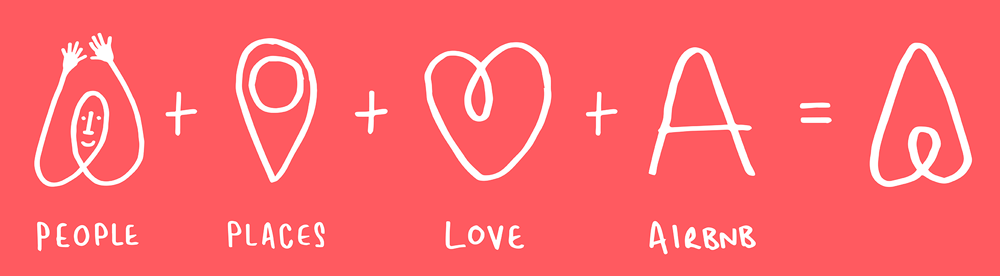
Multiple Linear Regression Model for Airbnb Listings Price



Contents

[1. Introduction 3](#_Toc129897937)

[2. Data Preparation 4](#_Toc129897938)

[2.1 Data Overview 4](#_Toc129897939)

[2.2 Data Cleaning & Transformation 4](#_Toc129897940)

[2.3 Descriptive Statistics of cleaned data 7](#_Toc129897941)

[3. Modelling 8](#_Toc129897942)

[3.1 Correlation Analysis 8](#_Toc129897943)

[3.2 Model Development 9](#_Toc129897944)

[3.3 Model Interpretation 10](#_Toc129897945)

[3.4 Model Evaluation 10](#_Toc129897946)

[4. Insights 12](#_Toc129897947)

[5. Limitation 13](#_Toc129897948)

[6. Further research 15](#_Toc129897949)

[7. Conclusion 15](#_Toc129897950)

[Reference 17](#_Toc129897951)

[Appendix A – Data Summary 19](#_Toc129897952)

[Appendix B – R Output 21](#_Toc129897953)

[B.1 Model Coefficients 21](#_Toc129897954)

[B.2 Model Assessment 23](#_Toc129897955)

[Appendix C – Geospatial 24](#_Toc129897956)

# 1. Introduction

Airbnb is an online marketplace for booking services and travel information. Airbnb has been a forerunner in terms of changing rental markets via the introduction of peer-to-peer accommodation services (Perez-Sanchez et al., 2018).

Airbnb’s product offering concerns the company’s two key areas: renters and hosts. Airbnb’s product offerings to renters range from affordable homes and shared living spaces to unique stays like castles. Airbnb’s product offering to hosts is the platform to earn income via the website and app Airbnb maintains. Airbnb is a global service in over 190 countries, with over 40,000 listings in New Zealand alone. This allows renters to find housing globally and in remote areas where traditional accommodations like hotels are unavailable. Airbnb’s global promotional reach encompasses both traditional marketing, like television and billboard advertisements, but also digital media like social media advertisements. Airbnb has been successful largely due to the rising of peer-to-peer services but also because Airbnb is able to price lower than traditional stays like hotels. Though Airbnb has become one of the stars of the sharing economy, Airbnb’s revenue stream may be impacted due to the inefficiency of Airbnb pricing and the hosts of Airbnb ineffectively price their accommodations on the platform (Gibbs et al., 2018).

This project report aims to use machine learning techniques to identify trends within the New Zealand Airbnb dataset from Inside Airbnb. Data collected from Airbnb shows that listings from Airbnb are majorly homes, which are disrupting housing and communities despite Airbnb’s claims that they are a part of the sharing economy (Zervas et al., 2017). In this report, an analysis of the listings’ detailed information on Airbnb within New Zealand will be explored to indicate what influences Airbnb property prices potentially. Thus, this report set the following hypothesis (null) that listing details such as the number of bedrooms, location, and amenities have no influence on the listing price.

# 2. Data Preparation

## 2.1 Data Overview

In order to find out what influences the daily price of Airbnb’s in New Zealand, the detailed listings dataset from January 2023 was retrieved from InsideAirbnb.com.

The dataset originally contained 42,323 rows where each row has one listing and all its information. There are 80 columns in the dataset, with a breakdown of 39 numerical fields, 35 categorical fields, and 6 logical fields.

The dependent price is the daily price for a listing with a mean of 392 (Table 1).

|  |  |
| --- | --- |
|  | Price $NZD – Daily price for a listing |
| Minimum Value | 14 |
| Median | 200 |
| Mean | 392 |
| Maximum Value | 145 280 |

Table 1 Summary Statistics for Price

## 2.2 Data Cleaning & Transformation

The accuracy and reliability of our predictive model for Airbnb listing prices heavily depends on the data quality. A rigorous data cleaning and transformation process was implemented.

