417	0.417	8341.24
	3	*N_list*	0.416	0.415	8367.06
	4	*l_Center*	0.415	0.415	8374.20
	5	*l_Reviews*	0.428	0.428	8152.78
	6	*l_Reviews; l_Satisf*	0.429	0.429	8146.81
	7	*Center; l_Reviews; l_Satisf*	0.430	0.429	8131.44
	8	*Center; Nr_metro; l_Reviews; l_Satisf*	0.432	0.432	8093.29
Room (#1)	1	*Reviews*	0.121	0.119	5230.51
	2	*l_Reveiws*	0.124	0.123	5216.50
	3	*l_satisf*	0.122	0.121	5222.24
	4	*nr_metro*	0.120	0.119	5232.02
	5	*l_Reveiws; l_satisf*	0.126	0.124	5211.68

Table A5 – Regression modelling test (OLS)

Input Features:	f	Dependent Variable:	F.L_PRICE
Number of Observations:	13143	Akaike's Information Criterion (AICc) [d]:	14140,807563
Multiple R-Squared [d]:	0,535365	Adjusted R-Squared [d]:	0,535117
Joint F-Statistic [e]:	2162,069137	Prob(>F), (7,13135) degrees of freedom:	0,000000*
Joint Wald Statistic [e]:	14426,577018	Prob(>chi-squared), (7) degrees of freedom:	0,000000*
Koenker (BP) Statistic [f]:	797,613629	Prob(>chi-squared), (7) degrees of freedom:	0,000000*
Jarque-Bera Statistic [g]:	9178,981280	Prob(>chi-squared), (2) degrees of freedom:	0,000000*

Figure A8 – OLS Diagnostics (All Data)

	Model 1: All Data			**Model 2: Home**			**Model 3: Room**		
Variables	**Coeff**	**StdE**	**P-value**	**Coeff**	**StdE**	**P-value**	**Coeff**	**StdE**	**P-value**
bedrooms	0.0990	0.0063	0.0000	0.1271	0.0058	0.0000	0.0173	0.0191	0.3126
capacity	0.1192	0.0029	0.0000	0.0698	0.0029	0.0000	0.0278	0.0066	0.0088
nr_mus	0.0217	0.0007	0.0000	0.0133	0.0007	0.0000	0.0190	0.0015	0.0000
center	0.0029	0.0031	0.0000	0.0122	0.0029	0.0003	0.0274	0.0063	0.0002
l_reviews	-0.0527	0.0042	0.0000	-0.0761	0.0038	0.0000	-0.0352	0.0102	0.0000
l_satisf	-0.0444	0.0100	0.0000	-0.0255	0.0093	0.0065	-0.1507	0.0218	0.000
	R2=0.42 AdjR2=0.42			**R2=0.43 AdjR2=0.43**			**R2=0.12 AdjR2=0.12**		
	AICc=17084,61			**AICc=8131.44**			**AICc=5215.31**		

Table A6 – OLS models' results (excluding variable type)

