Spring 2020

Price Prediction in the Sharing Economy: A Case Study with Airbnb data

Brandon McNeil
University of New Hampshire

Undergraduate Honors Thesis

Price Prediction in the Sharing Economy: A Case Study with Airbnb data

by

Brandon McNeil

Peter T. Paul College of Business and Economics

University of New Hampshire

Advisor:

Dr. Christopher Glynn

Assistant Professor of Decision Sciences

May 2020

Abstract:

Many sharing-economy companies like Airbnb have their profits rely on the dynamic pricing market that they participate in. Airbnb hosts can set their own price based on what they deem the market will buy. Recent research argues that almost all hosts fail to maximize their potential profit due to poorly pricing their listing (Gibbs et al., 2018). While previous studies have looked at how specific variables effect the price of an Airbnb listing, this study aims to be the first to group variables separately into two distinct categories based on the host's ability to control that variable. Looking at two U.S. cities with developed Airbnb markets, this study aims to use linear regression analysis to determine the significance that variables inside of the host's control have on price versus variables outside of the host's ability to control. The results show that variables within the host's control appear to have more of an impact on price versus variables outside of the host's control. Also, when variables inside and outside of the host's control are combined, they prove most accurate when predicting the price of a listing.

1. Introduction

Airbnb has grown into one of the largest companies in the sharing economy. As of November 2019, Airbnb had a valuation of $35 billion, over 150 million total number of users, over 6 million global Airbnb listings worldwide, and over 2 million people staying in an Airbnb per night. To go along with these statistics, Airbnb offers listings in over 191 countries with the United States leading all other countries with around 660,000 total listings.[1] Airbnb is unique compared to other companies in the sharing economy due to the peer-to-peer interaction between a host and a consumer and the ability for the host to set their own listing price. The host lists a property which the consumer then chooses out of the thousands of Airbnb listings across one location that best suits the individual's needs. When hosts are setting the price for their listing, Airbnb provides them with a price recommendation. Overtime Airbnb has continued to innovate and improve their price recommendation tool to be more dynamic. Using dynamic pricing provides a more effective price recommendation tool to Airbnb hosts.

Recent research has looked at the significance and effect of different variables on Airbnb listing prices. In this paper, I will categorize Airbnb variables into two types: host- controlled variables and out of host-controlled variables. The main differentiator between the two types of variables is whether the host has the ability to change the specific variable or if that variable is determined by the market. I aim to identify the significance of each of these types of variables and answer the following key question: Do variables controlled by the host or variables not controlled by the host have more of a significance on the listing price? Do out of host-controlled variables have more significance depending on location of the listing? Does obtaining "Superhost" status increase the impact that host-controlled variables have on listing price?

Below I will summarize previous research that has been conducted on Airbnb Dynamic Pricing and the significance of different variables on the listing price. While these studies have never categorized variables into these two types, they present unique findings on each individual variable that will be examined in this study. Then the following hypotheses relating to the stated key research questions will be presented followed with results and discussion.

[1] https://ipropertymanagement.com/airbnb-statistics/

2. Literature Review

2.1 Defining Dynamic Pricing in the Sharing Economy

Over the last decade, Airbnb has grown immensely in the sharing economy. The sharing economy is defined "as a peer-to-peer (P2P) based activity of acquiring, providing, or sharing access to goods and services that is often facilitated by a community-based on-line platform."[2] Airbnb offers a wide variety of lodging accommodations to consumers, through peer-to-peer (P2P) interaction between hosts and consumers. Hosts can set their own prices for each listing they post. However, one of the biggest challenges for Airbnb has been pricing. Gibbs et al. (2018) estimate that Airbnb forfeits 46% of revenues due to inefficient pricing. This loss in revenue is one consequence of hosts not pricing their listings appropriately. More experienced hosts attempt to maximize revenue by implementing a dynamic pricing approach to Airbnb listings. Dynamic pricing allows businesses to maximize profit by adjusting prices continuously in response to demand fluctuations (McGuire, 2015). Listing prices constantly change based on many internal and external variables to meet consumer demand.

2.2 Host-controlled Variables

There has been extensive research on variables that are within the control of the Airbnb host that determine the listing price. For this study, host-controlled variables will be defined as any features of the listing provided by the host and the details of the host. Prior research has associated these factors as potential drivers of price (Chen and Xie, 2017; Gibbs et al., 2017). Listing attributes considered include the type of accommodation, number of rooms, listing size, listing location, view from listing, and listing facilities. The host attributes considered include level of professionalism, years of experience, degree of trustworthiness, and the host's responsiveness. (Ert et al., 2016; Li et al., 2015; Wu, 2016).

2.3 Out of host-controlled Variables

Out of host-controlled variables will be defined in this study as any variables of the listing outside of the control of the host. Prior research on the impact of external variables on Airbnb listing prices focus around the effect of seasonality, day of the week, and social factors. Listing prices fluctuate according to both seasons and day of week, as well as holidays (Gibbs et al. 2018). One study also determined the importance of location through comparisons between a general linear model (GLM) and a geographically weighted regression (GWR) model, with the GWR model proving to be more accurate with a higher adjusted R-squared. (Zhang et al. 2017). Social factors, such as listing review score, host responsiveness, and total number of reviews, were shown to have a correlation with consumers spending more money on an Airbnb listing (Tang & Sangani 2015). It has also been shown that factors relating to listing size, property characteristics, amenities, services, rental rules, and customer reviews significantly affect listing prices (Dogru & Pekin 2017).