Chart

Description automatically generated with medium confidence

Figure 1 Listing Price Distribution

The distribution of the listing prices was analysed using a boxplot and histogram (see Figure 1). It was observed that the distribution pattern was heavily influenced by outliers that may not be representative of the underlying population. Therefore, the outliers in the price variable were removed by trimming the data to the 1%-99% range, resulting in a more accurate representation of the underlying distribution of prices.

Chart, histogram

Description automatically generated

Figure 2 1% - 99% Listing Price Distribution

It was observed that the distribution of trimmed prices still possessed a right-skewed distribution (see Figure 2). To normalise the data and meet the assumption of normality for linear regression models, the log transformation on the price will be used when fitting the model. Figure 3 shows that the distribution of log-transformed price is normally distributed.

Chart, histogram

Description automatically generated

Figure 3 Log Transformed Listing Price Distribution

Irrelevant or redundant variables were removed from the dataset, reducing the noise and complexity in the dataset and improving the accuracy and performance of predictive models (See Table 2).

|  |  |
| --- | --- |
| Variable Type | Removed Variable Names |
| Empty variables | bathrooms, calendar\_updated, requires\_license, license, region\_parent\_parent\_id, region\_parent\_parent\_name |
| URLs | 1listing\_url, picture\_url, host\_url, host\_thumbnail\_url, host\_picture\_url |
| IDs | scrape\_id, region\_id, region\_parent\_id, host\_id (Listing ID is kept for identification purposes) |
| Irrelevant variables | last\_scraped, last\_searched, host\_name, host\_neighbourhood, host\_location, neighbourhood, has\_availability, calendar\_last\_scraped |
| Variables highly correlated with others | minimum\_maximum\_nights, minimum\_minimum\_nights, minimum\_nights\_avg\_ntm, maximum\_minimum\_nights,-maximum\_maximum\_nights -maximum\_nights\_avg\_ntm |

Table 2 Summary of Dropped Variables

Certain variables were converted from their original data type. For example, price that was initially stored as strings had to be converted to numeric values for analysis and model fitting.

To extract more useful information for analysis, certain variables have been transformed into another variable. For instance, in our dataset, the values of amenities were represented as text descriptions of the amenities available in each listing. It would have been difficult to incorporate these textual descriptions into our model directly. Therefore, the variable representing the amenities was transformed to the number of amenities to assess whether the number of amenities impacted the listing price. This transformation allowed for more straightforward analysis and incorporation into our predictive model.

To improve the interpretability of our model, some variables such as property type, region name, and parent region name had many categories that could have led to overfitting or made it difficult to incorporate into the model directly. Therefore, based on intuitive justifications, we classified these variables into more insightful and interpretable categories.

For instance, 117 property types were classified into three categories: Standard accommodation, Nature accommodation, and Unique stays. This allowed us to capture additional information about these variables that may be relevant to the prediction of listing prices. Moreover, external sources, Urban Rural Profile Categories from Otago University (University of Otago, n.d.), were used to classify 239 regions into urban and rural areas. Additionally, 68 parent regions were classified as tourism and non-tourism destinations by incorporating data on the number of international visitors New Zealand 2019 by region (New Zealand: International Visitors by Region | Statista, 2023). These classifications were more intuitive and easier to interpret, which helped improve our model’s interpretability.

The exclusionary approach was employed to address missing values in the dataset, where any rows with missing values were removed. The processed dataset of 29,457 observations after the cleaning and transformation process was deemed sufficiently large for the analysis.

When analyzing the descriptive statistics of cleaned data, it is not uncommon to observe extreme values in the number of beds, number of bathrooms, and number of bedrooms, which can skew the data and make it less representative of typical New Zealand property listings. To address this issue, we trimmed observations with more than 10 bathrooms,10 bedrooms and 25 beds. This can help to ensure that the data is more representative of typical New Zealand property listings.

## 2.3 Descriptive Statistics of cleaned data

The cleaned dataset of the January 2023 Airbnb NZ details listings contains 29,418 observations and 54 columns of data. The breakdown of 45 numerical variables, 6 categorical variables, and logical variables. For a full breakdown of summary statistics please refer to Appendix A.

|  |  |
| --- | --- |
|  | Price $NZD – Daily price for a listing |
| Minimum Value | 274 |
| Median | 200 |
| Mean | 215 |
| Maximum Value | 1627 |

Table 3 Summary Statistics for Price after cleaning

# 3. Modelling

The processed dataset has been split into train data and test data, 80% of the data are used for training and the remaining 20% for testing.

## 3.1 Correlation Analysis

A correlation plot was created to investigate the correlation between all the numeric variables in the dataset to identify the most important predictor variables for the linear regression model. Variables with a correlation coefficient greater than 0.4 were considered to have at least moderate correlation with price, as shown in Figure 4, four variables were selected based on this criterion.