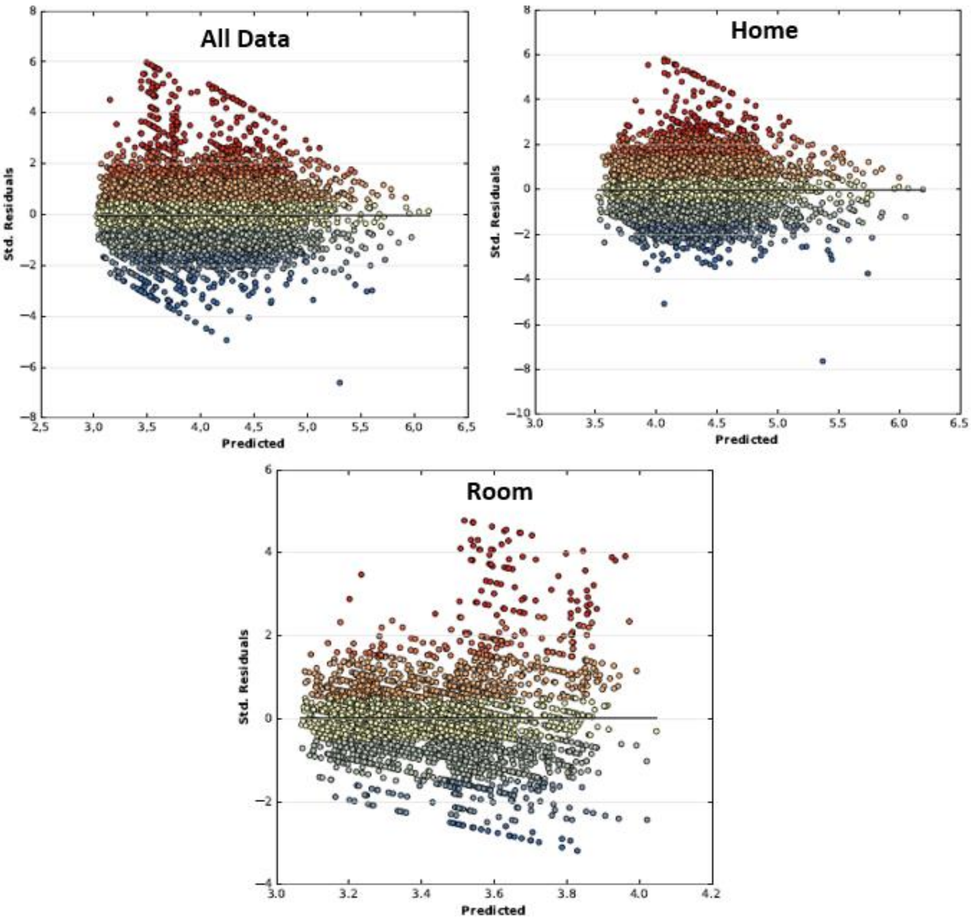

Figure A9 - Residual vs. Predicted Plot (final OLS models)

Moran's Index:	0.049402
Expected Index:	-0.000076
Variance:	0.000000
z-score:	226.874870
p-value:	0.000000

Table A7 – Global Moran's I Summary (dependent variable)

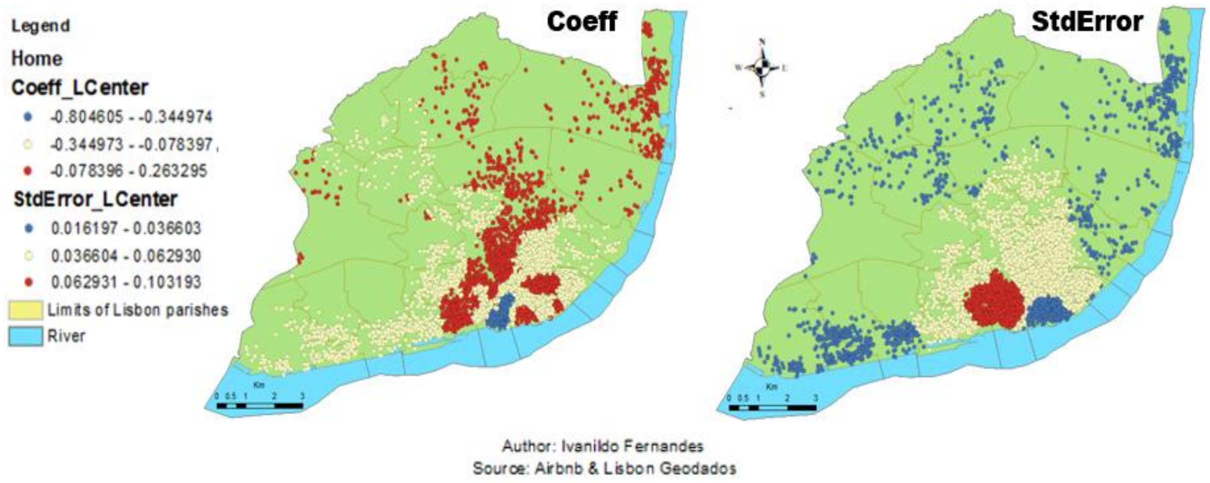

Figure A10 - Spatial distribution of coefficient values and Standard Errors (l_center)

