AIRBNB INSIGHTS: MULTI-REGRESSION ANALYSIS OF PRICE FACTORS IN TOP 10 MOST VISITED US STATES

Submitted by

HARSHAN RAGAVANDAR RAMESH	**11055132**
MUHAMMAD MIQDAD	**11044913**
SRIKAR VARMA NADIMPALLI	**11053614**
PRIYA BHAVANA V N DWARAM	**11055208**

Stetson-Hatcher School of Business

FEBRUARY 2024

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

CHAPTER - 1

INTRODUCTION

1.1 BACKGROUND

Our project looks at Airbnb prices in ten big U.S. cities, studying lots of different things like number of bedrooms and beds. We want to find out what makes prices go up or down and if we can use math to predict them. While we're mostly interested in what's happening now, we know it's hard to guess what will happen in the future. By looking at many different factors, we hope to help both hosts and guests make better choices. Even though we're only looking at ten cities, what we find might help in other places too. We looked at data from 2021 to 2023, after COVID-19, to understand how things changed. Our main goal is to give people the information they need to do well in the Airbnb world.

1.2 PREVIOUS STUDY

The advent of technology, particularly the internet, has brought about significant shifts across various sectors, including the way people travel and seek accommodations. Platforms such as Airbnb have revolutionized the hospitality industry, prompting researchers to investigate the factors influencing pricing strategies. Below is a concise overview of three studies that explore Airbnb pricing dynamics using different methodologies:

Study 1: This research focuses on understanding the determinants of Airbnb prices in Metro Nashville, Tennessee. It employs two statistical methods: the General Linear Model (GLM) and Geographically Weighted Regression (GWR). GWR proves superior in capturing how prices vary across different areas of the city. ("Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach by Dr. zhihua zhang") [1]

Study 2: This study examines the factors contributing to price disparities among Airbnb listings in Spanish cities. Utilizing multilinear regression analysis, it identifies key influencers such as property size, guest ratings, and available amenities. Furthermore, the study highlights significant price discrepancies observed across various cities in Spain. (Driving Factors Behind Airbnb Pricing A Multilinear Regression Analysis by Jonathan Herrg˚ ard, JohanFl¨ ojs) [2]

Study 3: This research aims to predict Airbnb listing prices in major cities like New York City, Paris, and Berlin using advanced computational models. Techniques such as Random Forest and neural networks are employed, resulting in accurate price predictions. Additionally, optimizing data preprocessing techniques enhances the performance of these predictive models. (Predicting Airbnb Listing Price Across Different Cities by Yuanhang Luo) [3] These studies underscore the complexity of Airbnb pricing, emphasizing the importance of location, property attributes, and advanced analytical tools in understanding and forecasting pricing dynamics. By unraveling these factors, researchers can contribute to the development of more effective pricing strategies and predictive models, thereby enhancing decision-making processes in the hospitality sector.

1.3 PURPOSE

Our purpose with this project is to understand the pricing of Airbnb short term rental spaces in ten big cities of United States with an aim to better comprehend the driving factors behind the prices of short-term rental spaces like that of Airbnb. We analyzed a huge set of data which had more than 80 variables which included different factors, some of which were quite essential like the bedroom numbers, accommodates and number of beds and there were some not so useful factors in the variable set which through our analysis and acumen weren't of much influence for understanding and building a good regression model. The variables we used are not useful for understanding the current market of Airbnb pricing for these big cities but can be used for every other short term rental space of every other region in the world. It can also be insightful for the rental space providers, who could observe these factors and offer more of the same services and amenities in order to get a better price out of their rental space. Customers on the other hand can understand the factors behind the high prices and can make more efficient choices. Moreover, technically this project will get into the use of linear regression, involving its basic methodology, multiple linear regression, collinearity, and residual analysis to not only get a better model but also to understand the use of these technics for getting a better model for understanding effects of independent variable on dependent variable.

1.4 FUTURE SCOPE AND LIMITATION

Our model is based on the current market situation with the assumption that the market is free and efficient. There might come times when the demand is of the highest level and their prices, as per the free market, get higher and higher as per the law of supply and demand. So, the host might raise the prices which would not be explained or understood by this model if this model is used for future studies on similar data for the future on the short-term rental spaces. We have tried our best to get as much data as possible and incorporate as much as possible variables which have had a significant effect on the pricing of the rental spaces but still there might be situations where this model won't be necessarily the best model to understand the effects of provided variables on short term rental spaces.

For physical location, we have only incorporated data from ten big cities of US because: i, they were the most visited cities here in US. ii, the data would get huge for a project of this magnitude if we involved more data from other cities. While the cities were limited the variables involved were not limited and we used all the available variables to see their effect on our Y variable, which is Pricing for us. So, even if we would want to understand the impacts of these variables on any other location we can most probably easily do so with this model.