Chart, histogram

Description automatically generated

Figure 4 Correlation Heatmap

However, accommodates, bedrooms, and number of beds were found to be highly correlated, which may imply multicollinearity, making it difficult to determine the true effect of each variable on the dependent variable, as the coefficients may be biased or unstable. To address this issue, a principal component analysis (PCA) was conducted to identify the most important factors, and as shown in Figure 5, the variable of beds and bathroom number has the highest importance and ultimately selected to be retained.

Chart

Description automatically generated

Figure 5 PCA Variable Importance Indicator

## 3.2 Model Development

There are five phases in the model development. Firstly, according to the above correlation analysis and PCA analysis, the number of beds and bathrooms are predictors in model 1. Secondly, all categorical variables were included to fit model 2. The host\_reponse\_time was not statistically significant. Thirdly, as a result, host\_response\_time was removed in the subsequent model, Model 3.. Fourthly, amenities, review\_score\_ranking and number\_of\_reviews were added to build model 4, which were considered to be factors that influence the price. Finally, host\_is\_superhost was dropped in model 5 as it was not significant in model 4. Comparing models 1,2,3,4 and 5 using Adjusted R squared and AICs metrics , model 5 was attributed with the highest adjusted R squared and lowest AIC. In turn model 5 was selected as the final model (model comparison is shown in Appendix B.1). The equation of the multiple linear regression model is expressed as follows:

## 3.3 Model Interpretation

Provided with the linear regression equation and model. The dataset was analysed to determine the validity of the null hypothesis. Essentially, the null hypothesis stated that no relationship existed between the independent variables and listing price. On the contrary, the alternative hypothesis states that a relationship does in fact exist between the independent variables and listing price. Furthermore, as shown in Figure B.1 1 in Appendix B.1, all the independent variables were statistically significant. This provided evidence to reject the null hypothesis and accept the alternative hypothesis.

The model coefficients shows fourteen variables along with the level of impact each independent variable has on the listing price. Out of the fourteen variables, nine had a significant positive impact on price, which is shown in Table B.1 1. It is insightful to note that within Table B.1 1, the number of bathrooms, private baths and unique stay variables had the highest impact on price. Furthermore, as shown in Table B.1 2, five of the fourteen variables had a significant negative impact on price, with shared room and private room variables having the most impact.

## 3.4 Model Evaluation

Adjusted R-squared, and root mean square error (RMSE) was used to evaluate the performance of the regression model.

The adjusted R-squared measures the proportion of variance in the dependent variable that can be explained by the independent variables included in the model; the model has an adjusted R-squared of 0.54 on training data, meaning that 54% of the variability in the price can be explained by the model. However, it also suggests that there may be other factors outside of the model that is influencing the listings’ price.

The RMSE measures the average difference between the predicted and actual values in the dataset. The model arrived with an RMSE of 0.46 on training data, $165 after transforming back to the original scale. For the listing with a median price of $500, This means that, on average, the model's predictions are off by $165 for listings with a median price of $500.

The adjusted R-squared and RMSE on test data are 0.53 and 0.47($157,) respectively, the difference is not significant, which suggests that overfitting the model in train data was avoided.

Several diagnostic tests, such as observing the regression plots, were further conducted to assess the quality of the regression model. This included an insight into the linearity, homoscedasticity, influential observations, and normality, as illustrated in Figure B.2 3 in Appendix B.2, which indicated that the model, which was fitted to the logged price variable, satisfied all of the underlying assumptions.

4. Insights

As shown in Table B.1, the factors that had a significant impact on listing prices could be divided into two categories. A category that was related to the listing’s physical features, and this included factors such as number of bathrooms, beds and amenities, as well as the type of bathroom (Private or Shared), property (Standard or Unique) and room (Private or Shared). The second category focused on the listing’s location, and this revolved around whether the listing was in a tourist region or not and whether the listing was located in an urban or rural area.

From Table B.1, Tourist areas had a positive and significant impact on the listing price. Similarly, listings located in an urban area had a positive and significant impact on the listing price. From those findings, it is inferred that listings located in urban, and tourist dense regions tend to have a higher listing price than their counterparts. Furthermore, the geospatial information in Appendix C displays the location of all the New Zealand listings of January 2023 associated with their price as a colour heatmap. Additional to the fact that Urban based listings are higher priced than other listings, it is observed that as listings move further away from urban areas, their listing price drops. However, the price starts to increase as listings get closer to coastal areas. This pattern is observed across multiple major cities in New Zealand, such as Auckland, Tauranga, and Wellington. Additional to the regression model, emphasising the positive relationship between urban and tourist locations with price, the geospatial graphs provide price trends and further emphasis on the positive relationship between coastal and urban areas with listing price.