[2] https://www.investopedia.com/terms/s/sharing-economy.asp

2.4 Superhost Status

One of the most important host attributes: "Superhost status" is defined by Airbnb as "hosts who provide a shining example for other hosts, and extraordinary experiences for their guests."[3] A host must meet a certain criterion set by Airbnb to receive the "Superhost" badge.[4] Based on previous studies, hosts with a "Superhost" badge typically post their listings at higher prices when they receive more reviews and higher ratings. (Liang et al. 2017). High quality host photos, "Superhost" Status, ratings, and reviews were confirmed to have a significant contribution to listing prices (Liang et al., 2017).

2.5 Progression of Airbnb Pricing Model over time

To understand dynamic pricing for Airbnb listings, it's important to study the progression of Airbnb pricing over the company's history. In 2012, Airbnb offered a price recommendation tool based off simple characteristics related to the listing. To make the price recommendation tool more effective, in 2015, Airbnb factored in expected listing demand into the tool (Hill 2015). This makes the recommended listing price more dynamic. Although the host is recommended prices generated by the tool provided by Airbnb, the decision on the listing price is ultimately up to them.

3. Exploratory Data Analysis

3.1 Data collection

Data collection was conducted through the website: insideairbnb.com, which sources its datasets directly from the Airbnb site. Preliminary data munging has already been conducted by Inside Airbnb to give users the ability to present their research findings and discuss with peers. Inside Airbnb compiles Airbnb data into three separate files for each city to categorize data through listing, calendar, and reviews. For this study, I have chosen the U.S. cities of Boston, Massachusetts and Seattle, Washington for research due to similarity in size and number of listings. For each city, listing data was scraped from three separate time points to encapsulate the entire years' worth of listings in 2019. These dates were January 17th, July 14th, and December 4th. Additional data cleaning was conducted in Microsoft Excel. This included determining which data fields would be involved in this study and removing any duplicate listings that were on the spreadsheet. I then compiled this into one large data set. Once compiled, all duplicate listings were removed. The two cities were combined into one large dataset with approximately 21,100 rows of unique listings between the cities. This study will aim to identify significant variables that are both in and out of the host's control regarding the listing price.

3.2 Variables Defined

For this study, the dependent variable chosen was listing price. The independent variables were selected based on their definition of being in the host's control or out of the host's control. The host-controlled variables that were selected include: # of bedrooms, # of bathrooms, room type, and # of people the listing accommodates. The variables that were deemed out of the host's control included in the study were: city, neighborhood, and number of reviews a listing had. For this study, "Superhost" status was included as an independent variable due to findings from

[3] https://www.airbnb.com/help/article/828/what-is-a-superhost
[4] https://www.airbnb.com/help/article/829/how-do-i-become-a-superhost

Liang et al. (2017), despite it not being classified as a variable within the host's control or out of the host's control. See Variables table below for a description of each variable and whether the variable was determined to be within the host's control.

Variable Name	Host Controlled?	Variable Description
Dependent Variables: - Listing Price	- Y	- The total price (room rate) of each accommodation
Independent Variables: - Bedrooms - Bathrooms - Room Type - Accommodates - # of Reviews - City - Neighborhood - Superhost Status	- Y - Y - Y - Y - N - N - N - N	- # of bedrooms in the listing - # of bathrooms in the listing - Type of room listing offers - # of people that listing is intended to host - # of reviews a listing has received - City location of listing - Neighborhood location of listing - Whether a host has obtained and kept the "Superhost" badge

Figure 1. *Variable Description Table*

3.3 Data Analysis Tools

Afterwards, Extensive statistical analysis was conducted in R statistical software which included linear regression analysis, accuracy measures that included: mean absolute percentage error (MAPE), and an Anova test. It was determined after preliminary data exploration, that linear regression analysis would best show each variable's significance to listing price. Accuracy measures were used to examine each regression model's soundness. Data analysis was then visualized through various R packages including ggplot2, ggfortify, dplyr, and sjPlot. Using these R libraries, visualizations made included boxplots, histograms, and other visualizations that aimed to extract the most information out of the dataset.

3.4 Data Visualization

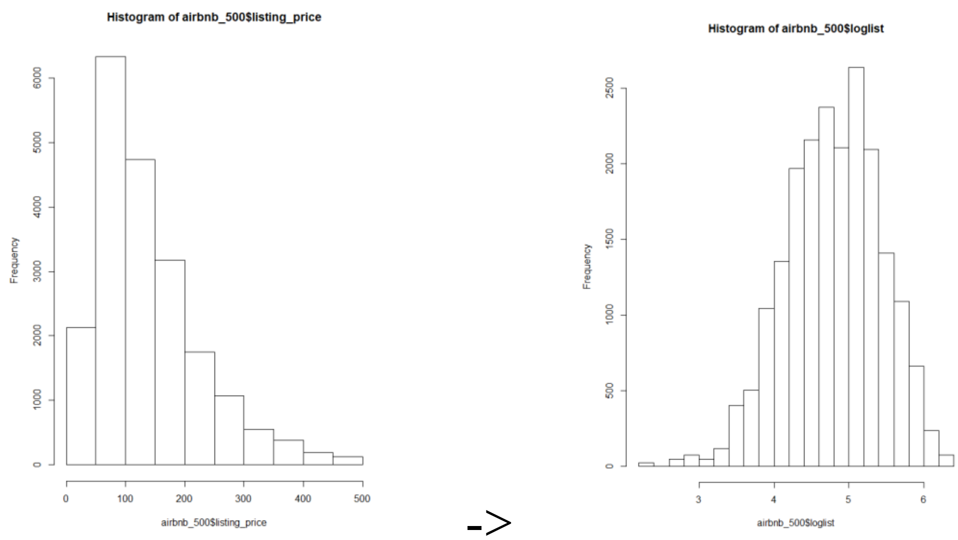


Figure 2. *Histograms of list price (left) and log list price (right)*

In Figure 2, we illustrate listing price variable in the dataset, both the absolute price (left) and the log price (right). Looking at the histogram on the left side of the page which visualizes the listing price distribution across the dataset. The histogram, however, doesn't follow a normal bell-shaped curve due to it skewing to the left side. In order to correct this distribution, the log transform of listing price was taken to normalize the distribution curve which is illustrated in the histogram on the right. This improved distribution will provide a more accurate regression analysis of the data.