Our data is based on time period from 2021 to 2023 for this project. Though most of this period can be termed as economically normal but one can understand the impact of COVID from 2019 and the spillover effect of that macro economical event. We did understand about this economical event but in the meantime, we wanted our data and the associated model to be as up to date as possible because with the new technologies there are always changing amenities and facilities in this sector of the business. So, we assumed this time period to be of a normal macro economical background in order to better understand the effects of this model and effects of the variables involved in this project

1.5 PROBLEM STATEMENT

The problem statement of this study is:

• Which are the main driving factors behind Airbnb pricing in Top 10 tourist place in USA.

These statement together with previous studies will form the model for the research done within this research.

CHAPHTER – 2

THEORY

2.1 MICRO ECONOMIC THEORY OF SUPPLY & DEMAND

In microeconomics, we talk about how supply and demand work together. The supply curve shows that when prices go up, businesses tend to make more of a product or service, assuming everything else stays the same. This is called the law of supply. On the other hand, the demand curve shows that when prices go up, people usually want to buy less of something, assuming everything else stays the same too. This is called the principle of demand. Where these two curves meet is called the equilibrium point, where supply and demand are balanced. If anything changes in either curve, like prices or people's preferences, it can move this equilibrium point.

2.2 HEDONIC PRICING MODEL

A hedonic pricing model uses a type of math called linear regression to guess how much something should cost based on what it's like. This is often used in real estate to figure out what makes a house worth a certain amount of money. The math looks at all the different things about the house, like how big it is or where it's located, and decides how much each of those things should affect the price. There are different ways to do this math, some more complicated than others, but for this study, we're sticking to the simpler linear method.

CHAPTER – 3

METHODS

3.1 DATA SET

We gathered our data from "Opendatasoft" [4], a project with a clear mission to offer insights and advocacy regarding Airbnb's influence on residential areas. This platform serves as a reliable source, known for its accuracy in providing data. Through various stages and platforms, our data collection efforts culminated in sourcing our information from Inside Airbnb, ensuring a dependable and accurate dataset for our analysis.

3.1.1 CATAGORIGAL GROUPING

In our dataset, we had both numbers (quantitative variables) and categories (qualitative variables). We kept the numbers as they were, but for the categories, we needed to organize them into groups and then turn them into numbers. This was done by creating dummy variables. For example, when we looked at the types of beds available, we grouped them into two categories: "Real Bed" and "Other Bed." We did the same for all other categories where there were too many options compared to the number of things we were studying. This helped us analyze the data more effectively.

3.1.2 VARIABLE SELECTION

The variables included in the initial data set, “Amenities”, contained small informational texts about the amenities of the listing. Amenities explains what appliances, services and facilities that are included in the stay. To represent these variables as their own we scraped these texts and created dummy variables for those that were of particular interest. Examples of variables found here are TV, Washer, AC, Superhot and more.

3.2 INTIAL SELECTION OF VARIABLES

Our data set is too huge. Each of our data point is explained by 85 different variables, so we decided to use subset of the initial data set. In this research we targeted on consumer values & micro-economic factors, variables those fit this criterion will be taken consideration while creating this subset. Variables was removed, based on domain knowledge.

We considered variables on 3 major factors:

1. Variables about the main physical factors of the accommodation.
2. Host profile and characteristics.
3. Amenities and general properties

Variables	Range	Description
Accommodates	1-16 people	Number of people the accommodation can house during the stay
Bathroom	0-8 bathrooms	Number of bathrooms
Bedrooms	0-10 bedrooms	Number of bedrooms
Beds	0-16	Number of beds
Days as Host	11-2942 days	Number of days since the host published its first listing on the platform
Host Response Rate	0-100%	Describes how many booking requests the host re- Spon
Minimum Nights	1-200 days	Minimum number of nights allowed per booking
Number of reviews	1-600 reviews	Number of reviews on the listing
Review Scores Rating	0-100	Average review score of the listing
Reviews Per Month	0.02-30 reviews/month	The number of reviews left on the listing per month
Check-in 24	Features	0,1 Whether it is possible to check in 24/7 or not
Exact Location	Features	0,1 Whether the exact location of the property is reveled before booking or not
Instant Booking	Features	0,1 Whether it is possible to book without conversation with host before confirmation.
super host	Features	0,1 Whether the host is a super host or not
Apartment	Property Type	0,1 Whether it is an apartment or not
House	Property Type	0,1 Whether it is a house or not
Entire Home	Room Type	041 Whether the accommodation is a entire home or not
Private Room	Room Type	01 Whether the accommodation is just a private room
Shared Room	Room Type	0,1 Whether the accommodation is a shared room or not

Figure 3.2*: Showing the quantitative variables included in the study.*

3.3 REGRESSION MODEL

We have used multiple regression analysis in this model which used the least square approach, and this fits well with the hedonic pricing assumption. As this model is elaborating the effects of different variables on pricing of the Airbnb rental spaces, the response variable would be price per night.