The geographic findings provide useful insights to Airbnb hosts and could be utilized to drive increased customer traffic and optimally price their listing amongst their competitors. Based on the geographic segmentation findings, Tourist areas significantly increased in listing price. In turn, hosts in these areas could focus their marketing channels on popular tourist sites and pages and local travelling pages. For example, advertising a listing on popular tourist social media channels or New Zealand Domestic traveller pages. According to the ministry of Business, tourism in New Zealand tends to peak during summer and it was reported that spending during summer was two to three times higher than during other seasons (2016). This provides hosts with coastal based listings an opportunity to rebrand their listing names to include keywords that match accordingly with the summer season. For example, emphasising on the keywords such as ‘Beach View’, ‘Beach House’, and ‘Summer House’.

Furthermore, given the physical features that are of significant impact on the listing price, Airbnb hosts with listings in the same location as their competitors can compare with other host’s listings based on those physical features mentioned in Table B.1 1 to optimally price their listing in comparison to their competitors.

Furthermore, Airbnb could use the fact that review scores positively impact price to encourage hosts to improve their ratings. This could be performed through blog posts on the Airbnb site and Articles that are focused on methods to provide a positive experience to their guests.

5. Limitation   
Although the final model conducted is robust enough for prediction, it is important to acknowledge the limitations of the analysis. These limitations stem from various factors, including a lack of property information in the dataset, the exclusion of certain variables, the grouping of categorical variables, seasonal trends in prices, and multicollinearities. In the following sections, these limitations and their potential impact on the results will be discussed.  
  
Firstly, there are two limitations on data, one is that the given dataset lacks some essential information about properties. For example, the property condition, the style, or the year they were built is unavailable. These factors can significantly impact the renting price because they influence the property value and appeal in the market. As noted by DeLisle that newer properties with modern designs and amenities may be more attractive to tenants and command higher rental rates than older properties with outdated features (2010).  
  
Another limitation of the data is that some data in the dataset was not utilized to build the model that might have an impact on price. For instance, image quality of the property and information of host. According to Guttentag, the presentation of the property using photographs and descriptions is essential to the success of the Airbnb model (2015). In most cases, high-quality photographs and detailed descriptions of the property and surrounding area are provided, enabling potential guests to assess the suitability of the accommodation for their requirements. Additionally, the role of the host in the peer-to-peer accommodation market is critical. By providing information, hosts create a sense of trust and personal connection with potential guests (Zervas, Proserpio, & Byers, 2017). Therefore, the variables like image quality of the property and information of host could play an essential role on the efficiency of marketing, which potentially influences the renting price. People tend to trust properties that have appealing images, and hosts with good reviews, which might affect their willingness to pay a higher price.

The second limitation is that the final model is only based on January 2023 data, which means that the findings did not entail the seasonal trend of Airbnb accommodations. Figure 6 below shows the quarterly consumers price index (CPI) changes of domestic accommodation services from September 2017 to December 2020 in New Zealand (Higher Accommodation Prices Contribute to Inflation | Stats NZ, n.d.). It is clear from the graph that there is a seasonal trend in accommodation prices in New Zealand. One of the main factors that contribute to these seasonal trends is tourism. New Zealand has many beautiful destinations that attract tourists throughout the year. For example, in beach destinations, the summer months are the busiest time of year, while in ski resorts, the winter months are the most popular (Seasons in New Zealand | 100% Pure New Zealand, n.d.). The increased demand in the busy season can drive up the price of accommodations.  However, the seasonal trend is a factor we did not account for in our prediction model.

