# Analysis of Seattle's AirBnB Data

https://github.com/ryan-odea/678-AirBnB-Analysis

Ryan O'Dea
Boston University GRS MA678

Abstract

Begun in August of 2008, AirBnB is an American vacation rental marketplace. Without owning any of the real estate listings, AirBnB connects hosts, users that are willing to share their home for payment, to guests. With data scraped from InsideAirbnb This project seeks to understand the relationship between how a host will price their home or single room in the Seattle area compared to factors of location, type of room being offered, and if the host is considered a Super Host, someone who AirBnB has designated as providing a "shining example for other hosts." On the plot below, we can observe a very noisy data set when simply plotting of listing ID vs price; however, there are some key takeaways. It appears many of the listings are in the Downtown Area and we generally see a trend around $100 per night in most of the areas.

Analysis of Seattle's AirBnB Data
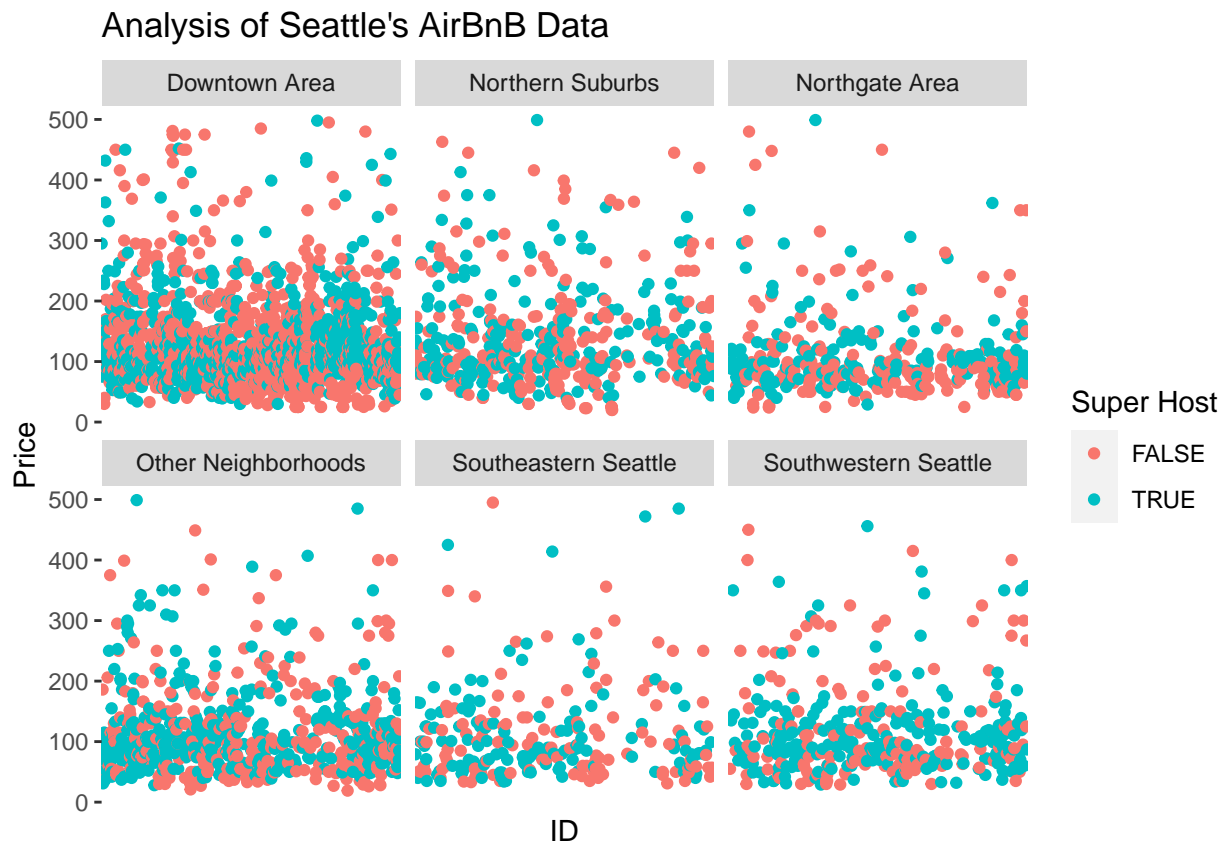
Basic EDA

