Distribution Patterns and Price Determinants of Airbnb Listings: A New York City Case Study

Daniel K Addai-Marnu

Abstract
In this study, we look at Airbnb data in New York City in the form of listings attribute and spatial distribution. I analyse spatial distribution in relation to price, listing type, and geographic information. Although boroughs are very different in terms of population composition, size and wealth, Airbnb listing density and price show a pattern of increase. Areas of high Airbnb presence and higher prices are those inhabited by the wealthier groups, the centre of the city and with good economic development. Finally, using original and newly engineered features, regression analysis is performed to determine which features affect the pricing Airbnb listings.

Keywords: Airbnb, Listing Type and Distribution, Borough Distribution, Regression Modelling

INTRODUCTION

Airbnb is a lodging service that allows people to rent their unused rooms or property directly via computer-based interactions to potential guests. Founded in 2008, Airbnb has grown exponentially over the last few years and now has more than 7 million accommodations worldwide.[1] Today, more people and investors have caught up in the trend and understood the success and prospect in Airbnb. As interest in Airbnb has increased, it is worth conducting analysis on pricing, type of accommodation and spatial location or distribution of listings to better inform investments strategies and maximise benefit.

1.0 DATA, QUESTION AND PLAN

1.1 DATA SOURCE

Airbnb data: The data was taken from Airbnb data made available on "Inside Airbnb: Adding data to the debate" website.[2] The website periodically publishes data on Airbnb listings around the world. The data used was the summary data on New York City, New York listings compiled in September 2019.

Spatial data: The spatial data is from data made available on "NYC OpenData" aimed at engaging New Yorkers in the information produced and used by the city government.[3]

1.2 RESEARCH QUESTION

In this study, we analyse the Airbnb and spatial data of New York City, with the aim of addressing the following research questions:

- Which factors explain the distribution of Airbnb listings across the boroughs of New York City? We investigate the concentration of listings, types of listings distribution, the effect of subway connectivity and places of interest in the distribution of listings in the five boroughs of New York City.

- Price determinants and their effect on Airbnb listing pricing in New York City? We investigate the possibility of using predictive models to find the features that affect the pricing of Airbnb listings.

1.3 Analysis strategy

In the light of these tasks, our first goal is to explore the data to provide insight. Secondly, modelling of features to determine what factors drive price. The following approach will, therefore, guide the achievement of this goal:

- Data loading and preparation: load and study data characteristics.
- Data preprocessing: clean data for exploration and analysis.
- Engineer new features.
- Data analysis: descriptive, exploratory and spatial analysis.
- Data modelling: preprocessing for modelling (encoding, standardization) and modelling.

The subsequent section summarises the findings obtained and comments on their usefulness.

2.0 Findings and reflections

2.1 Findings

2.1.1 Distribution of listings

The first thing investigated was the factors that explain the distribution of listings. We analysed the distribution, concentration and type of listings, then the influence of subway connectivity and tourist attraction on the distribution.

Manhattan has the highest number of listings followed closely by Brooklyn. These two boroughs share more than 80% of the listings. Queens, Bronx and Staten Island follow respectively in number with Staten Island having less than 1% of the total listings.