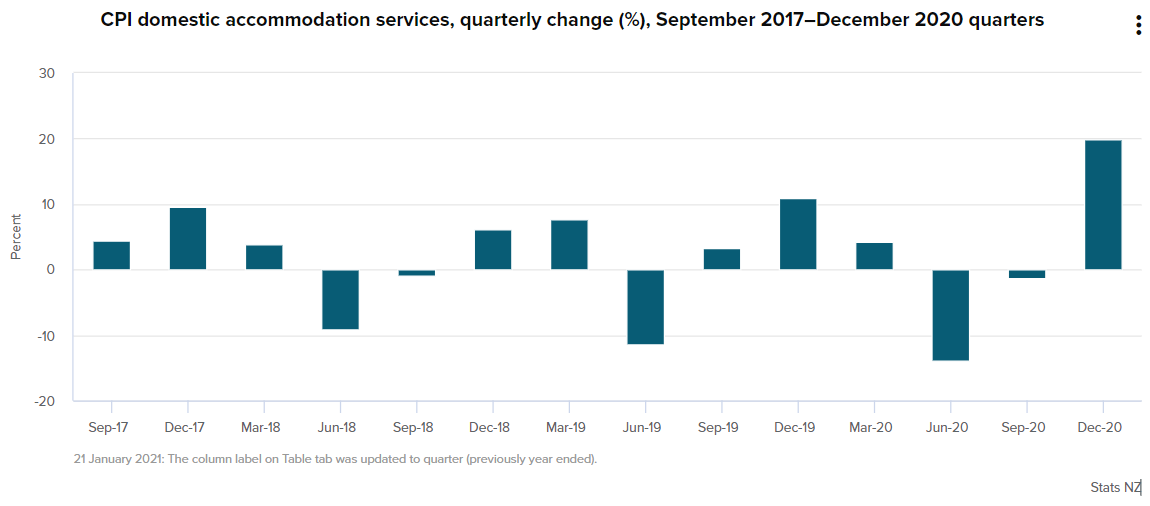


Figure 6 Quarterly change (%) of CPI accommodation services

Furthermore, as discussed in the methods and results section, some categorical variables were grouped into fewer categories. Grouping categorical variables improves interpreting the regression model result and reduces data noise in the dataset. However, it also has the negative influence of leading to loss of information, which could increase bias since grouping variables could be influenced by subjective criteria (Agresti, 2003). For example, the variable "region parent" was grouped into tourist area and non-tourist area according to the most popular tourism cities in New Zealand (New Zealand: International Visitors by Region | Statista, 2023). However, the approach of justifying whether a city is a tourist destination or not can be subjective and may vary. For instance, an alternative way of classifying tourist cities can be based on the number of tourist attractions or activities in each city. Similarly, variable property types were categorised into standard accommodations, unique livings, and nature accommodations. However, for properties like a cave or a boathouse, the classification can be ambiguous as answers can be vary among individuals.

# 6. Further research

From the discussion on limitations, it is clear to note that further research can be developed to gain a more robust prediction model for determining the listing price. Firstly, research can be conducted on more variables that could impact the price, such as occupancy rate, image quality of listings and host information. Also, combining prediction with time series trend analysis could help identify historical patterns that can be useful in predicting future prices. To address the disadvantage of grouping categorical variables, one approach is to use a more granular classification system, this involves breaking down larger categories into smaller subcategories or using more specific criteria for classification (Agresti, 2013a). By incorporating these suggestions, a more robust prediction model can be developed that includes a wider range of variables and is better equipped to make accurate predictions.

# 7. Conclusion

In conclusion, this report has provided an analysis of the factors that affect the price of Airbnb listings in New Zealand. Through data exploration and regression analysis, several variables were identified that significantly impact price, including the property location, features, and review scores. The findings of this report have important implications for both Airbnb hosts and Airbnb itself. Hosts can use the insights gained from this analysis to optimize their pricing strategies and improve the competitiveness of their listings. Airbnb can use this information to develop pricing strategies that increases revenue and improves the competitiveness of the platform.

However, it is important to note that this analysis is limited by the availability and quality of the data, as well as the assumptions and limitations of the regression model used. Further research is needed to confirm and expand upon these findings, including examining additional variables that could influence price and using more advanced approaches to group categorical variables.

Overall, this report provides a valuable contribution to the understanding of the factors that affect the price of Airbnb listings in New Zealand and suggests on marketing strategies, as well as some areas of further research and development of a more accurate prediction model.

# Reference

Agresti, A. (2013a). Categorical Data Analysis (3rd ed.). John Wiley & Sons. <https://mybiostats.files.wordpress.com/2015/03/3rd-ed-alan_agresti_categorical_data_analysis.pdf>

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.

Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192–1217. <https://doi.org/10.1080/13683500.2013.827159>

*Higher accommodation prices contribute to inflation | Stats NZ*. (n.d.). <https://www.stats.govt.nz/news/higher-accommodation-prices-contribute-to-inflation>

*Large-scale WebGL-powered Geospatial Data Visualization Tool*. (n.d.). Kepler.gl. Retrieved March 16, 2023, from <https://kepler.gl/demo/map?mapUrl=https%3A%2F%2Fdl.dropboxusercontent.com%2Fs%2Fzu7st0nszaqu8iy%2Fkeplergl_a3sbec9.json>

Ministry of Business, Innovation & Employment. *2 Regional and Seasonal Dispersal of International Tourists Tourism Insight Series*. 2016, www.mbie.govt.nz/assets/1f4279cc57/regional-and-seasonal-dispersal-report.pdf.

*New Zealand: international visitors by region | Statista*. (2023, January 17). Statista. <https://www.statista.com/statistics/687393/new-zealand-international-visitors-by-region/>

Perez-Sanchez, V. R., Serrano-Estrada, L., Marti, P., & Mora-Garcia, R. T. (2018). The what, where, and why of Airbnb price determinants. Sustainability, 10(12), 4596.