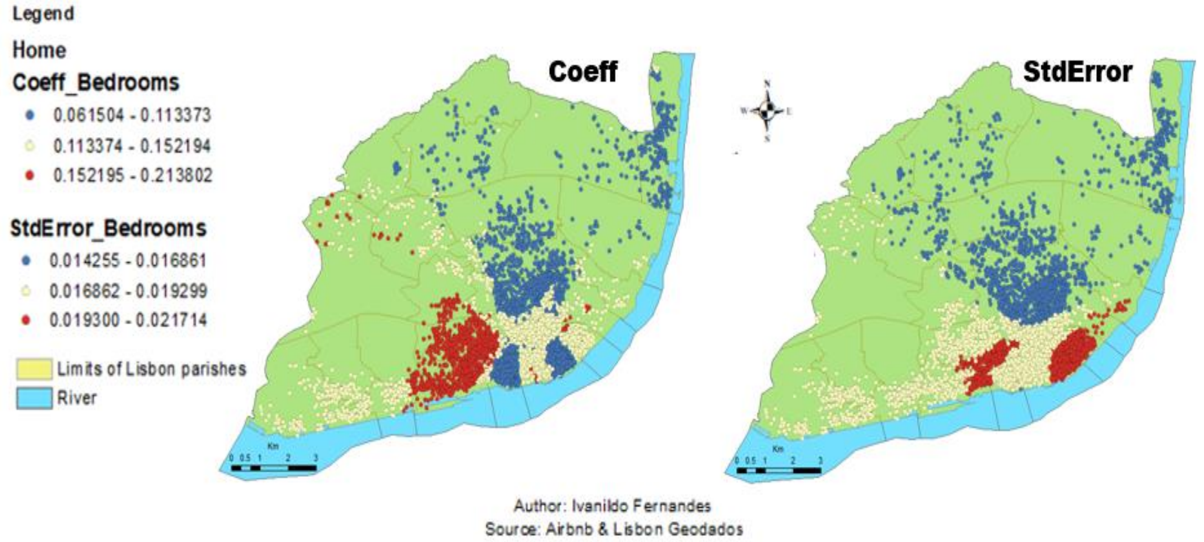

Figure A11 - Spatial distribution of coefficient values and Standard Errors (bedrooms)

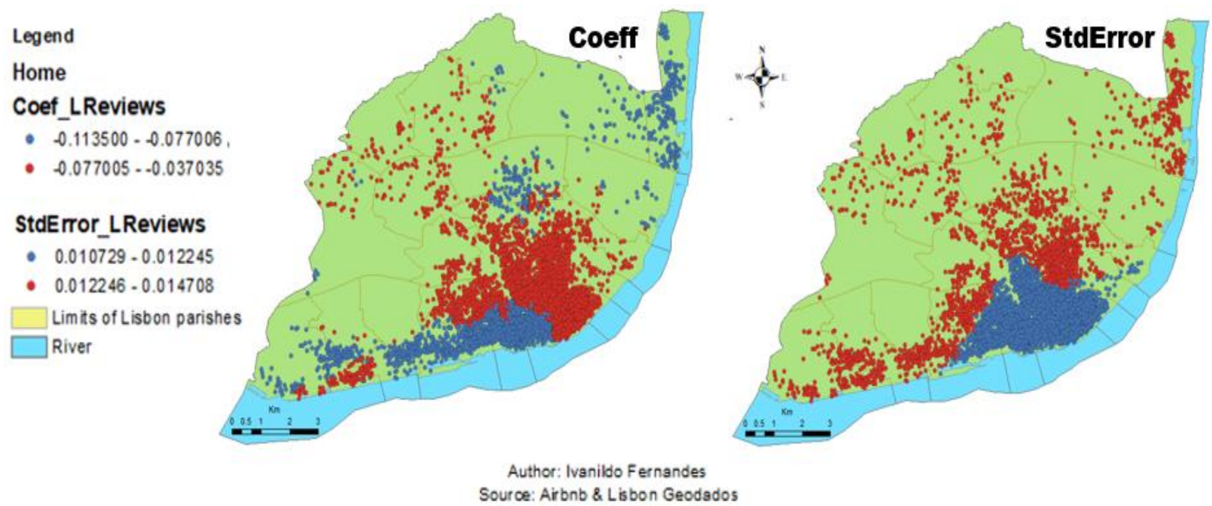

Figure A12 - Spatial distribution of coefficient values and Standard Errors (l_reviews)