3.4 ASSUMPTIONS ON LINEAR MODEL

Several assumptions are there to be made when using or applying a linear regression model and these will have to be validated. The assumptions are [5]:

1. **Linearity:** We assume that there are linear relationships between the dependent and independent variables.
2. **Independence**: The observations in the data set are independent from each other in this assumption.
3. **Homoscedascity:** This assumption is that constant variance of the errors across all levels of the independent variables is there.
4. **Normality**: The errors follow a normal distribution with mean zero in this assumption.
5. **No Multicollinearity:** This assumption explains that the independent variables

CHAPTER – 4

RESULTS

4.1 INITAIL MODEL

```
Call:
lm(formula = Price ~ Accommodates + Bathrooms + Bedrooms + Beds +
    Security.Deposit + Cleaning.Fee + Minimum.Nights + Reviews.per.Month +
    Review.Scores.Rating + Room.Type + Bed.Type + Cancellation.Policy +
    Host.Listings.Count + Host.Response.Rate + TV + Dryer + Washer +
    AC + Free.parking.on.premises + Pool + Elevator + X24.hour.check.in +
    Gym, data = K)

Residuals:
    Min      1Q  Median      3Q     Max
-567.22  -41.20   -8.31   25.70  866.53

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                 -56.718218   5.391434 -10.520  < 2e-16 ***
Accommodates                  7.444091   0.300296  24.789  < 2e-16 ***
Bathrooms                    39.235173   0.724668  54.142  < 2e-16 ***
Bedrooms                     34.258265   0.735097  46.604  < 2e-16 ***
Beds                         -1.558545   0.457228  -3.409 0.000653 ***
Security.Deposit             -0.014963   0.001900  -7.875 3.45e-15 ***
Cleaning.Fee                  0.594860   0.008074  73.673  < 2e-16 ***
Minimum.Nights               -0.701633   0.054853 -12.791  < 2e-16 ***
Reviews.per.Month            -3.692431   0.158272 -23.330  < 2e-16 ***
Review.Scores.Rating          0.944636   0.047047  20.078  < 2e-16 ***
Room.TypePrivate room       -33.075564   0.833763 -39.670  < 2e-16 ***
Bed.TypeReal Bed              3.607686   2.379767   1.516 0.129528
Cancellation.Policymoderate  -4.289712   0.945850  -4.535 5.76e-06 ***
Cancellation.Policystrict    -0.578290   0.917155  -0.631 0.528353
Host.Listings.Count          -0.029616   0.007503  -3.947 7.92e-05 ***
Host.Response.Rate           -0.161091   0.023371  -6.893 5.52e-12 ***
TV1                           7.533459   0.829373   9.083  < 2e-16 ***
Dryer1                        0.387883   1.154879   0.336 0.736974
Washer1                       1.195178   0.847119   1.411 0.158286
AC1                           2.125307   0.773818   2.747 0.006025 **
Free.parking.on.premises1   -20.043929   0.695585 -28.816  < 2e-16 ***
Pool1                         5.651383   1.233805   4.580 4.65e-06 ***
Elevator1                    19.567010   0.988771  19.789  < 2e-16 ***
X24.hour.check.in1           -2.669928   0.725655  -3.679 0.000234 ***
Gym1                          2.750728   1.369054   2.009 0.044518 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 83.76 on 69619 degrees of freedom
  (5745 observations deleted due to missingness)
Multiple R-squared:  0.5626,    Adjusted R-squared:  0.5625
F-statistic:  3732 on 24 and 69619 DF,  p-value: < 2.2e-16
```

Figure 4.1: *Illustrating all included variables, their estimated coefficients, errors, t-values and p-values as well as model R-square and F-statistics in the initial model.*

In the initial model, we have 34 variables which are categorical and quantitative. With the initial model, we have a few variables that are not significant even though which has causality. With further modeling, we will remove the variables that are not significant and build a model where all the independent variables are significant. For this model, we have an Adjusted R-square of 0.5809 and a P-value of 2.2e-16.

Now comes the interpretation of some of the variables.

***Accommodates*:** Here the coefficient is 7.67, which tells us that with one more accommodate involved the price is going to increase by $7.67 for the room in Airbnb.