*Seasons in New Zealand | 100% Pure New Zealand*. (n.d.). <https://www.newzealand.com/int/seasons-in-new-zealand/>

University of Otago. (n.d.). *Urban Rural Profile Categories* [Dataset]. <https://www.otago.ac.nz/healthsciences/otago706060.pdf>

Zervas, G., Proserpio, D. M., & Byers, J. W. (2017). The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry. *Journal of Marketing Research*, *54*(5), 687–705. <https://doi.org/10.1509/jmr.15.0204>

# Appendix A – Data Summary

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Mean | Std dev | Min | Max |
| TECTs | 71355 | 40932 | 0 | 142783 |
| Description | 817 | 264 | 0 | 1196 |
| Latitude | -40 | 3.3 | -47 | -35 |
| Longitude | 174 | 3.9 | -177 | 178 |
| neighborhood\_overview | 227 | 267 | 0 | 1001 |
| Host\_since | 5.2 | 2.4 | 0 | 14 |
| host\_response\_rate | 98 | 9.5 | 0 | 100 |
| host\_acceptance\_rate | 92 | 16 | 0 | 100 |
| host\_listings\_count | 113 | 469 | 1 | 2217 |
| host\_total\_listings\_count | 153 | 632 | 1 | 2988 |
| host\_verifications | 2.1 | 0.45 | 1 | 3 |
| host\_has\_profile\_pic | 0.98 | 0.13 | 0 | 1 |
| host\_identity\_verified | 0.88 | 0.32 | 0 | 1 |
| Accommodates | 4.7 | 2.7 | 1 | 16 |
| bathroom\_number | 1.4 | 0.71 | 0 | 8 |
| bedrooms | 2.2 | 1.2 | 1 | 9 |
| beds | 3 | 2.1 | 1 | 22 |
| amenities | 36 | 15 | 1 | 102 |
| Price | 274 | 215 | 41 | 1627 |
| minimum\_nights | 2.4 | 19 | 2 | 1124 |
| maximum\_nights | 3941 | 583030 | 1 | 100000000 |
| has\_availability | 1 | 0.03 | 0 | 1 |
| availability\_30 | 11 | 9.2 | 0 | 30 |
| availability\_60 | 26 | 19 | 0 | 60 |
| availability\_90 | 45 | 28 | 0 | 90 |
| availability\_365 | 187 | 125 | 0 | 365 |
| number\_of\_reviews | 56 | 88 | 1 | 1103 |
| number\_of\_reviews\_ltm | 16 | 21 | 0 | 315 |
| number\_of\_reviews\_l30d | 2.1 | 2.50 | 0 | 32 |
| first\_review | 1035 | 833 | -1 | 4278 |
| last\_review | 99 | 255 | -1 | 4107 |
| review\_scores\_rating | 4.8 | 0.31 | 1 | 5 |
| review\_scores\_accuracy | 4.8 | 0.31 | 0 | 5 |
| review\_scores\_cleanliness | 4.8 | 0.34 | 0 | 5 |
| review\_scores\_checkin | 4.9 | 0.28 | 0 | 5 |
| review\_scores\_communication | 4.9 | 0.22 | 0 | 5 |
| review\_scores\_location | 4.9 | 0.22 | 0 | 5 |
| review\_scores\_value | 4.7 | 0.36 | 0 | 5 |
| instant\_bookable | 0.45 | 0.5 | 0 | 1 |
| calculated\_host\_listings\_count | 113 | 468 | 1 | 2212 |
| calculated\_host\_listings\_count\_entire\_homes | 112 | 468 | 0 | 55 |
| calculated\_host\_listings\_count\_private\_rooms | 0.67 | 2.9 | 0 | 55 |
| calculated\_host\_listings\_count\_shared\_rooms | 0.02 | 0.32 | 0 | 11 |
| reviews\_per\_month | 1.8 | 1.9 | 0.01 | 17 |

Table A 1 Summary Statistics of the cleaned dataset

# Appendix B – R Output

## B.1 Model Coefficients

Text

Description automatically generated

Figure B.1 1 The Coefficients of Multiple Linear Regression Model

|  |  |
| --- | --- |
| **Variables** | **Estimate** |
| Bathroom\_Number | 0.31551829 \*\*\* |
| bathroom\_typeprivate bath | 0.13065857\*\*\* |
| property\_defUnique stays | 0.09209887\* |
| review\_scores\_rating | 0.07583973\*\*\* |
| beds | 0.07059738\*\*\* |
| region\_typeUrban | 0.04984047\*\*\* |
| touristTourist | 0.04388052\*\*\* |
| property\_defStandard accomodation | 0.04022200\*\* |
| amenities | 0.00216670\*\*\* |

Table B.1 1 Variables with positive impact on price

|  |  |
| --- | --- |
| **Variables** | **Estimate** |
| number\_of\_reviews | -0.00127867\*\*\* |
| room\_typeHotel room | -0.08679251\* |
| bathroom\_typeshared bath | -0.37992369\*\*\* |
| room\_typePrivate room | -0.51698774\*\*\* |
| room\_typeShared room | -1.00368523\*\*\* |