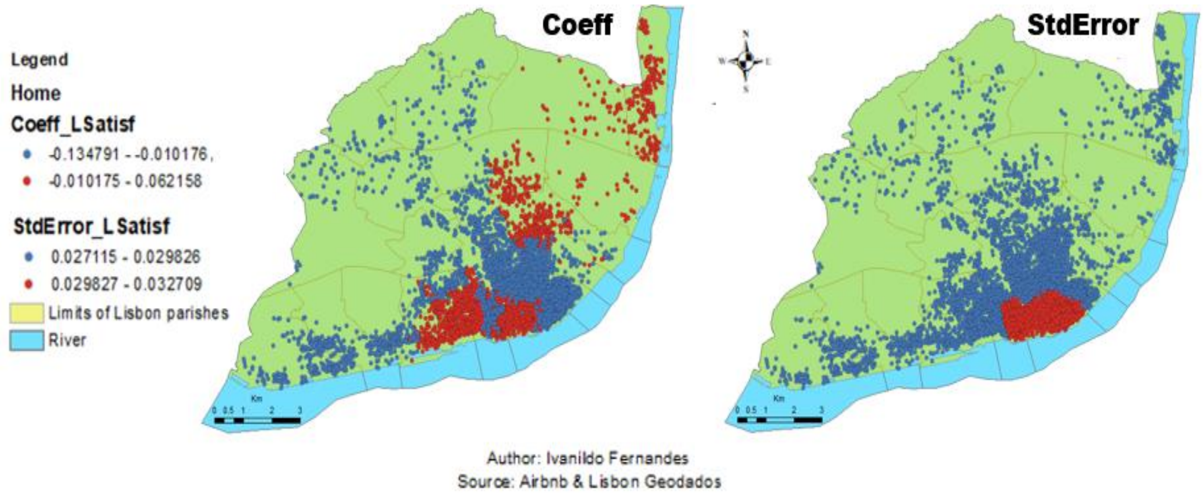

Figure A13 - Spatial distribution of coefficient values and Standard Errors (l_satisf)

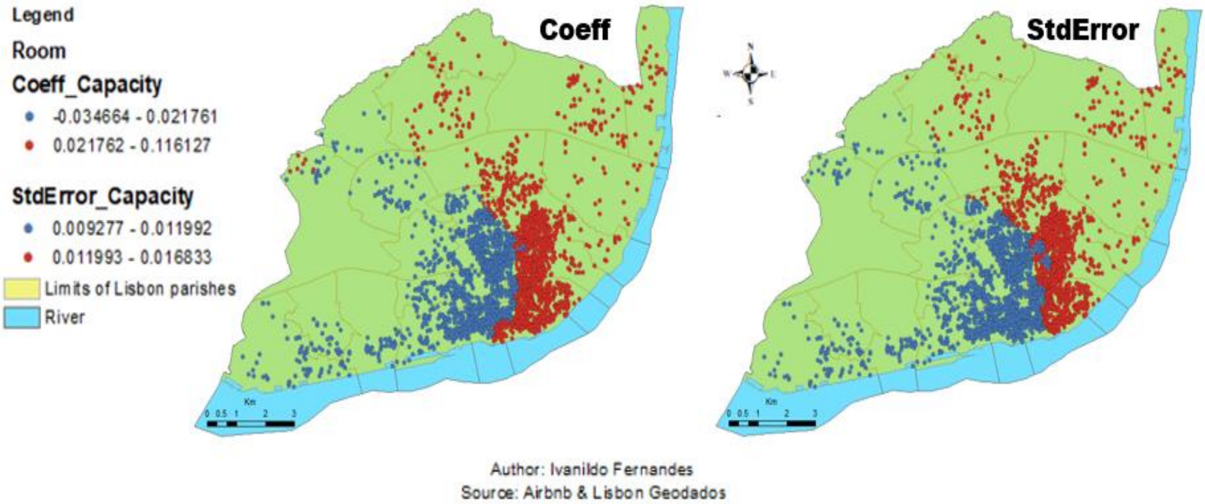

Figure A14 - Spatial distribution of coefficient values and Standard Errors (capacity)

