MSc Data Science & Analytics

NCG613 Group Project - May 2022

London Airbnb Price Analysis using Machine Learning

Abhishek Acharya
Eliza Majumder
Jinsee Payyeri
Srishti Mishra

Abstract

This report explores the Airbnb pricing trends to help us understand important attributes that need to be taken towards the prediction of the price for London Airbnb listings. The importance of the features that can help customers decide on the type of Airbnb has been taken into consideration and the best model to select the price prediction of the listings which can help owners decide on the pricing has been identified. Several Geo-Spatial, Data Analytics and Machine learning techniques is used in this report. Price prediction models such as KNN, and Random Forest were built to predict the model were implemented. Random Forest model demonstrated the best performance with the test data collected from the Airbnb website.The accuracy of the model was calculated to be 63.56%.

Contents

1 Introduction

Hotels have always dominated the hospitality industry until sharing business such as Airbnb arrived. Founded in 2008, Airbnb is an online housing platform that provides users to list, discover, and book accommodations for their users across the world. Hosts can offer their property spaces to guests for short or long periods of time [1]. These accommodations differ from the typical hotel room because of the ability to live and witness the culture of place chosen to live. Airbnb's listings offer a huge range of options ranging from community bedrooms to luxury housing, all in one platform. There is a peer-review system where guests can leave reviews after reservations are made. These reviews can often dictate how much owners can book/earn. Unlike the traditional hospitality industry, listings on Airbnb provide more diversified experiences and prices. While the star rating system is primarily used by hotels to develop their pricing strategies, no explicit pricing guideline is available for Airbnb hosts. Hence, pricing becomes more complicated in the context of the sharing economy. However, it is also critical to understand the pricing structure of Airbnb because it drives consumer's decision making as well as stakeholder's profitability [1]. Furthermore, doing so will also help researchers gain profound insights on the sharing economy.

Some prior work has been done to predict the prices using NLP and Sentiment analysis to help property owners and customer to decide on the prices and to evaluate an offered price respectively. However, few studies have evaluated and predicted the price based on the machine learning techniques. This is important to consider as there can be more accuracy for these predictions from the results of using machine learning techniques. Pricing a rental property on Airbnb is a challenging task for the owner as it determines the number of customers for the place. On the other hand, customers must evaluate an offered price with minimal knowledge of an optimal value for the property . These predictions can help hosts get the price right for their properties as it solely depends on the host to set their own prices for their apartment listings [2]. There are a lot of factors that alter the daily prices around the base prices such as the availability in the area, neighbourhood, number of people looking for a place, seasonality or which day of the week it is. We focus on knowing the graphs of the market of Airbnb and based on the past data we try to understand what the customers of Airbnb are looking for. Then, we co- relate all the results by making some important predictions for the reference of future customers.

Pandemic has created a lot of uncertainty about Airbnb prices. We propose to analyse Airbnb's publicly available listing information from the Airbnb website for the recent listings available for

2021 and predict the prices of the Airbnb in London to help hosts and the customers . This report aims to develop a reliable price prediction model using machine learning techniques to aid both the property owners and the customers with price evaluation given minimal available information about the property. Although, there are fewer research that predict the price for different cities using different models, here we have used Machine learning models. The models will provide important information for the pricing strategies of the hosts and other stakeholders.

In this analysis, we are going to use different machine learning techniques to predict the listing prices of Airbnb in London. In spatial data analysis, Hedonic modelling is a technique performed to understand the effect of multiple factors that can lead to ultimate pricing of the house. For example, almost all buildings are different from one another, however, it can be further broken down to its constituent parts like the number of available bedrooms, parking space availability and distance from specific public places. The hedonic model calculates these attributes separately in case of an additive model or elasticity in case of a log model. Finally, the output from this model can be used to estimate prices of different cities or conduct time series analysis of the Airbnb pricing over the years.

From previous research, the spatial factor of distance to a tourist spot was not considered. This analysis also considers a popular borough in London(Westminster) which has major tourist attractions along with other property specific parameters. The analysis dives deeper in unearthing various factors that have major impact on the prices of the Airbnb such as bedrooms, host listing counts and distance from famous tourist spots. Finally, the data was fit into k-nearest neighbour regression and Random Forest Regressor and predictions were made. The accuracy of the Random Forest model came out to be 63.56%. We also found out that for this topic, Random Forest model is the best model and is strongly recommended for price prediction.