Table B.1 2 Variables with a negative impact on price

## B.2 Model Assessment

Text

Description automatically generated

Figure B.2 1 Models RMSEs and Adjusted R-squared

Text

Description automatically generated

Figure B.2 2 Models AICs

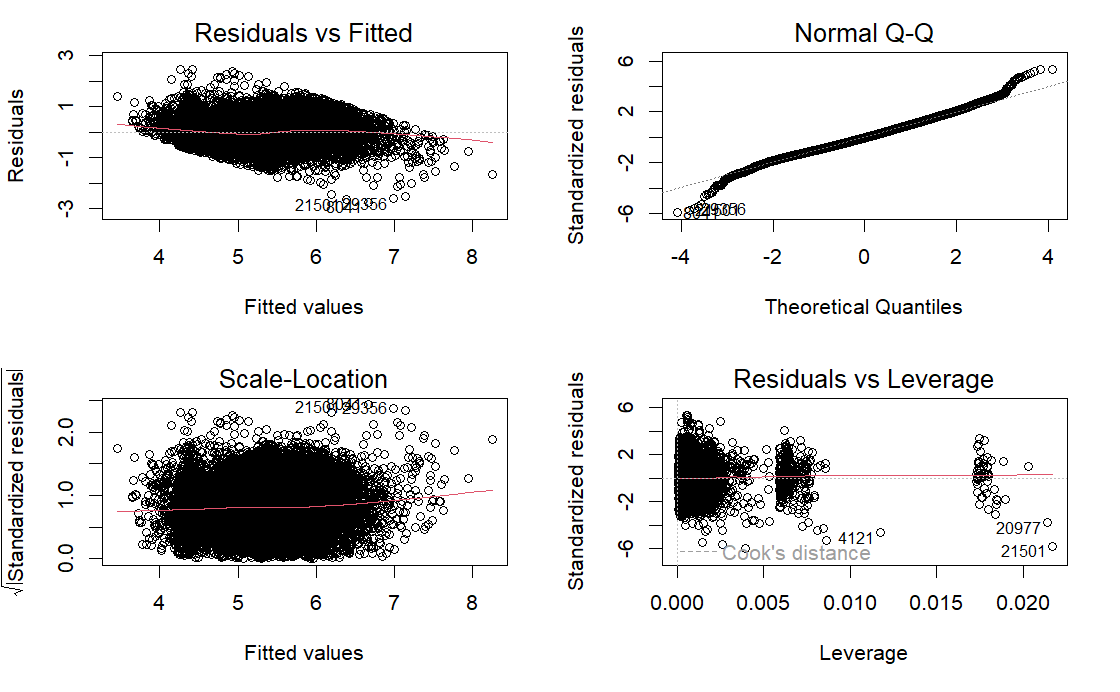