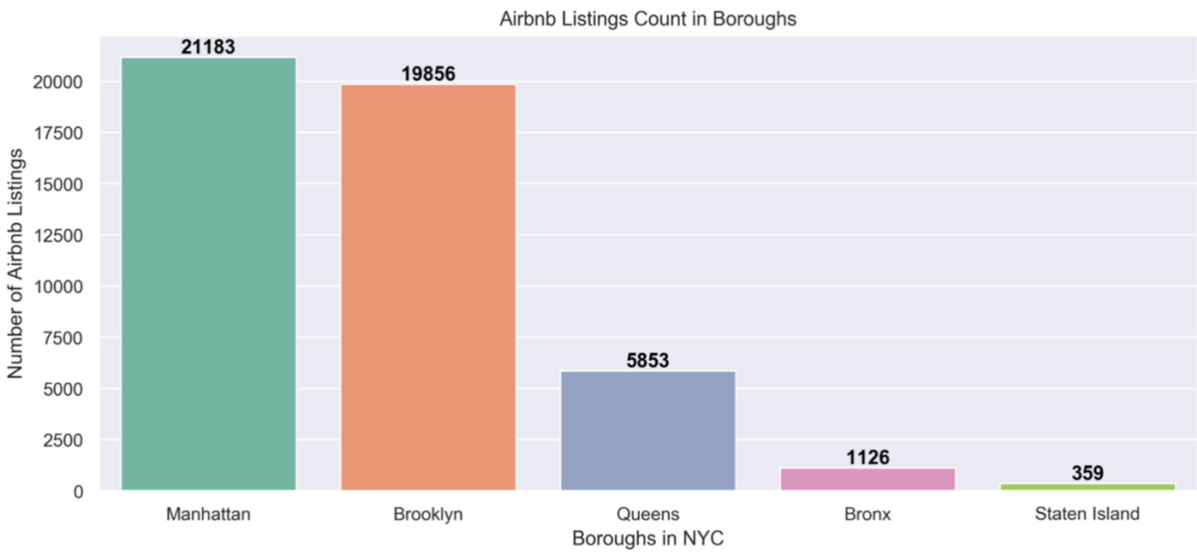

Figure 1: Airbnb Listings Count By Borough

Further analysis of listing type distribution by boroughs shows that most of the listings are either an entire home/apartment or private room. Few of the listings are shared room or hotel room type. Bronx and Staten Island have no hotel room listing.

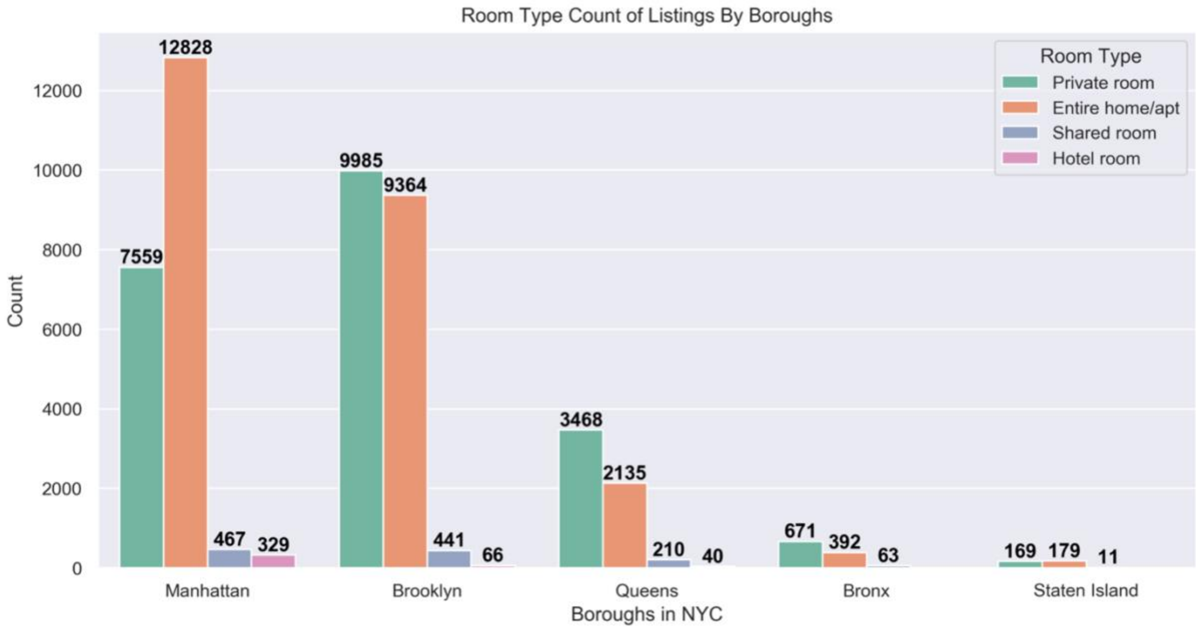

Figure 2: Airbnb Listings Type Count By Borough

Availability of listings type and availability by borough was investigated to see why more of the listings are entire home/apartment and private room. From figure 3, most of the entire home/apartment and private room listings have higher short availability over the year whereas that of shared room has almost equal distribution of long and short availability over the year. Hotel rooms are mostly available throughout the year with a heavy head in the chart.