2 Literature Review

Pricing a rental property on Airbnb is a tough challenge for the owner since it determines the number of customers who will stay there. Hosts in many areas are presented with a good selection of listings and can filter by criteria such as price, number of bedrooms, room type, and more. Customers, on the other hand, must evaluate an offered price with minimal knowledge of an optimal value for the property. There are also other factors that are affecting the prices on Airbnb like Competitor's Prices, Weekends, High Season, Special Events, and Special Amenities. There are existing Machine Learning predictions for Airbnb listing prices for different cities that will not well fit with the London Airbnb housing dataset [3]. The existing literature on housing prices focuses mainly on non-shared purchases or rental price predictions. Formerly a reliable price prediction model was developed using machine learning, deep learning, and natural language processing techniques to aid both the property owners and the customers with price evaluation given minimal available information about the property [4]. As per their implementation and models tested, they have declared Support Vector Regression (SVR) performed the best and resulted in an R^2 score of 0.69 and MSE of 0.147. In another paper, Machine learning techniques like Linear Regression, Polynomial Regression, Regression Tree, as well as Neural Network, and SVM were used in order to predict the housing price in Melbourne City, Australia. They concluded that a combination of the Step-wise and SVM model is a competitive approach [5]. Another study was done for the Real Estate Price Prediction using Regression and Classification methods, suggesting that for the classification problem, the best-performing model is SVC with linear kernel, with an accuracy of 0.6740 and for the regression problem, the best-performing model is SVR with Gaussian kernel, with RMSE of 0.5271, however, visualization for SVR is difficult due to its high-dimensionality [6]. In recent work a method of Multi-Scale Affinity Propagation (MSAP) aggregating the listing appropriately by the landmark and the facility was proposed. During their study on reasonable price recommendations, MSAP can cluster the house effectively and aggregate the listing into different price zone, which gives the reasonable price uniquely [7]. Various attributes of the listing affect the prices of Airbnb which makes it important to have a look at the relation between several attributes and how they differ while reflecting the price. The host thereby can decide the attributes which help the community grow more by keeping the check on the prices as well. Analysing the attitude of the hosts with the Airbnb property also plays a major role while renting. Most of the people do not have information on how to price the property [8]. All these studies show that the best result is around 60 to 70 percent for the prediction of Airbnb listing prices in different cities. After reading and experimenting with these studies, we see that the results obtained from these methods are not much effective for predicting the Airbnb houses in London. The algorithms which are used do not stand well with the London Airbnb datasets. Our research study tries to improve and use different machine learning algorithms like Hedonic Price modelling, k-nearest neighbours (KNN) algorithm, Random Forest algorithm in the experiments, focusing on the variety of factors that may help us provide the best results in predicting the Airbnb house prices for London, City.

3 Data, Study Area, and Methods

3.1 Study Area

Airbnb is an online platform for posting a rental properties for tourism experiences. So essentially, Airbnb connects people looking to rent their houses with people who are searching for accommodations. We have chosen London city for our study area as it is the largest city in UK and a major tourist destination. Around 2.8 million households across UK casually rented out part of or their entire home on platforms such as Airbnb in 2017-18. According to a research entire home Airbnb listing in London have increased by 571% in 5 years and more of the city's housing stock has been gobbled up by short-term rental companies.[1]. The demand for the listings as per the city, most common property type preferred by the travellers, the price for the accommodation, duration of the stay, availability of the place, etc. are among many factors that help decide what finally drives a user to make the final reservation. Distance of a listing from a major tourist attraction is also a major factor which affects the price of a listing. Many tourists prefer to stay near to the city centre around which major tourist attractions are.

3.2 Data

The public Airbnb dataset for London City [2] was used as the main data source for this study. The dataset includes 66,641 entries, each with 74 features. Table 1 contains the description of major attributes in the Airbnb file.

For the initial prepossessing, we inspected each feature of the dataset to (i) remove features with frequent and irreparable missing fields or set the missing values to zero where appropriate, (ii) convert some features into floats (e.g., by removing the dollar sign in prices), (iii) change Boolean features to binaries, (iv) remove irrelevant or uninformative features, e.g., host picture URL, constant valued fields, or duplicate features. In addition, the features were normalized, and the labels were converted into logarithm of the prices to mitigate the impact of the outliers in the dataset. The attribute with the most importance is geometry as it's the attribute which makes the data spatial(point data).