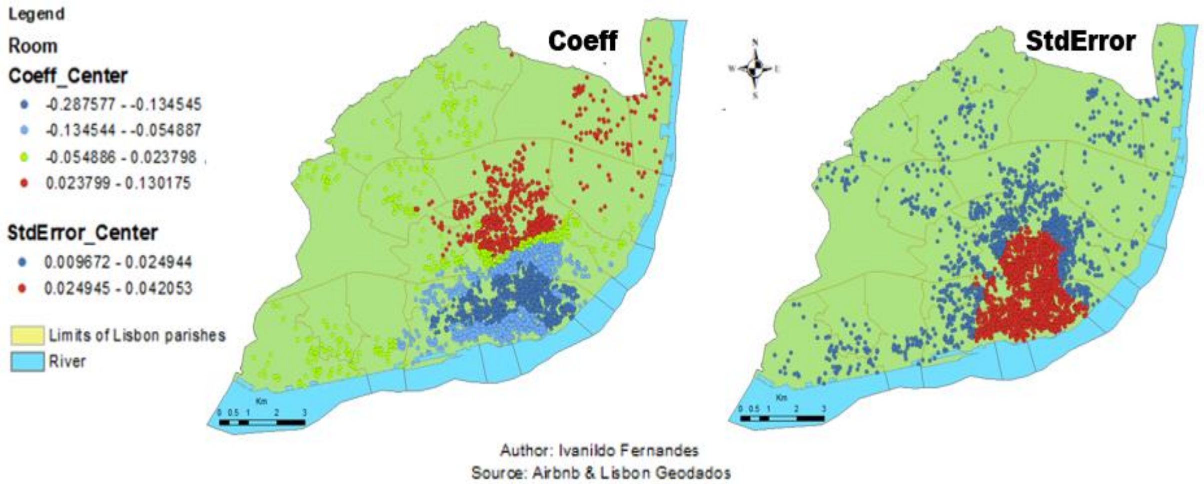

Figure A15 - Spatial distribution of coefficient values and Standard Errors (center)

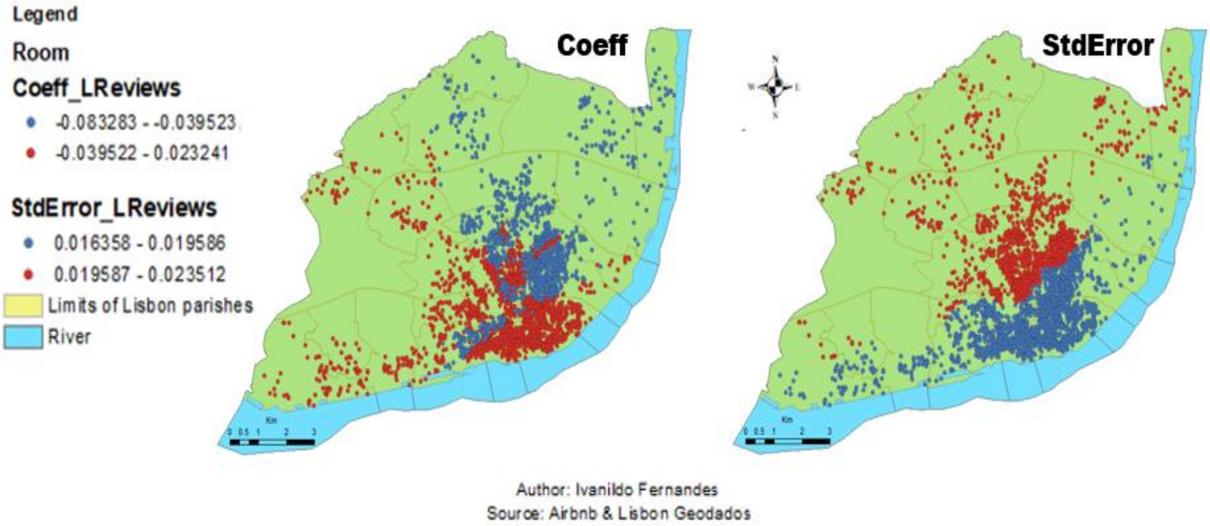

Figure A16 - Spatial distribution of coefficient values and Standard Errors (l_reviews)