Box plots of Independent Variables vs Dependent Variable:

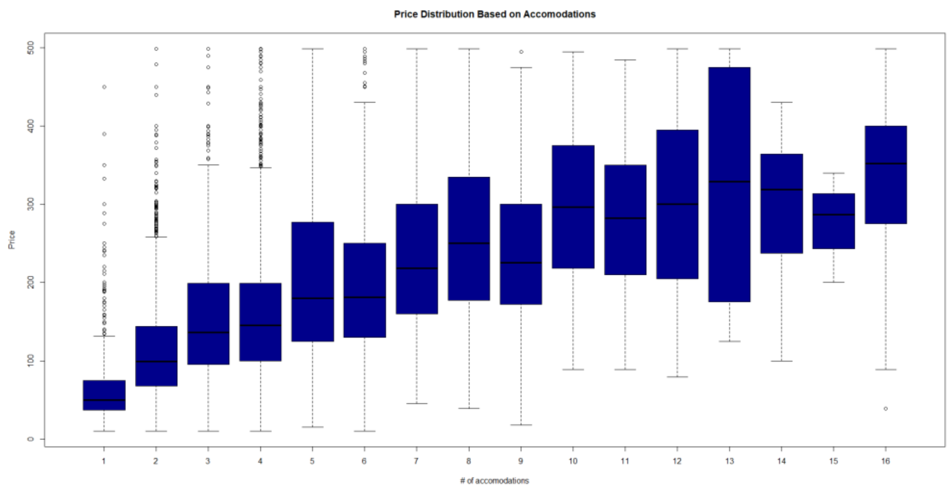


Figure 3. Accommodates/Listing Price Box plot

Looking at Figure 3, as the number of people the listing accommodates increases, the price distribution increases as well. This shows a clear upward trend that there is a direct positive correlation between accommodation and listing price. Based off this boxplot, it would be wise for future listing hosts to pay attention to the number of people their listing accommodates as this will heavily impact the listing price.

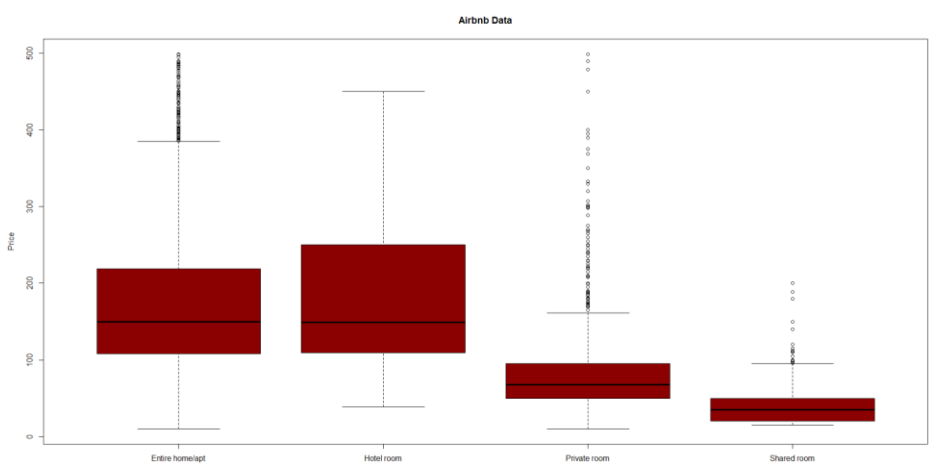


Figure 4. Room Type/Listing Price Box plot

Figure 4 illustrates the distribution of listing prices based off the type of Airbnb unit: Entire Home/Apt, Hotel Room, Private Room, and Shared Room. The price distribution varies substantially based on what the room type. Shared Room type has a much lower price distribution compared to Entire Home/Apt and Hotel Room. This is another important variable that a listing host needs to be aware of when calculating listing price.

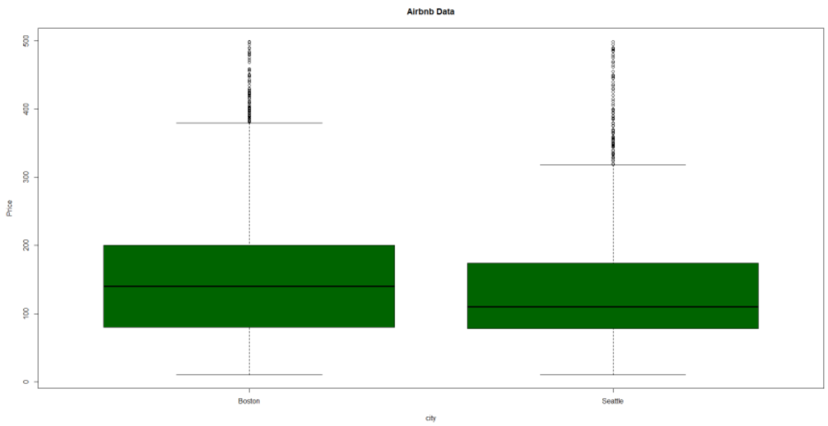


Figure 5. *City/Listing Price Box Plot*

Looking at the price distribution between cities in Figure 5, Boston appears to have not only a wider price distribution, but also a higher average price per listing. While price distribution will vary city to city, it is important for host's to be aware of the city their listing is in has a distinct price distribution.