We have also used the neighbourhood.geojson file [3] which consists of 33 entries and 3 variables.The latitude and longitude is used from this available list of neighbourhoods in the form

[1] More information:https://theconversation.com/entire-home-airbnb-listings-in-london-have-increased-by-571-in-5-years-new-research-172436

[2] The data can be downloaded from http://insideairbnb.com/get-the-data/.

[3] The geojson file can be downloaded from http://insideairbnb.com/get-the-data/.

of a geojson file. The data set has information regarding the districts constituting the London neighbourhood area. This data set has been majorly used to join the listing data with neighbourhood to form a geodataframe to perform spatial analysis. The geometry column consists of multi-polygon values. Refer to Table 2 for attribute description. A heatmap was generated to check the multicollinearity between the features to select which feature has more weightage.

3.3 Methods

All the features were thoroughly studied and changes were made to the required attributes before creating the models. As the dependent variable price was heavily skewed Figure 3, log transformation was performed on it Figure 3. In this study, we are building and comparing machine learning algorithms which includes Spatial Regression, Spatial Lag Models, Random Forest and KNN.

Spatial Regression : Before introducing explicitly spatial methods, we ran a linear regression model. This helped to set the main principles of hedonic modelling and in interpreting the coefficients. Also, this supported in understanding how the spatial extensions are useful.Essentially, the core of a linear regression is to explain a given variable which here is the price of a listing i on Airbnb (Pi),as a linear function of a set of other characteristics we will collectively call Xi. Mathematically this model is interpreted as shown below.

$$ln(Pi) = \alpha + \beta Xi + \epsilon i \tag{1}$$

Spatially lagged exogenous regressors (WX) : Here we have introduced space by spatially lagging one of the explanatory variables(d2westminster).

$$ln(Pi) = \alpha + \beta Xi + \delta \sum_j wijXi' + \epsilon i \tag{2}$$

where ln(Pi) is the dependent variable (logarithmic price), Xi' is a subset of Xi, although it could encompass all of the explanatory variables, and wij is the ij-th cell of a spatial weights matrix W. Since it is a spatial transformation of an explanatory variable, the standard estimation approach -OLS- is sufficient: spatially lagging the variables does not violate any of the assumptions on which OLS relies. Generally variables which could affect the price of the Airbnb (dependent variable) are specially lagged to see the impact on other areas associated with it.

Spatially lagged endogenous regressors (WY) : The prices of listings surrounding a given property also enter their own price function. In mathematical terms it is represented as following.

$$ln(Pi) = \alpha + \lambda \sum_j wijln(Pi) + \beta Xi + \epsilon i \tag{3}$$

In spatial econometrics, this is referred to as a spatial lag model. This specification does violate some of the assumptions on which OLS relies. Here we are including and endogenous variable. GM_Lag (PySAL) is used to obtain reliable coefficients in the estimation method. Here we aim to set an interdependent process by which each Airbnb owner sets their price taking into account the price that will be set in neighbouring locations. From this we were able to measure the statistical significance of estimate of spatial lag of price which in turn helps in understanding whether the owners interact with each other before setting price for their property.

Random Forest Regression : Random Forest is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For regression tasks, the mean or average prediction of the individual trees is returned. The Random Forest model was implemented using the RandomForestRegressor from Python sklearn ensemble Package.The data set was split into 75% training and 25% test set. We evaluated this model on the test set using root-mean-square error (RMSE) and R squared values.

KNN : The KNN algorithm assumes that similar things exist in close proximity. As we had few significant categorical predictors, some scaling had to be done on the data as KNN regressor model does not consider categorical variables. The data such as 'property type', 'room type', 'neighbourhood cleansed' are the categorical variables which was take into consideration. They were converted to numerical values to be used by KNN Algorithm. Again, the data split of 75% training and 25% test set was used. The model was run for different K nearest neighbours and optimal K value was obtained to be 8.