***Bathroom & Bedroom*:** Here the coefficients for bathroom and bedroom are 36.96 and 34.11 respectively. This says that with an extra bathroom and bedroom in the Airbnb house, the price is going to increase by $37 and $34 respectively. With interaction between the bedrooms and bathrooms, we found out that when both were combined the price increased $8.10.

With *TV* in the house the price is increased by $8.56, with Pool in the house the price is increased by $4, and with an Elevator as an amenity included in the House the price increased by $19.7. Interestingly all the amenities have an impact on the Price of Airbnb.

4.2 LINEARITY

For linearity, we have plotted graphs for all the key variables and as shown in figure 3, we can see the linearity from the graphs and see how they are moving in the same direction. Categorical Values cannot have graphical representation as they are in qualitative format, hence we can see them in 0s and 1s but for the representation, they are based on these values which are mostly binomially shown, which tells us there are linearly related to each other

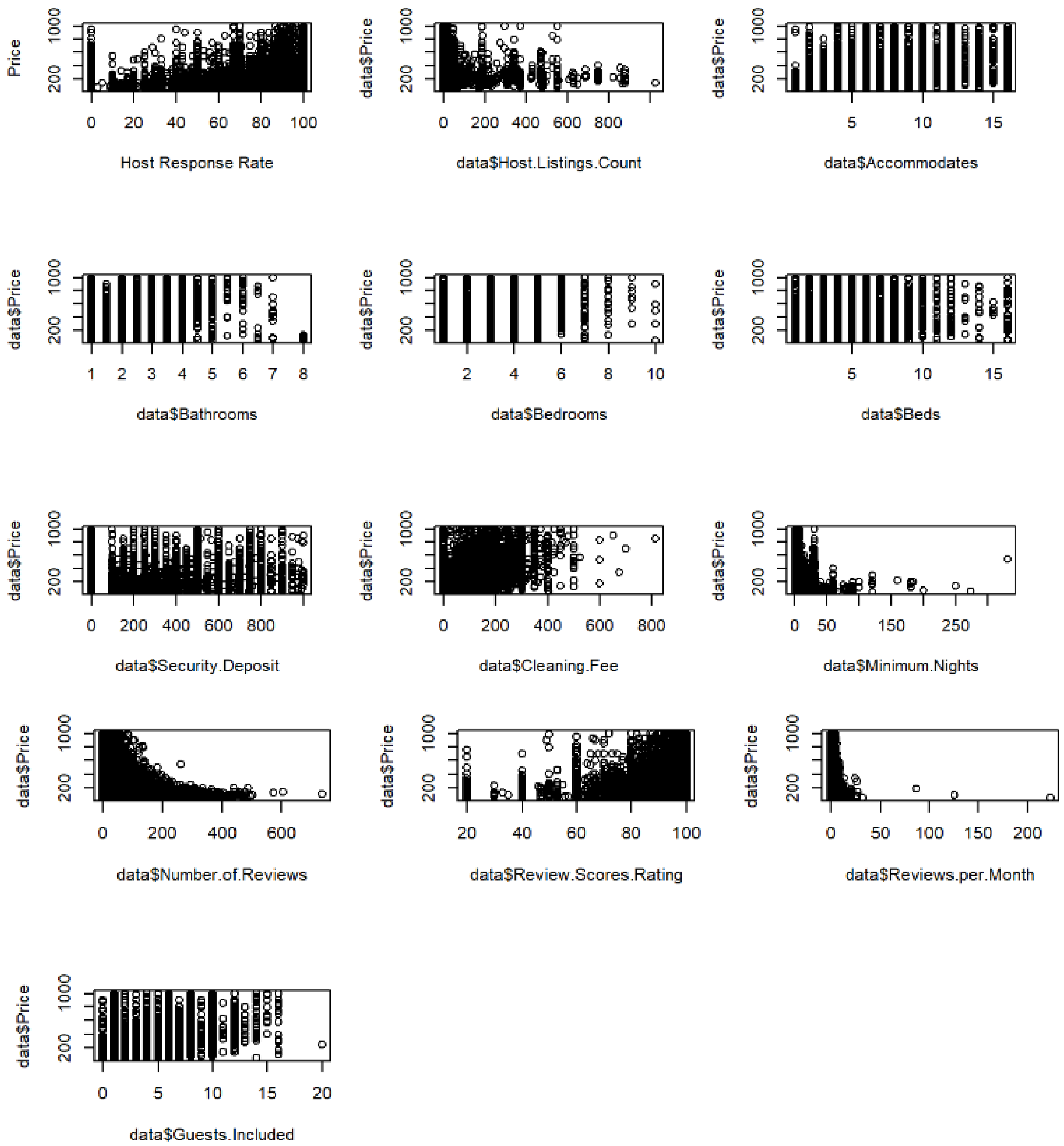
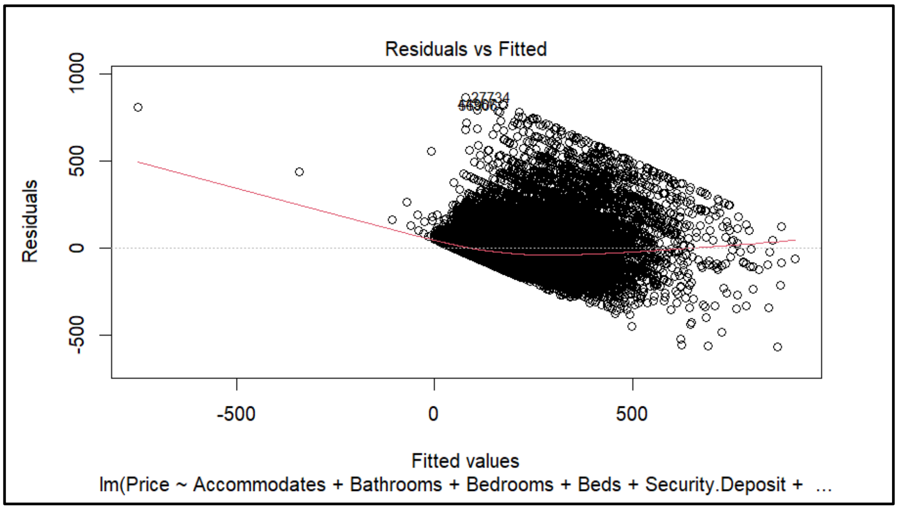