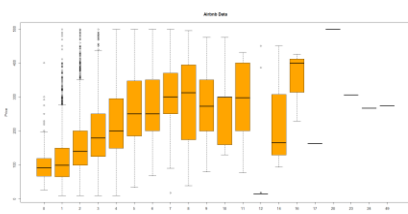


Figure 6. *Beds/Listing Price Box Plot*

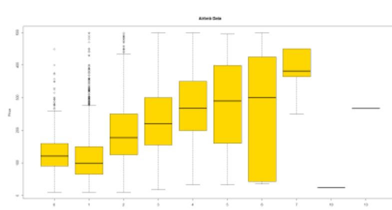


Figure 7. *Bedrooms/Listing Price Box Plot*

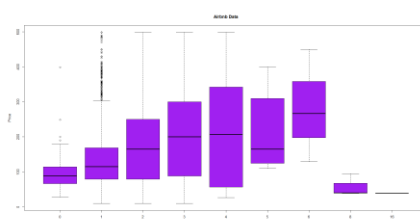


Figure 8. *Beds/Listing Price Box plot*

Number of Beds, Bedrooms, and Bathrooms all appear to have a similar correlation to listing price as they increase (see Figures 6-8 above). Like accommodation, these three variables share a positive correlation to listing price. When a host is calculating listing price, they should factor in the number of beds, bedrooms, and bathrooms that their listing has available as this is shown to increase listing price.

4. Hypotheses

In this section we present three different hypotheses related to the significance that variables controlled by the host and variables outside the host's control have on listing price. Previous studies have examined the individual impact of these specific variables in relation to price, but none have focused on grouping the variables into two distinct groups. These hypotheses aim to answer key research questions with the objective of learning the overall impact that each variable group has on listing price of an Airbnb.

4.1 H1

For the question: Do variables controlled by the host or variables not controlled by the host have more of a significance on the listing price? The hypothesis is that host-controlled variables will have a greater significance on the listing price compared to variables not controlled by the host. We predict that host-controlled variables have less of a degree of variability and limitations compared to variables outside the host's control. Out of host control variables are more sporadic when comparing other listings and have a greater degree of variability meaning they are less accurate price indicators.

4.2 H2:

For the question: Do out of host-controlled variables have more of a significance on listing price depending on the location of the listing? I hypothesis that variables out of the host's control will have a small degree in significance depending on the location of the listing. We predict that certain neighborhoods in a city will have different activity levels which will cause demand and number of review levels to fluctuate. Each city tends to have neighborhoods that are in a more attractive location. These specific neighborhoods will have a greater impact on listing price than others.

4.3 H3:

For the question: Does obtaining "Superhost" status increase the impact that host-controlled variables have on listing price? The hypothesis is that "Superhost" status will not increase the impact that a host-controlled variable will have on the listing price. "Superhost" status has already been linked to higher listing prices (Liang et al., 2017). However, this status is outside of the host's ability to obtain. Because of this reason, it is predicted that there won't be a strong correlation between variables within the host's control and "Superhost" status.

5. Methodology

5.1 Regression Models

Multiple Linear Regression Equation:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n$$

To determine the significance of host and out of host-controlled variables, two separate linear regression models were created. To provide more accuracy, a subset of listing price was created with prices < $500 a night. This provided more accuracy by causing the data to fit a normal distribution due to the original data set containing outliers up to $10,000. Based on further analysis of the distribution of listing prices in the dataset, it was determined that log transforming the listing price would also cause the data to fit a more normal distribution. The independent variables were changed between the two models. The Host Variable model was composed of only host-controlled variables: beds, bedrooms, bathrooms, room type, accommodations, and "Superhost" status. The Mixed Variable Model contained all the independent variables included in the Host Variable Model as well as out of host-controlled variables: city, neighborhood, and number of reviews.

6. Results

The paper investigated the relationships between the dependent variable (Airbnb listing price) and the independent variables (beds, bedrooms, bathrooms, room type, accommodates, neighborhood, city, number of reviews, and Superhost status) using two separate linear regression models. Below are the table of coefficients for each linear regression model.

	loglist		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.68	4.66 – 4.70	**<0.001**
bedrooms	0.07	0.06 – 0.08	**<0.001**
bathrooms	0.05	0.04 – 0.06	**<0.001**
room_type [Hotel room]	0.30	0.22 – 0.38	**<0.001**
room_type [Private room]	-0.66	-0.68 – -0.65	**<0.001**
room_type [Shared room]	-1.40	-1.45 – -1.35	**<0.001**
accommodates	0.06	0.05 – 0.06	**<0.001**
host_is_superhost [t]	-0.11	-0.12 – -0.10	**<0.001**
Observations	19305		
R^2 / R^2 adjusted	0.484 / 0.484		

Figure 9. *Host Variable Model Output*

	loglist		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.81	4.79 – 4.83	**<0.001**
bedrooms	0.06	0.05 – 0.07	**<0.001**
bathrooms	0.06	0.05 – 0.07	**<0.001**
room_type [Hotel room]	0.33	0.25 – 0.41	**<0.001**
room_type [Private room]	-0.70	-0.71 – -0.68	**<0.001**
room_type [Shared room]	-1.39	-1.44 – -1.35	**<0.001**
accommodates	0.06	0.06 – 0.07	**<0.001**
host_is_superhost [t]	-0.04	-0.05 – -0.02	**<0.001**
city [Seattle]	-0.23	-0.24 – -0.21	**<0.001**
number_of_reviews	-0.00	-0.00 – -0.00	**<0.001**
Observations	19305		
R^2 / R^2 adjusted	0.518 / 0.518		

Figure 10. *Mixed Variable Model Output*

The host variable model (Figure 9) had an adjusted R-squared of 0.484 and a MAPE of 7.863 (Figure 11). The mixed variable model (Figure 10) had an adjusted R-squared of 0.6129 and a MAPE of 6.593 (Figure 11).