4 Results

From an initial analysis of the data we found that there is an increase in the host joining Airbnb till 2016 as seen in Figure 8 . After that we can see a decrease in the trend.The oldest London listing that is currently live on Airbnb was first listed on the site in August 2008. From 2011 onward, the number of listings started increasing considerably. However, growth in the number of new hosts (of those currently listing on the site) has been decreasing since 2015, when the UK government introduced a law making it illegal to let short-term residential properties for more than 90 nights a year. From the figure Figure 1 , we can see that, when looking for accommodation, being near to certain touristic areas is important. Much of Airbnb listings are centred around Westminster area, which is consistent with the huge draw for tourists, especially during the peak seasons. Also from Figure 2, We can see that, the distribution of prices also reflects that the most expensive listings are on the City of London , Kensington and Chelsea followed by Westminster.A heatmap was also generated as shown in Figure 7.From the heatmap we can see that beds and bedrooms are correlated which means it provides a lot of similar information. But from the spatial results we can see both are significant which is related to other factors including magnitude of the error variance, variance of the variable itself, other variables included in the model. This suggests that bedroom and number of beds are major factors that customers consider while looking for a listing.

From the summary of Baseline Spatial Regression(OLS) in figure 4, We can see that Results are largely unsurprising, but nonetheless reassuring. Both an extra bedroom and an extra bathroom increase the final price around 30%. Accounting for those, an extra bed pushes the price about 14%. The number of listings the host has in total does not have a significant effect on the final price. Although, review_scores_rating has no impact, review_scores_cleanliness has a significant impact and increases the price by 14%. For Spatially lagged exogenous regressor model, results are largely consistent with the original model as we can see in the Figure 5. Also, incidentally, the distance between the listings and the top tourist spot area, Westminster has a slight significant effect of 2% on the price of a given property. As seen in the Figure 6, Spatially lagged endogenous regressor model have results similar in all the other variable except that in this model, the distance between the listings and the top tourist spot area, Westminster does not have a significant effect on the price of a given property. It is also very clear that the estimate of the spatial lag of price is statistically significant with value of 16%. This points to evidence that there are processes of spatial interaction between property owners when they set their price. It was observed that the Spatially lagged endogenous model performed the best with MSE of 0.6072 as compared to MSE of OLS and Spatially lagged exogenous model having values 0.669998 and 0.607458 respectively. We compared the MSE's for Random Forest and KNN model out of which Random Forest has the lower MSE value. The Random Forest model successfully predicts the price of the listings with an accuracy of 63.56%.

5 Discussion and Conclusion

Pricing a rental property on Airbnb is a challenging task for the owner as it determines the number of customers for the place. On the other hand, booking an Airbnb for a trip is a challenging task for customers as they consider multiple factors for the stay and evaluate an offered price with minimal knowledge of an optimal value for the property. This research aims to help both owners and customer upfront. We have tried to understand the importance of the features that can help customers to decide on the type of Airbnb and have come up with the best model to select the price prediction of the listings which can help owners decide on the pricing.

The exploratory data analysis performed on the dataset revealed that the median prices of the listings were higher in the city centre and the listings farther from the centre had comparatively lower prices. The City of London, Kensington and Chelsea and Westminster had the highest median prices of the listings. Also, it was observed that from 2011, the number of listings showed considerable increase. The growth in the number of new hosts has been declining since 2015. One of the major reasons behind this was the law imposed by UK government making it illegal to let short-term residential properties for more than 90 nights a year.

This paper attempts to analyse the Airbnb listing dataset in London. A hedonic model was constructed to understand how the different variables of the dataset are spatially related to one another which helped us to find out which variables in the dataset are affected by neighbouring variables. After the analysis, it was observed that the distance of the houses from the main tourist location becomes insignificant, have no significant impact on the listing price and do not fully explain the variation in the listing prices. It was also observed that the Spatial lag model(M3) performed the best representing the significance of estimate of spatial lag of price which in helped us understand that the owners interact with each other before setting price for their property. The spatial regression performed on the data helped us to understand that the price was dependent on features such as bedrooms, beds and host listing counts as well as not dependent on the features like distance to Westminster indicating that owners price the listing on the availability of the bedrooms and beds and the customers do not consider tourist attractions as a major factor while booking their accommodation respectively. We also came up with the best-performing model for predicting the Airbnb prices based on a limited set of features including property specifications, owner information, and customer reviews on the listings. Random forest model performed the best in predicting price of the Airbnb listings.