Figure 4.2: *Scatter plots for Variables with Price.*

4.3 RESIDUAL ANALYSIS

In the below figure3 we have graphs for residuals and fitted values from the model. From the Graph we can see that the residuals are surrounded zero, and the residuals are greater with the predicted price increased.

This says that the residuals are following heteroscedasticity. We definitely should overcome heteroscedasticity by either taking log or square of the variables. We did take the squared term for the Price and overcome with heteroscedasticity.



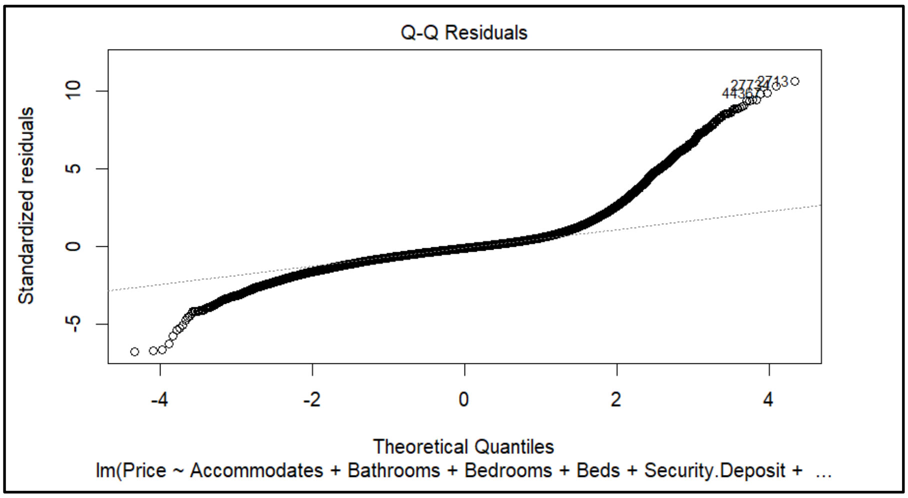
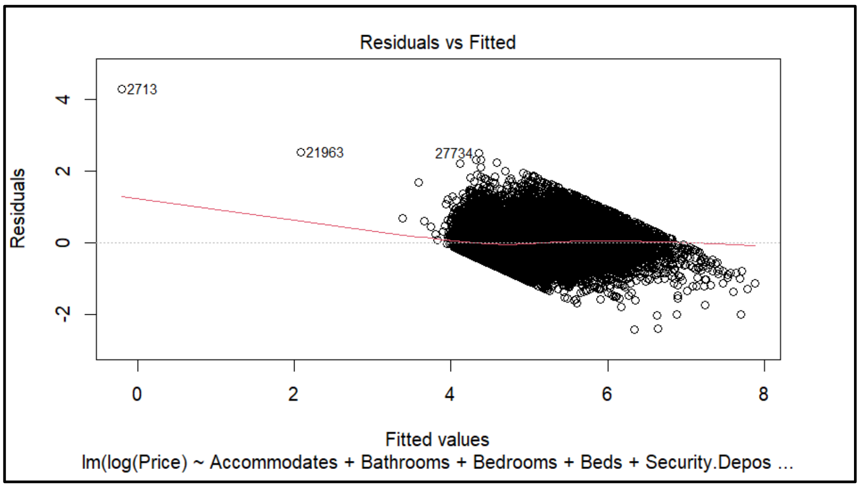


Figure 4.3: *Standardized residuals of the model with fitted values and residuals*

4.4 TRANSFORMATION

We have transformed the model to log transformation of the dependable variable (i.e, Price).