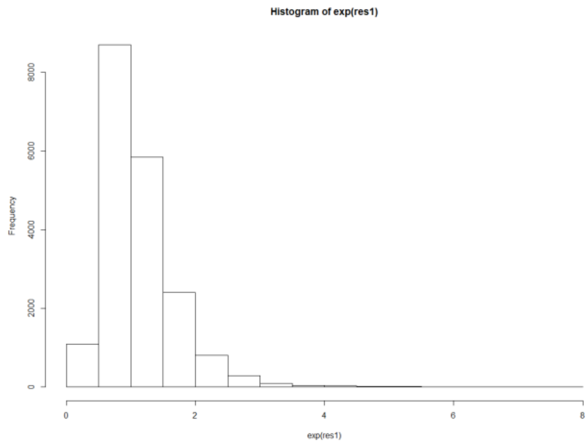
Accuracy Results:						
Host Variable Model:						
	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**
Training Set	-1.97E-16	0.4586405	0.3635829	-1.004257	7.86343	0.7057984
Mixed Variable Model:						
	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**
Training Set	-3.05E-17	0.3960998	0.3041784	-0.765731	6.593692	0.5904806

Figure 11. *Accuracy Results Table*

Looking at Figure 12, the P-value is 2.2 e-16. A low p-value indicates that the differences between some of the means are statistically significant. The F-statistic is greater than the F-critical value, which means the difference in results is likely caused by chance at the alpha level 0.05.

Anova Test						
	Res.Df	RSS	Df	Sum of Sq	F	Pr (>F)
1	19297	4060.8				
2	19183	3028.9	114	1032	57.332	< 2.2e-16 ***

Figure 12. *Anova Test Results Table*



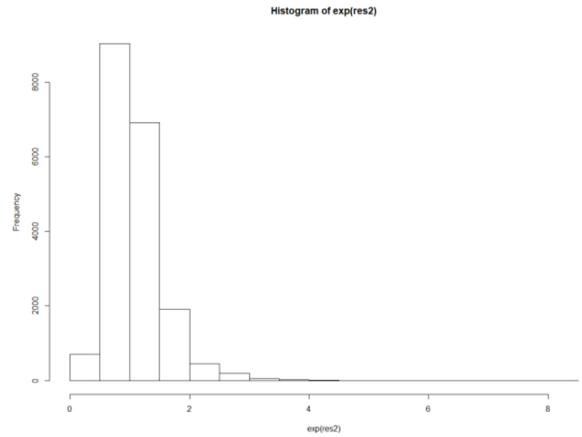


Figure 13. *Histogram of Residuals: Host Variable Model*

Figure 14. *Histogram of Residuals: Mixed Variable Model*

Histogram of residuals across each model appear to be almost identical (see Figures _ & _). This indicates that the distribution of pricing error was similar. One difference between the two histograms is that the Mixed Variable Model's distribution seems to fit a normal distribution slightly more accurately.

7. Discussion/Conclusion

This study uncovered three key findings in association to Airbnb listing prices. The first main result is that both host-controlled and out of host-controlled variables prove to have significance in determining Airbnb listing prices. Across both models, all independent variables proved to have significance except for certain neighborhoods. I believe this is due to the demand and activity that some neighborhoods have over others in each specific city. When run without neighborhoods, the Mixed Variable Model's adjusted R-squared dropped to .518 compared to .613. This indicates that even though some neighborhoods were insignificant, the significant neighborhoods had a strong correlation to listing price and contributed to a higher adjusted R-square. The second main finding was that when host-controlled and out of host-controlled variables are combined, they prove to have a greater significance on listing price accuracy. In the results, the mixed variable model performed better than the host-controlled model. While hosts have no control over certain variables, they should do their research on the market that they are in and be aware of the impact that certain variables outside of their control have on listing price. The final key finding from the results shows that host-controlled variables prove to have greater significance to listing price than out of host-controlled variables. Looking at the results, while the mixed variable model performed greater than the host-controlled model, the host-controlled variables made a more significant impact in creating the R-squared and MAPE of the model, proving to have a high correlation with the listing price variable.

While this study shows significance in the difference between these two types of variables in relation to listing price, there are some limitations and improvements that could be made. The dataset used in this study was limited to two U.S. cities with around 9,000 listings each. One way to further this study is to gather a larger collection of cities in the U.S. and take listing data from years prior to 2019 to more accurately impact the growth in Airbnb's popularity. Another limitation to this study was the collection of variables that were used for analysis. While I believe these variables best represent what this study was aiming to accomplish, adding additional variables to the study could have revealed better results.

As many are aware, Covid-19 has had a tremendous impact on the global economy. As of May 1, 2020, there are over 3 million confirmed global cases according to World Health Organization (WHO), and almost every country and every business has been affected by the virus. Airbnb, a company whose business model revolves around travel, has been impacted heavily. While Covid-19 is still relatively new and quickly developing around the world, it has already made a substantial impact on Airbnb listing prices. Pulling Airbnb listing data from Inside Airbnb on Rome, Italy, one of the hardest hit cities by the virus, the average price per listing appears to have come down significantly. Comparing listing data from March 2019 vs. March 2020, the average listing price in Rome had gone down from $110.50 to 99.32. This would mean that prices for listings in March 2020 in Rome were 11% lower than they were the same time the year before most certainly due to Covid-19.

While data is still limited on the Coronavirus due to it spreading globally recently, the impact that the virus will have on the sharing economy is severe. Even as the pandemic dies down, companies like Airbnb will most likely feel the impact that the virus had for many years to come.

In this paper, I focused on the effects that host-controlled and out of host-controlled variables had on Airbnb listing prices. While the results are promising, further analysis will always be needed due to the dynamic pricing nature of the Airbnb market. This study aimed to provide Airbnb hosts and Airbnb consumers with information on how certain variables impact the price of an Airbnb listing.