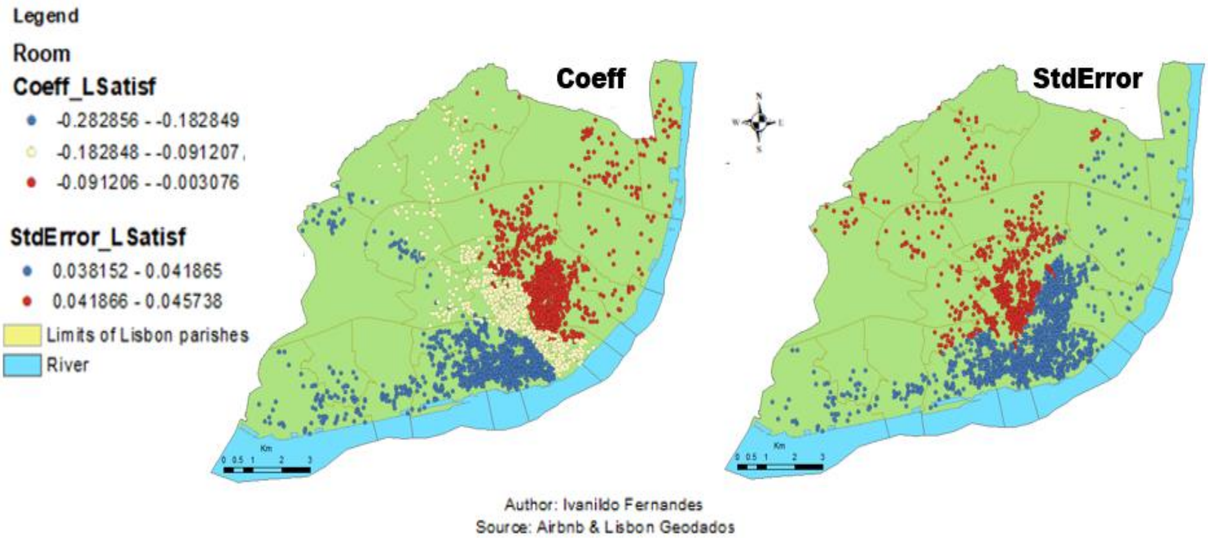

Figure A17 - Spatial distribution of coefficient values and Standard Errors (l_satisf)

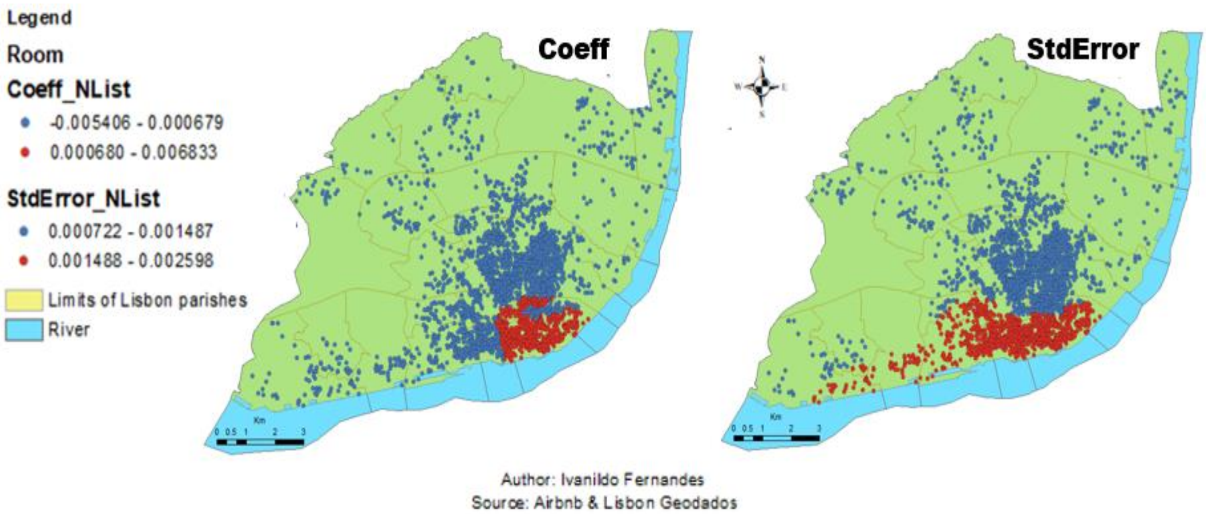

Figure A18 - Spatial distribution of coefficient values and Standard Errors (n_list)