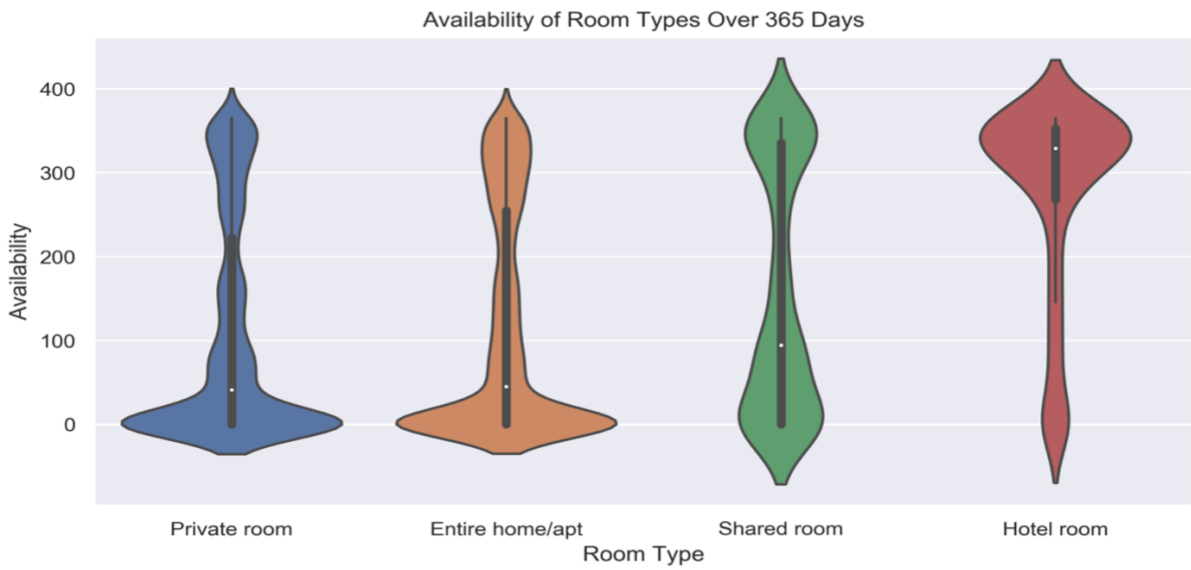

Figure 3: Listing Type Availability Over 365 Days

Looking at the availability of the listings by borough, most of the listings in Manhattan and Brooklyn which consist of mainly entire home/apartment and private room have short availability over the year. Queens, Bronx and Staten Island have similar availabilities with almost even spread of availability. There seems to be a correlation between availability of listing type and availability of listings by boroughs. Majority of the listings in Manhattan and Brooklyn have short availability which happen to be entire home/apartment and private room listing types.

Figure 4: Borough Listings Availability Over 365 Days

The heatmap below shows the concentration of listings. It can be seen that the listings are concentrated throughout Manhattan, considering Manhattan has the highest population of listings. There is also a high concentration of listings in part of Brooklyn neighbouring Manhattan and centre of the city.