8. Appendix

A. (Mixed Variable Model Output w/neighborhoods included):

Table of regression coefficients:

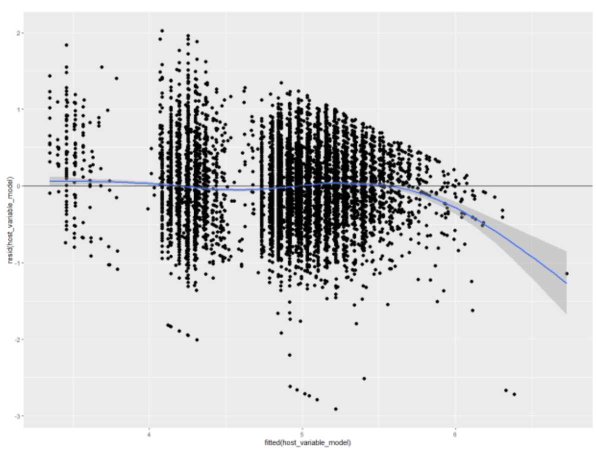
	loglist		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	4.61	4.49 – 4.72	**<0.001**
bedrooms	0.12	0.11 – 0.13	**<0.001**
bathrooms	0.05	0.04 – 0.06	**<0.001**
room_type [Hotel room]	0.16	0.08 – 0.23	**<0.001**
room_type [Private room]	-0.50	-0.51 – -0.48	**<0.001**
room_type [Shared room]	-1.22	-1.27 – -1.18	**<0.001**
accommodates	0.06	0.06 – 0.07	**<0.001**
host_is_superhost [t]	0.01	-0.00 – 0.02	0.175
neighbourhood_cleansed [Alki]	0.02	-0.07 – 0.11	0.669
neighbourhood_cleansed [Allston]	-0.12	-0.24 – 0.00	0.056
neighbourhood_cleansed [Arbor Heights]	-0.26	-0.41 – -0.11	**0.001**
neighbourhood_cleansed [Atlantic]	-0.03	-0.11 – 0.06	0.518
neighbourhood_cleansed [Back Bay]	0.36	0.24 – 0.48	**<0.001**
neighbourhood_cleansed [Bay Village]	0.24	0.09 – 0.39	**0.002**
neighbourhood_cleansed [Beacon Hill]	0.19	0.07 – 0.31	**0.003**
neighbourhood_cleansed [Belltown]	0.30	0.23 – 0.36	**<0.001**
neighbourhood_cleansed [Bitter Lake]	-0.24	-0.37 – -0.12	**<0.001**
neighbourhood_cleansed [Briarcliff]	0.04	-0.12 – 0.20	0.598
neighbourhood_cleansed [Brighton]	-0.24	-0.35 – -0.13	**<0.001**
neighbourhood_cleansed [Broadview]	-0.22	-0.35 – -0.09	**0.001**
neighbourhood_cleansed [Broadway]	0.18	0.11 – 0.24	**<0.001**
neighbourhood_cleansed [Bryant]	-0.10	-0.21 – 0.01	0.083
neighbourhood_cleansed [Cedar Park]	-0.13	-0.27 – 0.01	0.067
neighbourhood_cleansed [Central Business District]	0.53	0.45 – 0.60	**<0.001**
neighbourhood_cleansed [Charlestown]	0.15	0.02 – 0.28	**0.020**
neighbourhood_cleansed [Chinatown]	0.33	0.20 – 0.46	**<0.001**
neighbourhood_cleansed [Columbia City]	-0.14	-0.22 – -0.05	**0.001**
neighbourhood_cleansed [Crown Hill]	-0.10	-0.23 – 0.03	0.119
neighbourhood_cleansed [Dorchester]	-0.27	-0.38 – -0.15	**<0.001**
neighbourhood_cleansed [Downtown]	0.36	0.24 – 0.48	**<0.001**
neighbourhood_cleansed [Dunlap]	-0.29	-0.41 – -0.16	**<0.001**
neighbourhood_cleansed [East Boston]	-0.09	-0.21 – 0.03	0.160
neighbourhood_cleansed [East Queen Anne]	0.11	0.03 – 0.20	**0.008**
neighbourhood_cleansed [Eastlake]	0.05	-0.04 – 0.15	0.259
neighbourhood_cleansed [Fairmount Park]	-0.08	-0.20 – 0.04	0.212
neighbourhood_cleansed [Fauntleroy]	0.04	-0.10 – 0.17	0.597
neighbourhood_cleansed [Fenway]	0.25	0.13 – 0.37	**<0.001**
neighbourhood_cleansed [First Hill]	0.32	0.25 – 0.40	**<0.001**
neighbourhood_cleansed [Fremont]	0.04	-0.03 – 0.11	0.282
neighbourhood_cleansed [Gatewood]	-0.12	-0.23 – -0.01	**0.031**
neighbourhood_cleansed [Genesee]	-0.00	-0.11 – 0.10	0.951
neighbourhood_cleansed [Georgetown]	-0.12	-0.27 – 0.03	0.108
neighbourhood_cleansed [Green Lake]	-0.01	-0.10 – 0.07	0.772
neighbourhood_cleansed [Greenwood]	-0.21	-0.28 – -0.13	**<0.001**
neighbourhood_cleansed [Haller Lake]	-0.19	-0.30 – -0.08	**0.001**
neighbourhood_cleansed [Harbor Island]	-1.01	-1.79 – -0.23	**0.011**
neighbourhood_cleansed [Harrison/Denny-Blaine]	0.33	0.20 – 0.47	**<0.001**
neighbourhood_cleansed [High Point]	-0.19	-0.30 – -0.07	**0.002**
neighbourhood_cleansed [Highland Park]	-0.27	-0.38 – -0.15	**<0.001**
neighbourhood_cleansed [Holly Park]	-0.07	-0.29 – 0.15	0.537
neighbourhood_cleansed [Hyde Park]	-0.37	-0.51 – -0.22	**<0.001**
neighbourhood_cleansed [Industrial District]	0.36	-0.09 – 0.82	0.118
neighbourhood_cleansed [Interbay]	-0.17	-0.30 – -0.04	**0.011**
neighbourhood_cleansed [International District]	0.22	0.12 – 0.32	**<0.001**
neighbourhood_cleansed [Jamaica Plain]	-0.06	-0.18 – 0.06	0.365
neighbourhood_cleansed [Laurelhurst]	0.06	-0.10 – 0.23	0.454
neighbourhood_cleansed [Lawton Park]	0.05	-0.05 – 0.14	0.315
neighbourhood_cleansed [Leather District]	0.35	0.03 – 0.67	**0.030**
neighbourhood_cleansed [Leschi]	0.01	-0.08 – 0.10	0.838
neighbourhood_cleansed [Longwood Medical Area]	0.09	-0.14 – 0.32	0.443
neighbourhood_cleansed [Lower Queen Anne]	0.13	0.05 – 0.20	**0.001**
neighbourhood_cleansed [Loyal Heights]	-0.05	-0.15 – 0.04	0.282
neighbourhood_cleansed [Madison Park]	0.04	-0.11 – 0.19	0.580
neighbourhood_cleansed [Madrona]	0.14	0.03 – 0.25	**0.012**
neighbourhood_cleansed [Mann]	-0.11	-0.19 – -0.03	**0.010**
neighbourhood_cleansed [Maple Leaf]	-0.11	-0.21 – -0.00	**0.045**
neighbourhood_cleansed [Mattapan]	-0.37	-0.51 – -0.24	**<0.001**
neighbourhood_cleansed [Matthews Beach]	-0.11	-0.24 – 0.02	0.085
neighbourhood_cleansed [Meadowbrook]	-0.20	-0.36 – -0.03	**0.021**
neighbourhood_cleansed [Mid-Beacon Hill]	-0.19	-0.29 – -0.10	**<0.001**
neighbourhood_cleansed [Minor]	0.03	-0.04 – 0.11	0.359
neighbourhood_cleansed [Mission Hill]	0.01	-0.12 – 0.13	0.886
neighbourhood_cleansed [Montlake]	-0.16	-0.25 – -0.07	**<0.001**
neighbourhood_cleansed [Mount Baker]	-0.13	-0.23 – -0.04	**0.004**
neighbourhood_cleansed [North Admiral]	0.03	-0.06 – 0.11	0.555
neighbourhood_cleansed [North Beach/Blue Ridge]	0.03	-0.11 – 0.17	0.686
neighbourhood_cleansed [North Beacon Hill]	-0.05	-0.13 – 0.03	0.223
neighbourhood_cleansed [North College Park]	-0.29	-0.40 – -0.19	**<0.001**
neighbourhood_cleansed [North Delridge]	-0.13	-0.23 – -0.02	**0.015**
neighbourhood_cleansed [North End]	0.09	-0.03 – 0.22	0.149
neighbourhood_cleansed [North Queen Anne]	0.06	-0.02 – 0.14	0.157
neighbourhood_cleansed [Olympic Hills]	-0.22	-0.35 – -0.08	**0.002**
neighbourhood_cleansed [Phinney Ridge]	-0.08	-0.16 – 0.00	0.063
neighbourhood_cleansed [Pike-Market]	0.43	0.35 – 0.51	**<0.001**
neighbourhood_cleansed [Pinehurst]	-0.24	-0.37 – -0.11	**<0.001**
neighbourhood_cleansed [Pioneer Square]	0.41	0.31 – 0.51	**<0.001**
neighbourhood_cleansed [Portage Bay]	0.14	-0.01 – 0.29	0.066
neighbourhood_cleansed [Rainier Beach]	-0.25	-0.35 – -0.14	**<0.001**
neighbourhood_cleansed [Rainier View]	-0.36	-0.52 – -0.20	**<0.001**
neighbourhood_cleansed [Ravenna]	-0.17	-0.26 – -0.08	**<0.001**
neighbourhood_cleansed [Riverview]	-0.34	-0.45 – -0.23	**<0.001**
neighbourhood_cleansed [Roosevelt]	-0.17	-0.26 – -0.07	**<0.001**
neighbourhood_cleansed [Roslindale]	-0.37	-0.50 – -0.24	**<0.001**
neighbourhood_cleansed [Roxbury]	-0.24	-0.36 – -0.12	**<0.001**
neighbourhood_cleansed [Roxhill]	-0.16	-0.33 – 0.01	0.058
neighbourhood_cleansed [Seaview]	-0.14	-0.27 – -0.01	**0.031**
neighbourhood_cleansed [Seward Park]	-0.10	-0.19 – 0.00	0.058
neighbourhood_cleansed [South Beacon Hill]	-0.43	-0.56 – -0.30	**<0.001**
neighbourhood_cleansed [South Boston]	0.19	0.07 – 0.31	**0.002**
neighbourhood_cleansed [South Boston Waterfront]	0.55	0.42 – 0.68	**<0.001**
neighbourhood_cleansed [South Delridge]	-0.39	-0.51 – -0.27	**<0.001**
neighbourhood_cleansed [South End]	0.25	0.13 – 0.37	**<0.001**
neighbourhood_cleansed [South Lake Union]	0.26	0.18 – 0.34	**<0.001**
neighbourhood_cleansed [South Park]	-0.23	-0.41 – -0.05	**0.012**