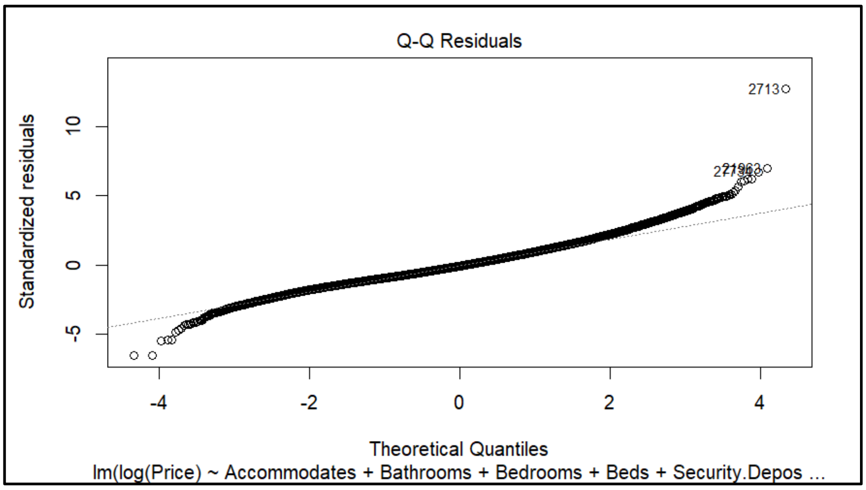


Figure 4.4: *Transformed Model*

After transforming the dependent variable in the data set, we can clearly see that the residuals are evenly spread and have overcome the variances in residuals just by taking the quadratic of the Dependent variable.

The new Q-Q plot has improved from the previous model by taking the Log term of the Dependent Variable. Thus, we can say that the model is now heteroscedasticity free and have normality in the model.

4.5 DEPENDENCIES B/W VARIABLES

4.5.1 CORRELATION

Determining the extent to which each predictor contributes to the response variable may be challenging due to the correlation between different regressors. It is crucial to eliminate any regressors that have a significant degree of association. Some of the variables in the model have high correlations with one another.

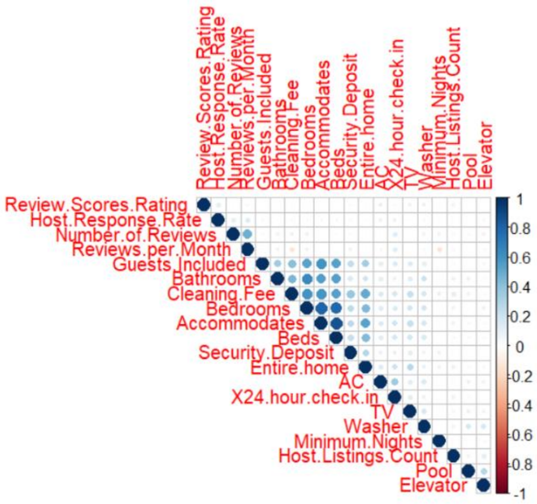


Figure 4.5.1: *Correlation matrix for all the variables against Price.*

4.5.2 MULTICOLLINEARITY

For multicollinearity we usually use the Variance Influence factor (VIF). VIF is a measurement for multicollinearity with the independent variables and Multiple Regression Model. Usually for the value higher than 10 we can say as a high VIF value. Looking at the figure6 we can say that all the variables in our model are less than our ideal.

	GVIF	Df	GVIF^(1/(2*Df))
Accommodates	5.251634	1	2.291644
Bathrooms	1.949439	1	1.396223
Bedrooms	4.062209	1	2.015492
Beds	4.327982	1	2.080380
Security.Deposit	1.207542	1	1.098882
Cleaning.Fee	2.019341	1	1.421035
Minimum.Nights	1.042623	1	1.021089
Reviews.per.Month	1.109966	1	1.053549
Review.Scores.Rating	1.054248	1	1.026766
Room.Type	1.596699	1	1.263605
Bed.Type	1.012772	1	1.006366
Cancellation.Policy	1.198925	2	1.046401
Host.Listings.Count	1.066123	1	1.032532
Host.Response.Rate	1.052330	1	1.025831
TV	1.124080	1	1.060226
Dryer	1.572052	1	1.253815
Washer	1.580263	1	1.257085
AC	1.268975	1	1.126488
Free.parking.on.premises	1.199757	1	1.095334
Pool	1.370622	1	1.170736
Elevator	1.439171	1	1.199655
X24.hour.check.in	1.246428	1	1.116435
Gym	1.633381	1	1.278038