Although the analysis states that the variable bedrooms, bed, host listing count were significant in deciding prices of the Airbnb and the distance to tourist spot Westminster was not significant in determining the prices of the Airbnb for the customers, there might still be other factors which could affect the pricing. This can be a topic of further research and can be combined with other existing resources to check if the Airbnb prices have any impact on residential properties prices surrounding them or other economic factors. This will also help researchers and statisticians better estimate the house prices of the boroughs while extending the current research work. Based on the research performed on the data, it can be concluded that the Random Forest model works best on our data.The project will be providing the host with accurate information according to their needs such as location, ratings, and price. The accuracy of the prediction models can be further improved using the reviews and the listing summary attributes. Further experimentation with neural net architectures can be done and more training examples can be included from other hospitality services such as VRBO to boost the performance of K-means clustering in particular with Ridge Regression model.

6 Figures

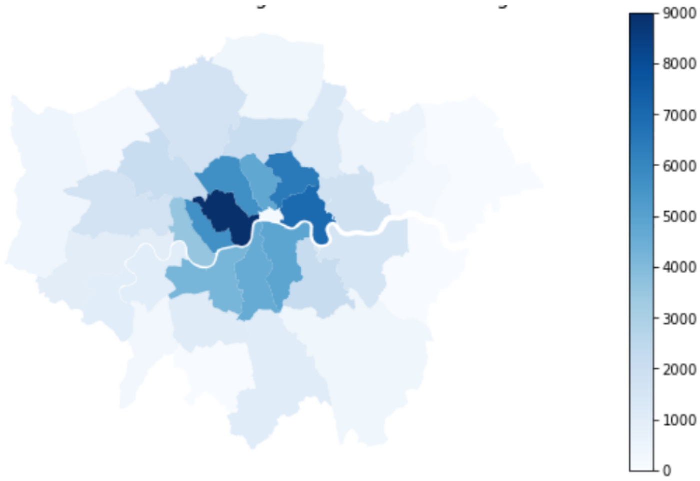

Figure 1: The description of all the attributes in the Airbnb file

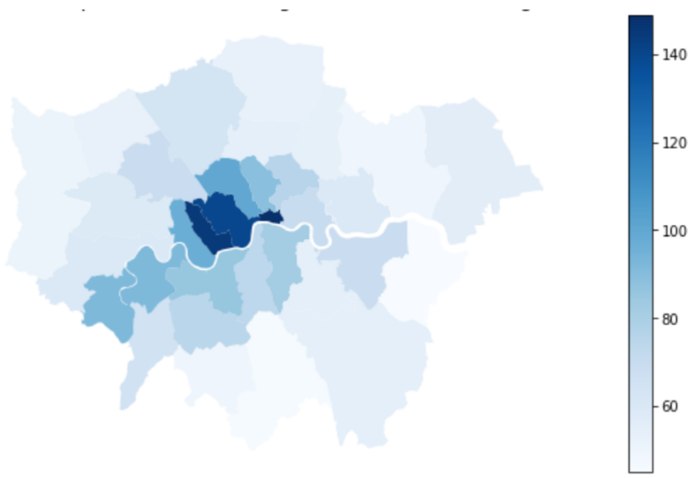

Figure 2: The description of all the attributes in the Airbnb file

Sno.	Attribute Name	Description
1.	host_since	date that the host first joined Airbnb
2.	host_response_time	average amount of time the host takes to reply to messages
3.	host_response_rate	proportion of messages that the host replies to
4.	host_is_superhost	whether or not the host is a superhost, which is a mark of quality for the top-rated and most experienced hosts, and can increase your search ranking on Airbnb
5.	host_listings_count	how many listings the host has in total
6.	host_identity_verified	whether or not the host has been verified with id
7.	neighbourhood_cleansed	the London borough the property is in
8.	property_type	type of property, e.g. house or flat
9.	room_type	type of listing, e.g. entire home, private room or shared room
10.	accommodates	how many people the property accommodates
11.	bathrooms	number of bathrooms
12.	bedrooms	number of bedrooms
13.	beds	number of beds
14.	amenities	list of amenities
15.	price	nightly advertised price (the target variable)
16.	extra_people	the price per additional guest above the guests_included price
17.	minimum_nights	the minimum length of stay
18.	maximum_nights	the maximum length of stay
19.	calendar_updated	when the host last updated the calendar
20.	availability_30	how many nights are available to be booked in the next 30 days
21.	availability_60	how many nights are available to be booked in the next 60 days
22.	availability_90	how many nights are available to be booked in the next 90 days