neighbourhood_cleansed [Southeast Magnolia]	0.07	-0.04 – 0.17	0.238
neighbourhood_cleansed [Stevens]	0.04	-0.03 – 0.12	0.264
neighbourhood_cleansed [Sunset Hill]	0.13	0.02 – 0.24	**0.021**
neighbourhood_cleansed [University District]	-0.09	-0.16 – -0.01	**0.019**
neighbourhood_cleansed [Victory Heights]	-0.23	-0.36 – -0.10	**<0.001**
neighbourhood_cleansed [View Ridge]	-0.07	-0.23 – 0.09	0.416
neighbourhood_cleansed [Wallingford]	-0.03	-0.10 – 0.04	0.368
neighbourhood_cleansed [Wedgwood]	-0.22	-0.34 – -0.11	**<0.001**
neighbourhood_cleansed [West End]	0.33	0.20 – 0.46	**<0.001**
neighbourhood_cleansed [West Queen Anne]	0.17	0.08 – 0.25	**<0.001**
neighbourhood_cleansed [West Roxbury]	-0.30	-0.44 – -0.16	**<0.001**
neighbourhood_cleansed [West Woodland]	-0.04	-0.12 – 0.05	0.422
neighbourhood_cleansed [Westlake]	0.01	-0.08 – 0.10	0.801
neighbourhood_cleansed [Whittier Heights]	0.01	-0.09 – 0.12	0.803
neighbourhood_cleansed [Windermere]	-0.14	-0.29 – 0.01	0.066
neighbourhood_cleansed [Yesler Terrace]	0.06	-0.05 – 0.18	0.273
city [Seattle]	-0.20	-0.30 – -0.10	**<0.001**
number_of_reviews	-0.00	-0.00 – -0.00	**<0.001**