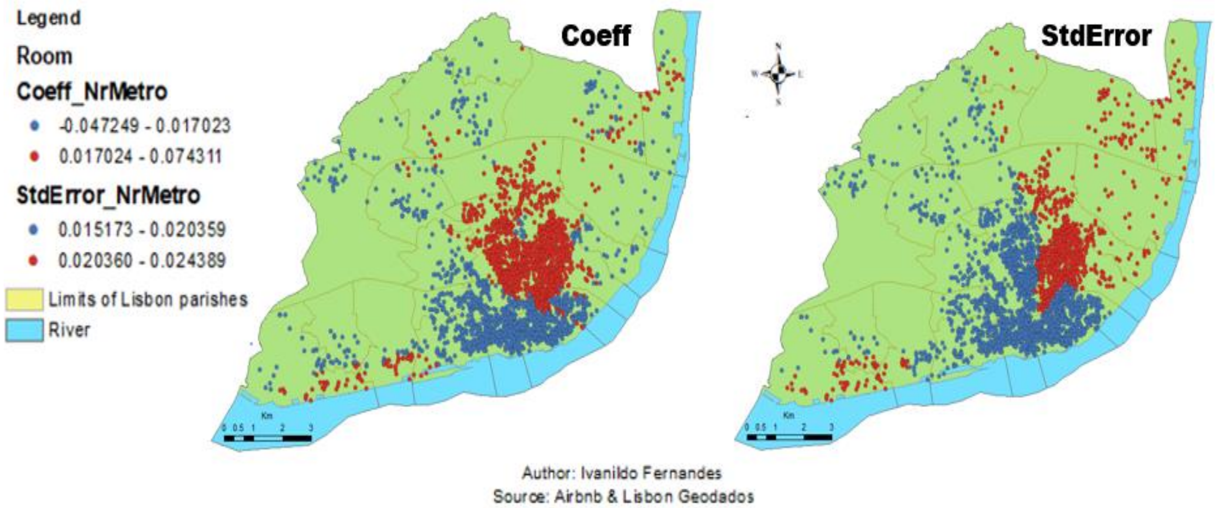

Figure A19 - Spatial distribution of coefficient values and Standard Errors (nr_metro)

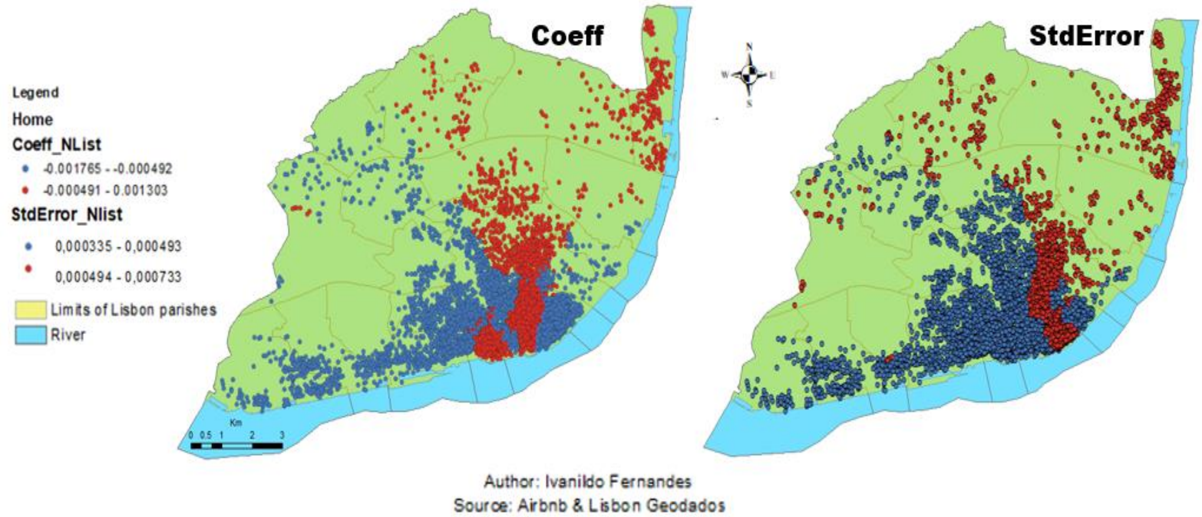

Figure A20 - Spatial distribution of coefficient values and Standard Errors (nr_list)

2018

Title: Modelling the Airbnb listings' price in Lisbon using local spatial regressions

Ivanildo Semedo Correia Fernandes

MEGI