23.	availability_365	how many nights are available to be booked in the next 365 days
24.	number_of_reviews	the number of reviews left for the property
25.	number_of_reviews_ltm	the number of reviews left for the property in the last twelve months
26.	first_review	the date of the first review
27.	last_review	the date of the most recent review
28.	review_scores_rating	guests can score properties overall from 1 to 5 stars
29.	review_scores_accuracy	guests can score the accuracy of a property's description from 1 to 5 stars
30.	review_scores_cleanliness	guests can score a property's cleanliness from 1 to 5 stars
31.	review_scores_checkin	guests can score their check-in from 1 to 5 stars
32.	review_scores_communication	guests can score a host's communication from 1 to 5 stars
33.	review_scores_location	guests can score a property's location from 1 to 5 stars
34.	review_scores_value	guests can score a booking's value for money from 1 to 5 stars
35.	instant_bookable	whether or not the property can be instant booked (i.e., booked straight away, without having to message the host first and wait to be accepted)
36.	reviews_per_month	calculated field of the average number of reviews left by guest each month

Table 1: The description of all the attributes in the Airbnb file

Sno.	Attribute Name	Description
1.	neighbourhood	Name of the area
2.	neighbourhood group	Neighbourhood group which is the area part of
3.	geometry	Spatial location of each area

Table 2: Description of all the attributes in the neighbourhood geojson file

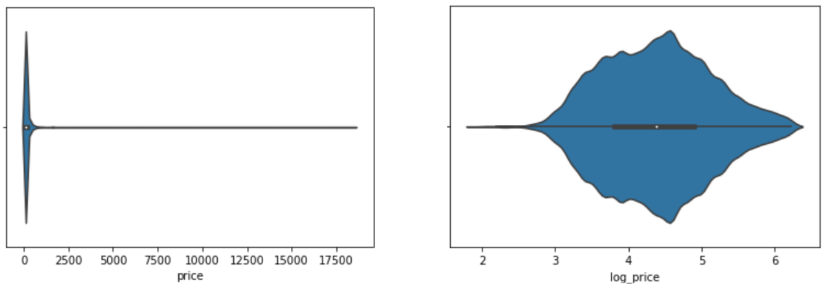

Figure 3: Violin plots of price variable before and after log transformation

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       66641
Mean dependent var  :      4.4648                Number of Variables   :           6
S.D. dependent var  :      0.9297                Degrees of Freedom    :       66635
R-squared           :      0.2249
Adjusted R-squared  :      0.2248
Sum squared residual:   44649.334                F-statistic           :   3866.5010
Sigma-square        :       0.670                Prob(F-statistic)     :           0
S.E. of regression  :       0.819                Log likelihood        :  -81215.269
Sigma-square ML     :       0.670                Akaike info criterion :  162442.537
S.E of regression ML:      0.8185                Schwarz criterion     :  162497.180

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       3.8771339       0.0076162     509.0660961       0.0000000
 host_listings_count       0.0002037       0.0000086      23.7422353       0.0000000
            bedrooms       0.2969318       0.0050670      58.6009362       0.0000000
review_scores_rating      -0.1685607       0.0096624     -17.4449863       0.0000000
                beds       0.1422380       0.0036412      39.0638743       0.0000000
review_scores_cleanliness       0.1463039       0.0096934      15.0931309       0.0000000
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER           34.511