Figure B.2 3 Model 5 Diagnostic Plots

# Appendix C – Geospatial



Figure C 1 Price Legend

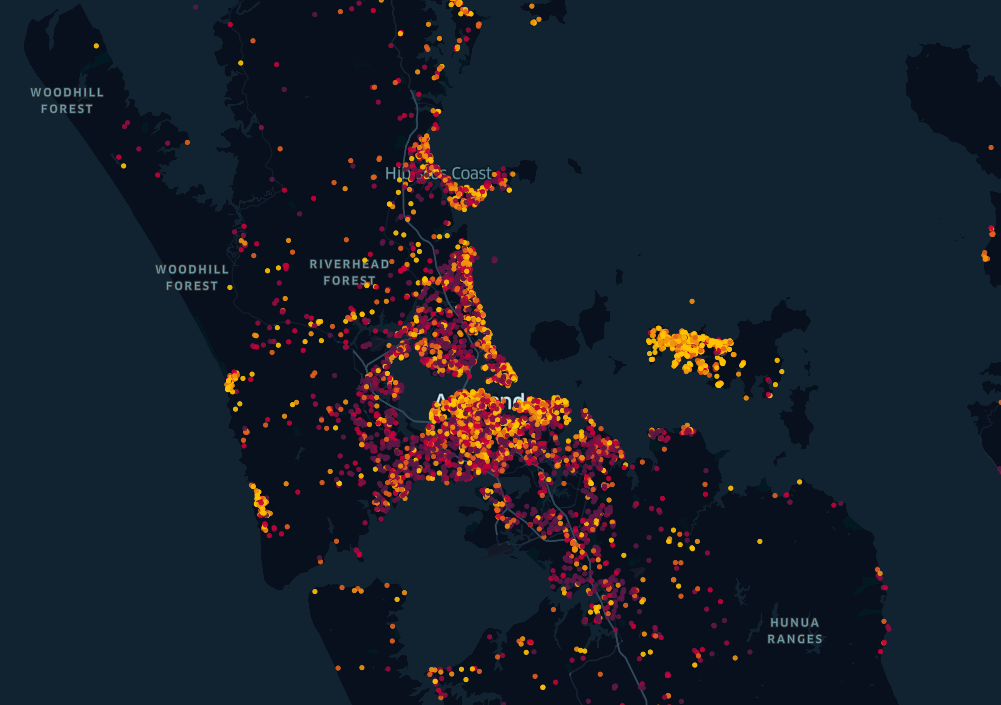


Figure C 2 Auckland Location

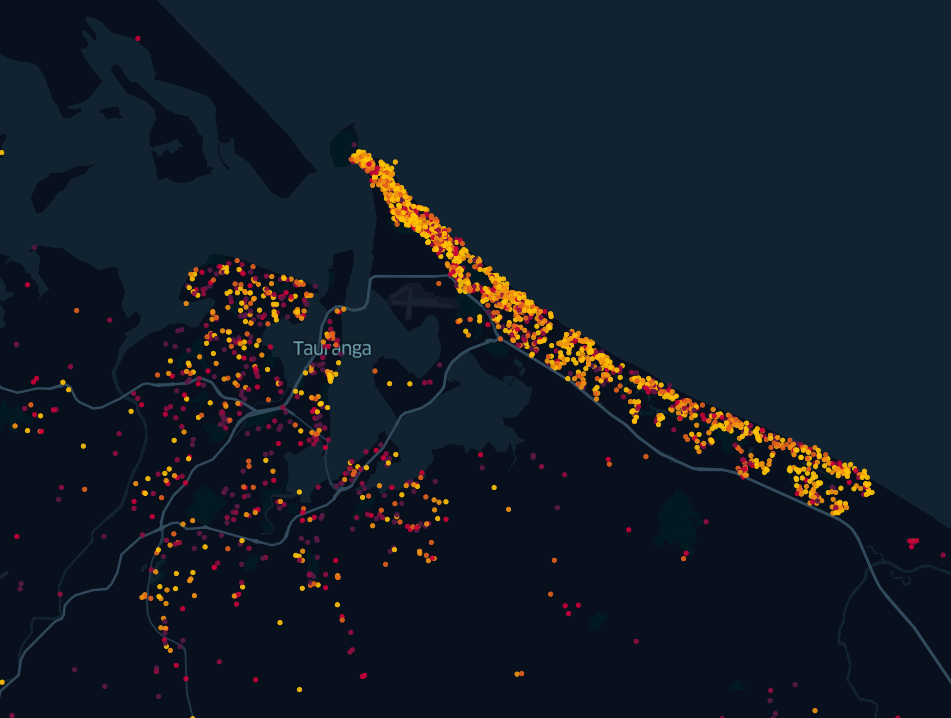


Figure C 3 Tauranga Location

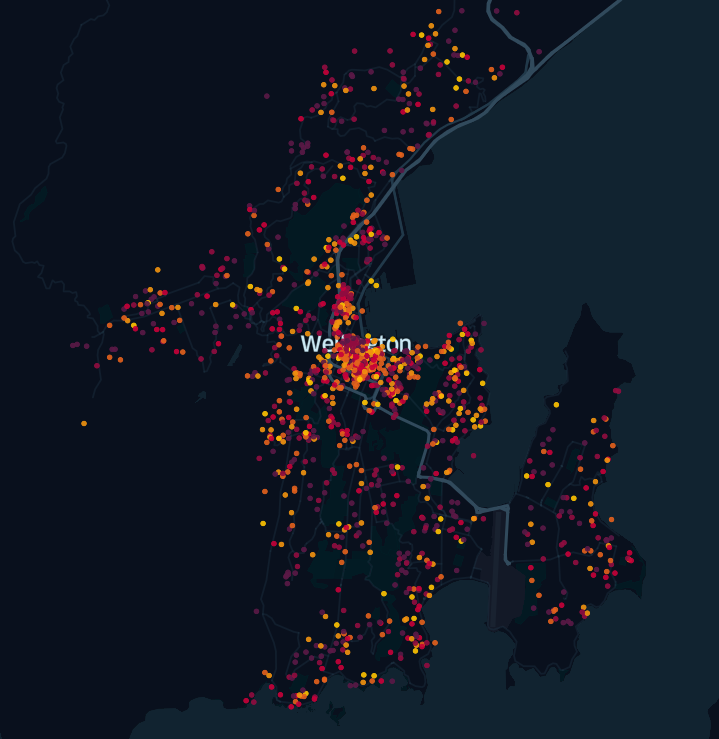


Figure C 4 Wellington Location