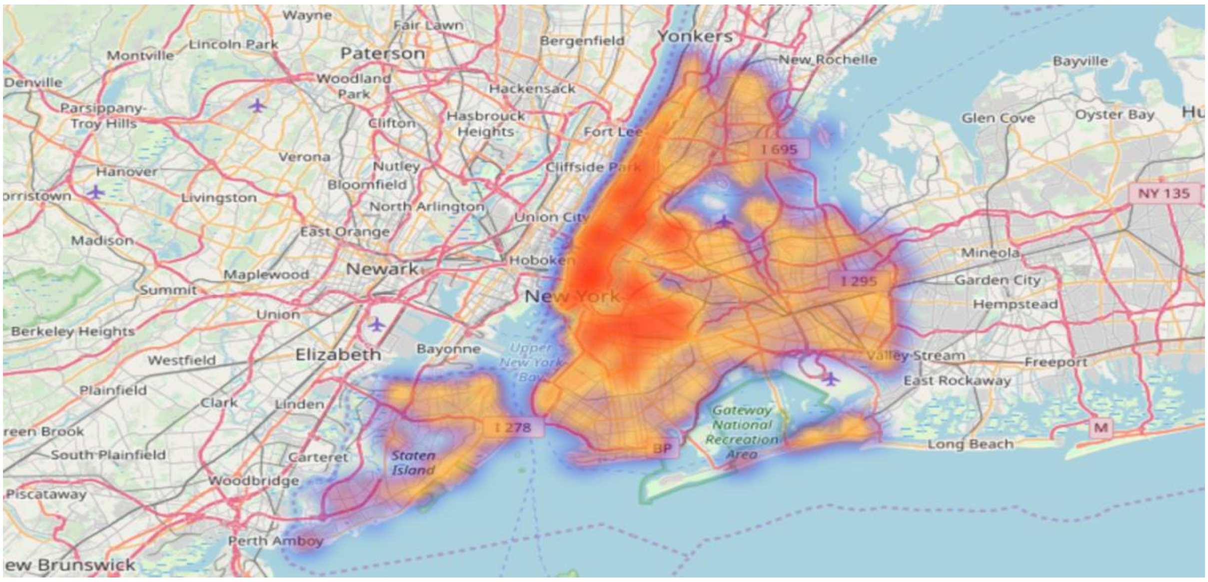

Figure 5: Heatmap of Listings Concentration

Figure 6 shows that Manhattan and Brooklyn have good subway connectivity, evident by the number of subway stations. These two boroughs happen to have the highest number of listings with most of them clustered in all Manhattan and part of Brooklyn adjacent to Manhattan. Queens, Bronx and Staten Island comparatively have less subway stations, thus, less number of listings.