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2   16679780.876           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                5       9000.262           0.0000
Koenker-Bassett test              5        227.010           0.0000
================================ END OF REPORT =====================================
```

Figure 4: The figure depicts the Summary of OLS model

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES
-----------------------------------------
Data set            :     unknown
Weights matrix      :        None
Dependent Variable  :   log_price                Number of Observations:       66641
Mean dependent var  :      4.4648                Number of Variables   :           8
S.D. dependent var  :      0.9297                Degrees of Freedom    :       66633
R-squared           :      0.2972
Adjusted R-squared  :      0.2972
Sum squared residual:   40481.633                F-statistic           :   4026.0374
Sigma-square        :       0.608                Prob(F-statistic)     :           0
S.E. of regression  :       0.779                Log likelihood        :  -77950.159
Sigma-square ML     :       0.607                Akaike info criterion :  155916.317
S.E of regression ML:      0.7794                Schwarz criterion     :  155989.174

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     t-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       4.3082878       0.0089282     482.5492794       0.0000000
 host_listings_count       0.0001661       0.0000082      20.3006014       0.0000000
            bedrooms       0.3109281       0.0048278      64.4036437       0.0000000
review_scores_rating      -0.1741145       0.0092008     -18.9238563       0.0000000
                beds       0.1327406       0.0034690      38.2643686       0.0000000
review_scores_cleanliness       0.1459924       0.0092301      15.8170524       0.0000000
       d2westminster       0.0190714       0.0596608       0.3196634       0.7492245
              w_west      -0.0247200       0.0199021      -1.2420797       0.2142115
------------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER          583.542

TEST ON NORMALITY OF ERRORS
TEST                             DF        VALUE           PROB
Jarque-Bera                       2    27163717.271           0.0000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                             DF        VALUE           PROB
Breusch-Pagan test                7       11326.445           0.0000
Koenker-Bassett test              7         225.065           0.0000
================================ END OF REPORT =====================================
```

Figure 5: The figure depicts the Summary of Spatial exogenous regressor model

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL TWO STAGE LEAST SQUARES
--------------------------------------------------
Data set            :     unknown
Weights matrix      :     unknown
Dependent Variable  :   ln(price)                Number of Observations:       66641
Mean dependent var  :      4.4648                Number of Variables   :           8
S.D. dependent var  :      0.9297                Degrees of Freedom    :       66633
Pseudo R-squared    :      0.3432
Spatial Pseudo R-squared:  0.2975