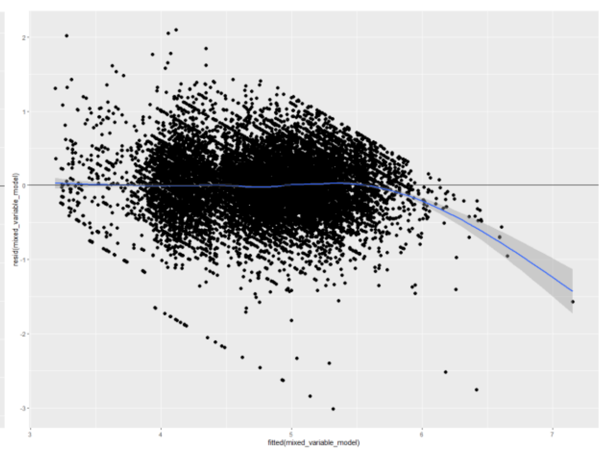
Observations	19305
R^2 / R^2 adjusted	0.615 / 0.613

B. (Fitted values vs. the residuals):

Host Variable Model:



Mixed Variable Model:



References

Chattopadhyay, M., & Mitra, S. K. (2019). Do airbnb host listing attributes influence room pricing homogenously? *International Journal of Hospitality Management*, *81*, 54–64. https://doi.org/10.1016/j.ijhm.2019.03.008

Chen, Y. and Xie, K. (2017), “Consumer valuation of Airbnb listings: a hedonic pricing approach”, International Journal of Contemporary Hospitality Management, Vol. 29 No. 9, pp. 2405-2424.

D. Hill, "How much is your spare room worth?," in *IEEE Spectrum*, vol. 52, no. 9, pp. 32-58, Sept. 2015.

Dogru, T., & Pekin, O. (2017). What do guests value most in Airbnb accommodations? An application of the hedonic pricing approach.

Ert, E., Fleischer, A. and Magen, N. (2016), “Trust and reputation in the sharing economy: the role of personal photos in Airbnb”, Tourism Management, Vol. 55, pp. 62-73

Gibbs, C., Guttentag, D., Gretzel, U., Yao, L., & Morton, J. (2018). Use of dynamic pricing strategies by Airbnb hosts. *International Journal of Contemporary Hospitality Management*, *30*(1), 2–20. https://doi.org/10.1108/IJCHM-09-2016-0540

Gibbs, C., Guttentag, D., Gretzel, U., Morton, J. and Goodwill, A. (2018), “Pricing in the sharing economy: a hedonic pricing model applied to Airbnb listings”, Journal of Travel & Tourism Marketing, pp. 1-11.

Li, J., Moreno, A. and Zhang, D.J. (2015), “Agent behavior in the sharing economy: evidence from Airbnb”, Working paper, [1298] Ross School of Business, University of Michigan, Ann Arbor, MI

Liang, S., Schuckert, M., Law, R., & Chen, C.-C. (2017). Be a “Superhost”: The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, *60*, 454–465. https://doi.org/10.1016/j.tourman.2017.01.007

McGuire, K. A. (2015). Hotel pricing in a social world: driving value in the digital economy. John Wiley & Sons.

Tang, E., & Sangani, K. (2015). *Neighborhood and Price Prediction for San Francisco Airbnb Listings*. 6.

TeuBner, T., Hawlitschek, F., & Dann, D. (2017). Price Determinants On Airbnb: How Reputation Pays Off In *The Sharing Economy. (2017). Journal of Self-Governance and Management Economics, 5(4), 53. https://doi.org/10.22381/JSME5420173*

Wang, D., & Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, *62*, 120–131. https://doi.org/10.1016/j.ijhm.2016.12.007

Wu, E. A. A. (2016). Learning in peer-to-peer markets: evidence from Airbnb (Doctoral dissertation).

Ye, P., Qian, J., Chen, J., Wu, C., Zhou, Y., De Mars, S., … Zhang, L. (2018). Customized Regression Model for Airbnb Dynamic Pricing. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD ’18*, 932–940. https://doi.org/10.1145/3219819.3219830

Zhang, Z., Chen, R., Han, L., & Yang, L. (2017). Key Factors Affecting the Price of Airbnb Listings: A Geographically Weighted Approach. *Sustainability*, *9*(9), 1635. https://doi.org/10.3390/su9091635