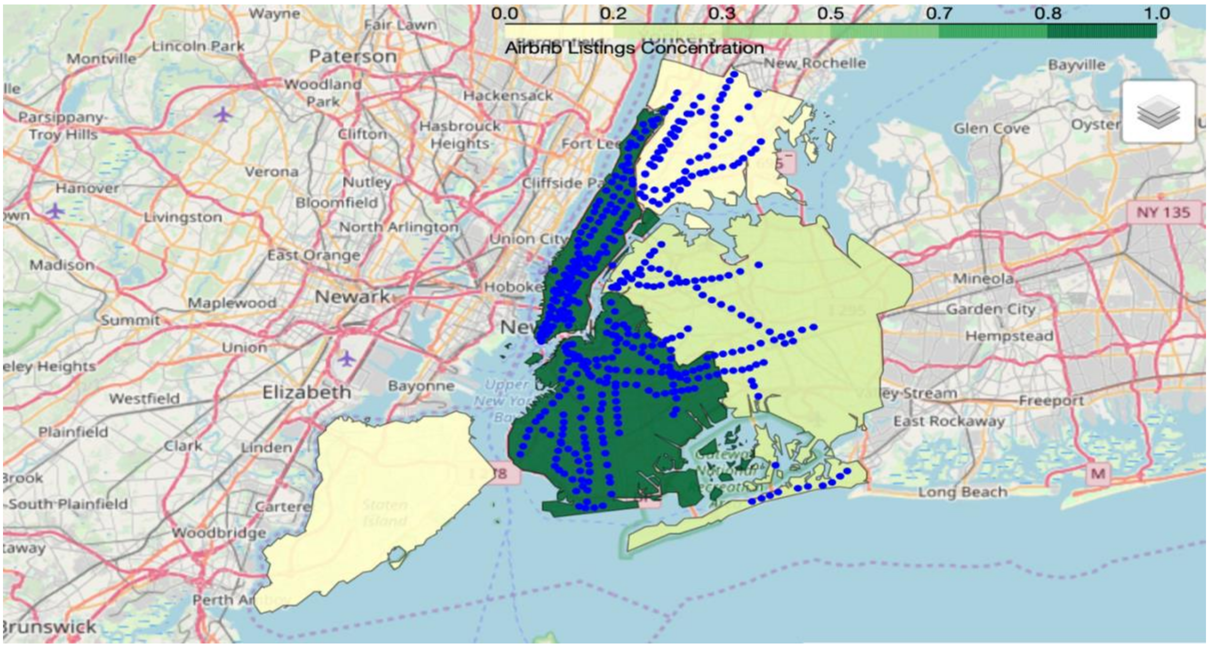

Figure 6: Subway Stations (Blue Dots) And Concentration of Listings

From figure 7, it is noticeable that places of interest are scattered throughout the city and in all boroughs. It can be seen that the location of places of interest, in general, has no bearing on the distribution of listings.

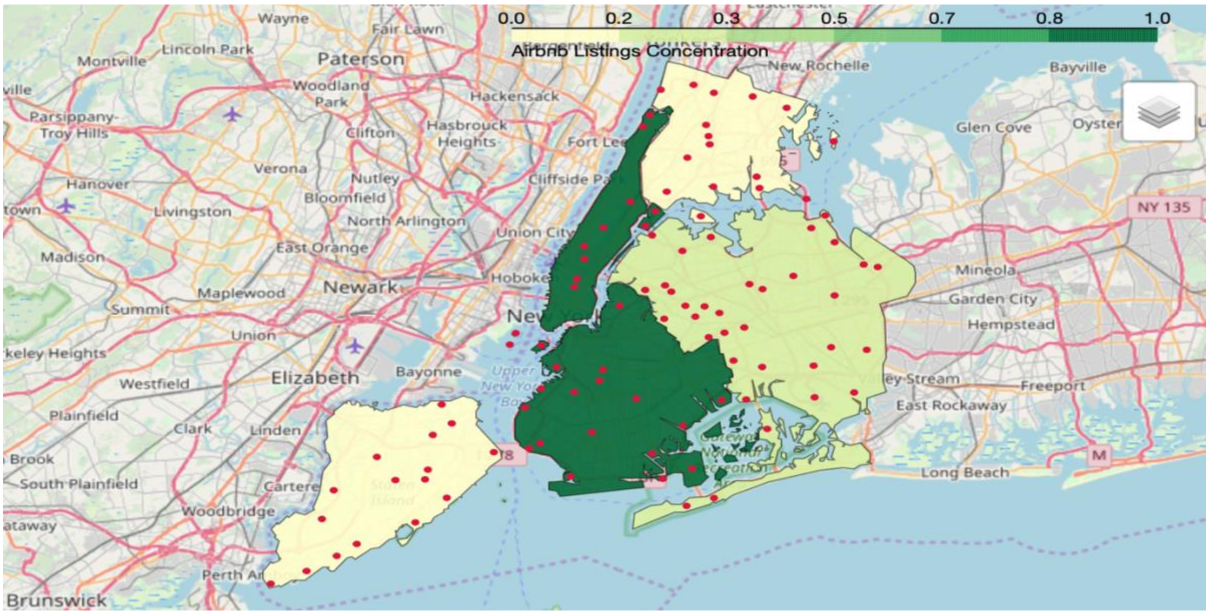

Figure 7: Places of Interest (Red Dots) and Concentration of Listings

2.1.2 Determinants of Price

In finding the determinants of price, two models were used to find the importance of features in predicting price of listings. Before modelling, three extra features were engineered; the minimum night spend (which is the amount spent per minimum stay), the estimated total annual income of the listing and the number of days after the last review. The two models used were xgboost regressor and random forest regressor for comparability.

Table 1 shows the R-squared of the two models, which is a measure of how close the data is to the fitted regression line. The random forest regressor performs better than the xgboost regressor, but they all perform well on the dataset.

R - Squared	XGBoost Regressor	Random Forest Regressor
Training	0.9675	0.9985
Validation	0.9656	0.9932

Table 1: R-Squared of Regressors

The determinants of price was arrived at by finding the feature importance of the regression models. The prominent features from the xgboost regression are listing type (notably entire home/apartment and private), borough/location (notably Manhattan, longitude and latitude), potential annual income from listing, minimum night price, minimum stay and availability over the year. Figure 8 shows in order how each feature is essential in determining the listing price of the xgboost regression.