------------------------------------------------------------------------------------
            Variable     Coefficient       Std.Error     z-Statistic     Probability
------------------------------------------------------------------------------------
            CONSTANT       3.5197044       0.0589380      59.7187254       0.0000000
 host_listings_count       0.0001624       0.0000079      20.5086214       0.0000000
            bedrooms       0.3086327       0.0046738      66.0351291       0.0000000
review_scores_rating      -0.1628136       0.0089405     -18.2108369       0.0000000
                beds       0.1283913       0.0033714      38.0828411       0.0000000
review_scores_cleanliness       0.1367888       0.0089555      15.2743500       0.0000000
       d2westminster      -0.0461465       0.0009190     -50.2114210       0.0000000
        W_ln(price)        0.1622832       0.0120023      13.5210506       0.0000000
------------------------------------------------------------------------------------
Instrumented: W_ln(price)
Instruments: W_bedrooms, W_beds, W_d2westminster, W_host_listings_count,
             W_review_scores_cleanliness, W_review_scores_rating
================================ END OF REPORT =====================================
```

Figure 6: The figure depicts the Summary of Spatial Lag model

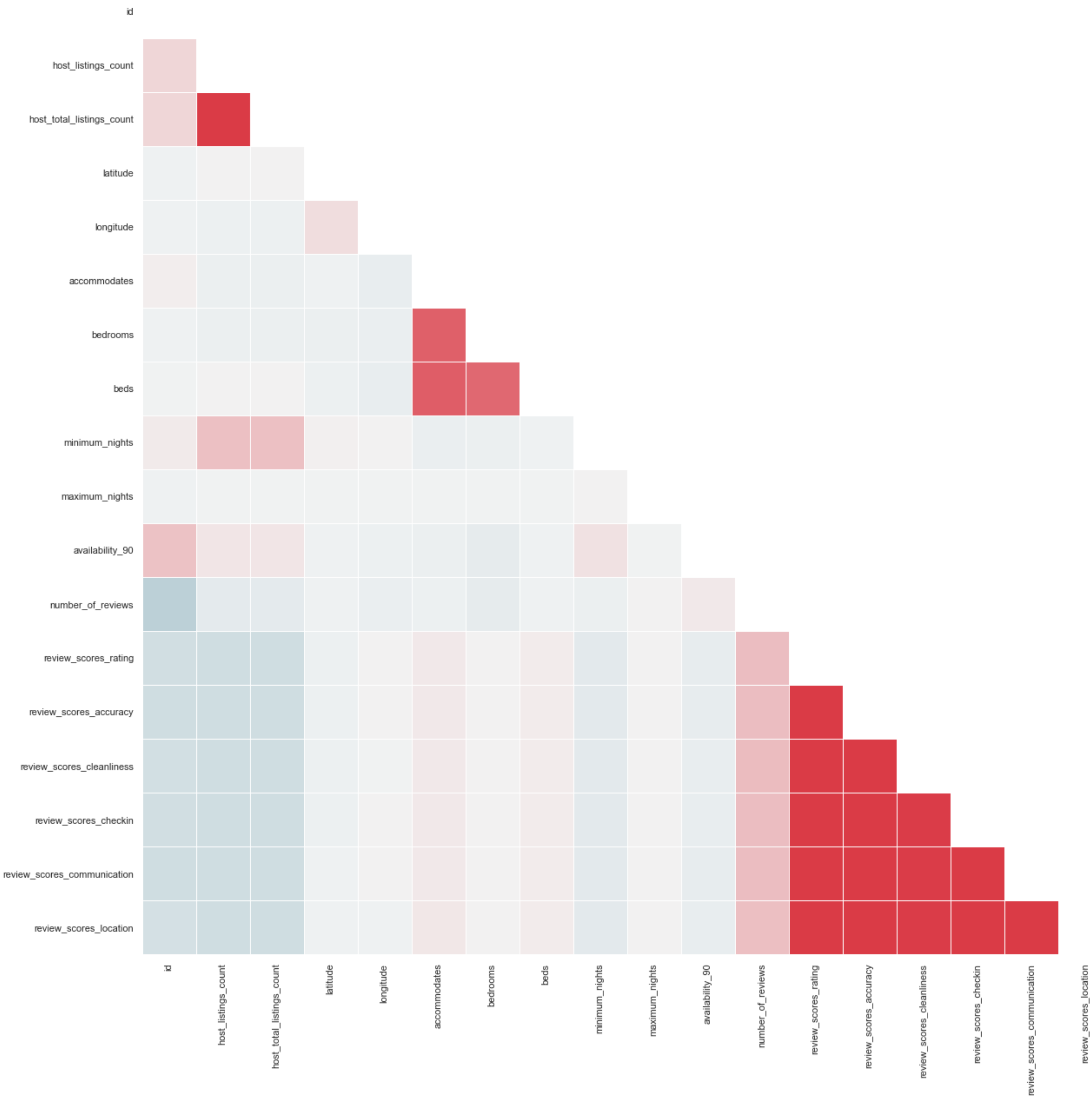

Figure 7: The figure depicts the correlation between variables in the London airbnb dataset

Figure 8: The figure depicts the correlation between variables in the London airbnb dataset

References

[1] Z. X. Ang Zhu, Rong Li, "Machine learning prediction of new york airbnb prices," p. 2, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9253078

[2] D. K. A. C. A. G. R. C. E. M. P. Jasleen Dhillon, Nandana Priyanka Eluri, "Analysis of airbnb prices using machine learning techniques," 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9376144

[3] Y. Li, "Price recommendation on vacation rental websites," 2017. [Online]. Available: Proceedingsofthe2017SIAMInternationalConferenceonDataMining. SocietyforIndustrialandAppliedMathematics

[4] L. N. Pouya Rezazadeh Kalehbasti and H. Rezaei, "Airbnb price prediction using machine learning and sentiment analysis," p. 2, 2019. [Online]. Available: https://arxiv.org/pdf/1907.12665.pdf

[5] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of melbourne city, australia," pp. 35–37, 2019. [Online]. Available: https://ieeexplore-ieee-org.jproxy.nuim.ie/stamp/stamp.jsp?tp=&arnumber=8614000&tag=1

[6] J. W. Hujia Yu, "Real estate price prediction with regression and classification," pp. 35–37, 2016. [Online]. Available: http://cs229.stanford.edu/proj2016/report/WuYu_HousingPrice_report.pdf

[7] Q. P. T. Y. L. G. Yang Li, "Reasonable price recommendation on airbnb using multi-scale clustering," 2016. [Online]. Available: https://ieeexplore.ieee.org/document/7554467

[8] S. Sheppard and A. Udell, "Do airbnb properties affect house prices," 2016. [Online]. Available: WilliamsCollegeDepartmentofEconomicsWorkingPapers