Figure 4.5.2*: VIF table for all the variables in the multiple regression model.*

4.6 FINAL MODEL

Our final model after transformation is presented below:

$$Log\ (Price) = \beta 0 + \beta Accommodates \cdot x1 + \beta Bathrooms \cdot x2 + \beta Bedrooms \cdot x3 + \beta Beds \cdot x4 + \beta Minimum\ Nights \cdot x5 + \beta Cleaning\ Fee \cdot x6 + \beta Reviews\ Per\ Month \cdot x7 + \beta Review\ Scores\ Rating \cdot x8 + \beta Room\ Type \cdot x9 + \beta Bed\ Type \cdot x10 + \beta Host\ Respond\ Rate \cdot x11 + \beta Free\ Parking\ on\ Premises \cdot x12 + \beta TV \cdot x13 + \beta Dryer \cdot x14 + \beta Gym \cdot x15 + \beta Elevator \cdot x16 + \beta Free\ parking\ on\ premisis \cdot x17$$

```
Call:
lm(formula = log(Price) ~ Accommodates + Bathrooms + Bedrooms +
    Beds + Cleaning.Fee + Security.Deposit + Minimum.Nights +
    Reviews.per.Month + Review.Scores.Rating + Room.Type + Bed.Type +
    Host.Response.Rate + TV + Dryer + Free.parking.on.premises +
    Elevator + Gym, data = K)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4272 -0.2520 -0.0274  0.2238  4.3210

Coefficients:
                           Estimate Std. Error  t value Pr(>|t|)
(Intercept)               3.941e+00  2.368e-02  166.398  < 2e-16 ***
Accommodates              3.919e-02  1.328e-03   29.516  < 2e-16 ***
Bathrooms                 1.173e-01  3.193e-03   36.739  < 2e-16 ***
Bedrooms                  1.404e-01  3.249e-03   43.202  < 2e-16 ***
Beds                     -1.301e-02  2.025e-03   -6.424 1.34e-10 ***
Cleaning.Fee              2.406e-03  3.500e-05   68.750  < 2e-16 ***
Security.Deposit          6.347e-05  8.309e-06    7.639 2.22e-14 ***
Minimum.Nights           -3.418e-03  2.429e-04  -14.074  < 2e-16 ***
Reviews.per.Month        -2.061e-02  6.928e-04  -29.743  < 2e-16 ***
Review.Scores.Rating      4.596e-03  2.068e-04   22.225  < 2e-16 ***
Room.TypePrivate room    -3.881e-01  3.674e-03 -105.622  < 2e-16 ***
Bed.TypeReal Bed          7.742e-02  1.055e-02    7.341 2.14e-13 ***
Host.Response.Rate       -7.376e-04  1.031e-04   -7.155 8.47e-13 ***
TV1                       6.536e-02  3.666e-03   17.827  < 2e-16 ***
Dryer1                    2.395e-02  4.259e-03    5.623 1.88e-08 ***
Free.parking.on.premises1 -1.077e-01  2.981e-03  -36.137  < 2e-16 ***
Elevator1                 1.223e-01  4.319e-03   28.311  < 2e-16 ***
Gym1                      3.647e-02  5.541e-03    6.582 4.67e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3712 on 69626 degrees of freedom
  (5745 observations deleted due to missingness)
Multiple R-squared:  0.6252,    Adjusted R-squared:  0.6251
F-statistic:  6833 on 17 and 69626 DF,  p-value: < 2.2e-16
```

Figure 4.6: *Summary including final variables in the model and the r-square and adj.R-square and f-stat with standard error in the final model.*

4.7 DISCUSSION

Our final choice of model includes 19 independent variables of which 11 of them are statistically significant with a p-value of ≈ 0.

4.7.1 POSITIVELY INFLUENTIAL FACTORS

Our model explains a lot of positive affecting variables on the Price. Such of those are Accommodates, which have a core depending intuition for the independent variable Price. When the number of accommodates increases then obviously the rooms and bathrooms should be more too! Which is why this is a positive influencing factor. There was almost a $4 increase in price for one accommodation per night. Bathrooms and Bedrooms were also the most important factors for this study. We have seen almost $12 and $14 increase for every other bedroom and bathroom added. People had to pay an extra cumulative of $30 for every bedroom and bathroom per night added. There were a lot of features for Airbnb like “Security. Deposit” and “Review Score Ratings” even though important, but didn’t have much effect on the Price with a $0.006 and $0.46 increase per night. People were interested in the type of bed they chose. With every different bed, there is a significant change of $7 upwards per night. The customers wanted amenities like “A TV”, “Dryer”, “Elevator”, and “Gym”. These facilities per night would cost $25.But for an Elevator alone there was a $12 increase which means that people were willing to have an elevator in their home for a luxurious stay-over.