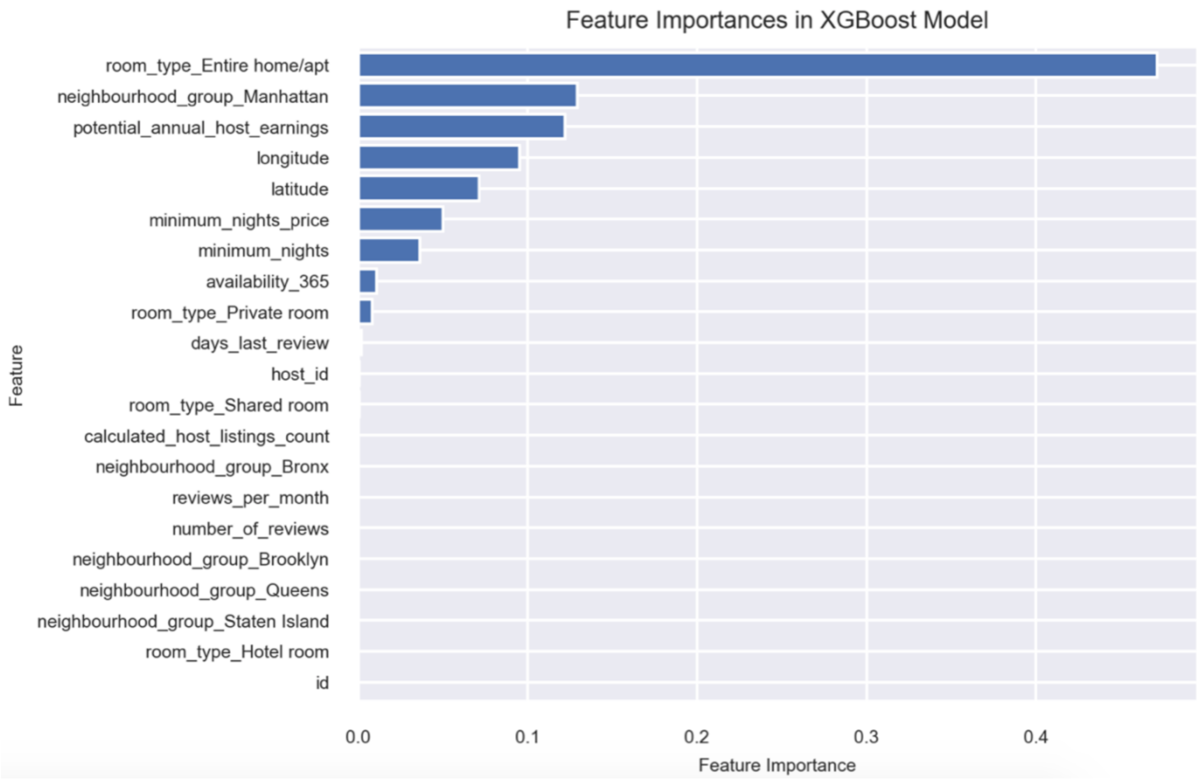

Figure 8: Feature Importance of Xgboost Regression

The prominent features from the random forest regression are listing type (notably entire home/apartment), minimum night price, potential annual income from listing, minimum stay, borough/location (notably Manhattan and latitude) and listings available. Figure 9 shows the feature importance from the random forest regressor.

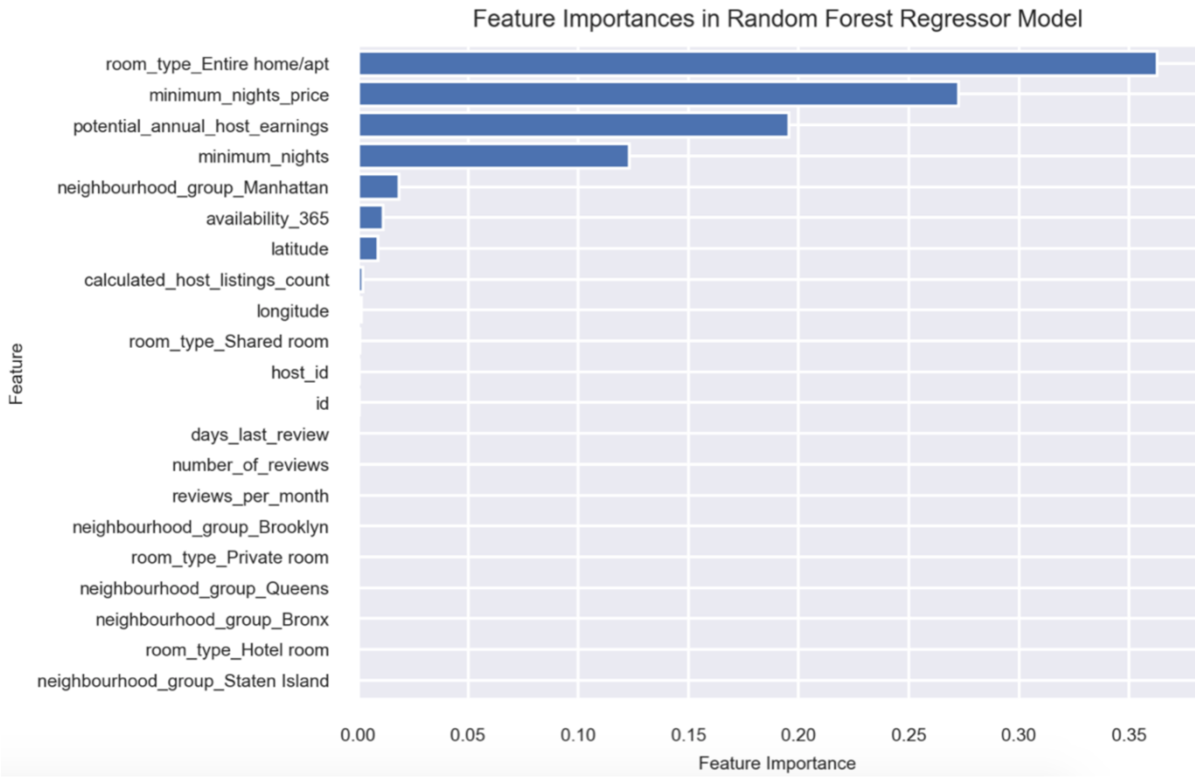

Figure 9: Feature Importance of Random Forest Regression

2.2 REFLECTIONS

From the analysis, Airbnb listings follows a general pattern distribution. Listings are concentrated in Manhattan and centrally in Brooklyn. Majority of the listings are entire home/apartment and private room and these dominate in all boroughs. Few of the listings are hotel room which indicate hotels are realising the benefit and competition Airbnb brings in the accommodation industry. Most of the entire home/apartment and private room listings have shorter availability over the year, which shows that hosts who have acquired such properties solely for Airbnb are not many. Most hosts of such properties list them for extra income outside their primary use.

The analysis did not produce any positive correlation between the distribution of listings and the places of interest, but this is not definitive. Potential study of individual tourist sites may help find how they affect the distribution of listings. The analysis shows, however, that subway accessibility influences the distribution of listings, in that the boroughs (Manhattan and Brooklyn) with good subway connectivity have a high number of listings.

Finally, the two price models demonstrate that price is primarily determined by the type of listing, the location, the cost of minimum stay, the estimated annual income from listing, the minimum stay and the availability over 365 days. Entire home/apartment and private room listings affect the price of the listings as they offer more privacy than shared or hotel room listings. The minimum stay and its associated cost also affect the price of listings as more extended minimum stay will apply discounts. Likewise the longer the listing is available over the year, the higher the estimated annual income.

References

1. Airbnb About-Us, accessed 10 December 2019, https://news.airbnb.com/about-us/
2. Inside Airbnb: Adding data to the debate, accessed 30 November 2019, http://insideairbnb.com/get-the-data.html
3. NYC OpenData, accessed 1 December 2019, https://opendata.cityofnewyork.us/data/