4.7.2 NEGATIVELY INFLUENTIAL FACTORS

In our model, rigorous data cleaning and model refinement procedures were instrumental in eliminating most negative coefficients, with a focus on factors demonstrating significant p-values. Nevertheless, six negative influences persisted, warranting further investigation: Beds, minimum nights, Review per month, room type, host response rate, and on-premises parking. While beds are typically seen as a positive contributor to rental prices, the nuanced impact of bed type and quantity, such as air mattresses versus memory foam, introduces variability. Similarly, a higher minimum night's requirement may discourage potential guests seeking shorter stays, thus diminishing rental attractiveness. The surprising negative correlation between a higher number of reviews and rental prices suggests a deeper relationship, possibly indicating more frequent rentals and subsequently lower prices. Room type comparisons highlight the inherent pricing differences between single rooms and entire spaces, with the former typically commanding lower rates. Negative coefficients associated with host response rates may signify hosts managing cheaper rentals facing higher demand, resulting in more inquiries but potentially lower response rates due to increased busyness. Lastly, the negative influence of on-premises parking could be attributed to accessibility challenges or safety concerns, influencing guest perceptions and impacting rental prices. Delving into these intricacies not only enhances the refinement of our predictive model but also underscores the importance of holistic insights derived from both data analysis and comprehensive literature review.

CHAPTER – 5

CONCLUTION

This study explains how pricing is affected for Airbnb in the top 10 most visited cities in the USA. In the beginning, we had a total of 88 independent variables and after cleaning the data we ended up with 16 independent variables for the multiple regression analysis. With this analysis, we had an in-depth understanding of the qualitative and quantitative variables. With this model, any user can predict the price of an Airbnb accommodation with all the variables affecting the price and can pick the best hotel / private room in those cities that they want to rent.

A hedonic pricing model could be created because the final model had 16 physical and non-physical attribute factors that had a big impact on the price. Additionally, we could distinguish between cities, supporting our hypothesis that the cost of living varies by city. Aside from the city, the most important factors of cost were the size of the accommodations, including the number of bedrooms, bathrooms, and private housing; they also included review ratings and whether or not the accommodations featured a pool or an elevator. However, more investigation into the small differences in how various factors influence price in different places and how different factors explain price at different price points can lead to a far better understanding of pricing in Airbnb.

CHAPTER-6

APPENDIX – A

6.1 REFERENCES:

[1] Dr. zhihua zhang., Rachel., Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. Available at: https://www.researchgate.net/publication/319865700_Key_Factors_Affecting_the_Price_of_Airbnb_Listings_A_Geographically_Weighted_Approach

[2] Jonathan Herrg˚ ard., JohanFl¨ ojs., Driving Factors Behind Airbnb Pricing A Multilinear Regression Analysis. Available at: https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1823748&dswid=-3142

[3] Yuanhang Luo., Predicting Airbnb Listing Price Across Different Cities.

[4] OpenDataSoft (2023). *Airbnb Listings*. Available at: https://data.opendatasoft.com/explore/dataset/airbnb-listings%40public/table/?disjunctive.host-verificationsdisjunctive.amenities&disjunctive.features

[4] Opendatasoft (2023) *Customers - Opendatasoft*. Available at: https://www.opendatasoft.com/en/customers/

REVIEWER COMMENTS AND QUERIES FROM PRESENTATION

No Queries was given by the reviewers

INDIVIDUAL CONTRIBUTION

INDIVIDUAL CONTRIBUTION OF SRIKAR VARMA NADIMPALLI

- Data Pre-Processing.
- Data Transformation.
- Developing the model R Studio
- Presentation & Report.

INDIVIDUAL CONTRIBUTION OF PRIYA BHAVANA DWARAM

- Data Pre-Processing.
- Data Transformation.
- Developing the model R Studio
- Presentation & Report.

INDIVIDUAL CONTRIBUTION OF MUHAMMAD MIQDAD

- Data pre-Processing.
- Data Transformation.
- Developing the model in R Studio
- Presentation & Report.

INDIVIDUAL CONTRIBUTION OF HARSHAN RAGAVANDAR R

- Data pre-Processing.
- Data Transformation.
- Developing the model R Studio
- Presentation & Report.