Introductory Analysis

With the beginning exploratory data analysis, we can confirm that the majority of rooms are in the Downtown Area and renting an entire home/apartment is generally more expensive than renting a private room with the average entire home/apt with a mean of $136 per night while private rooms are about half with a mean of $68. We also observe that the overwhelming majority of listings are full home/apt options with few private rooms available. As conjecture, this could be attributed to the COVID-19 pandemic - less hosts want to invite users into their homes while they are there, and would like to have little contact with them.
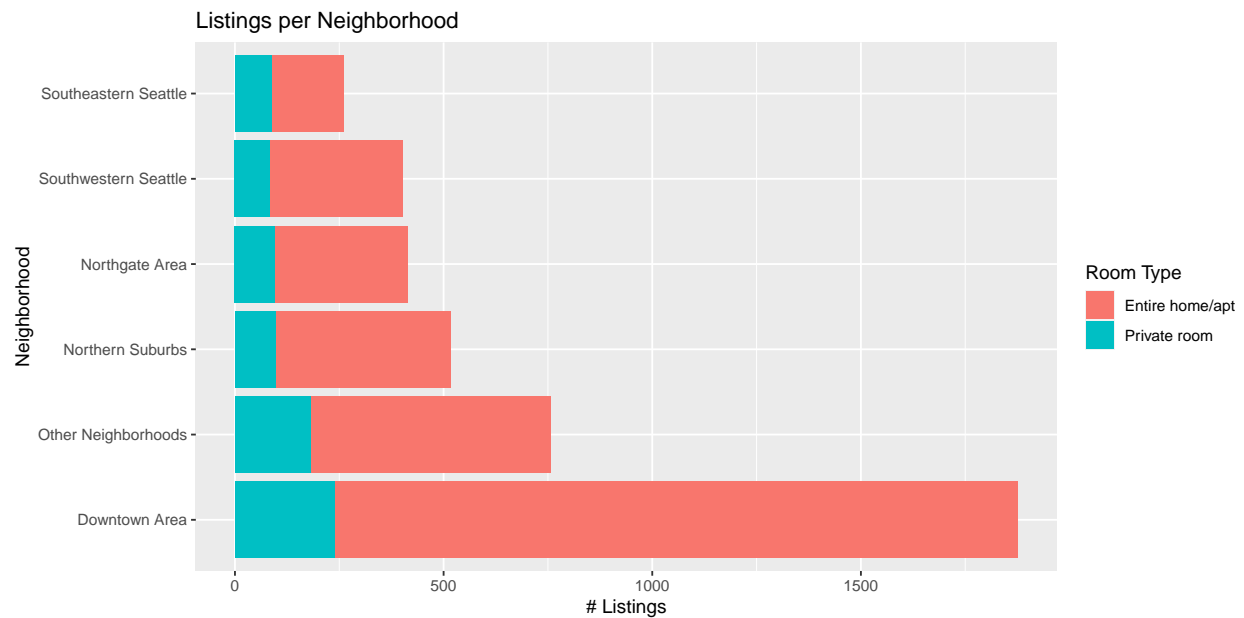
Spatial Analysis EDA

As expected, the location of the home also plays an important role in determining the price per night. Sorting by zipcode, the downtown area is generally more expensive than the other areas in King Country. We also see Lake City (categorized into "Northgate Area") in the north falls into the high average price as an outlier for zip codes in it's similar grouping. The area around SEATAC is the lowest average priced.

Expensiveness of Seattle's AirBnB's

Average Price
- [81 to 102)
- [102 to 107)
- [107 to 115)
- [115 to 121)
- [121 to 125)
- [125 to 149)
- [149 to 280]
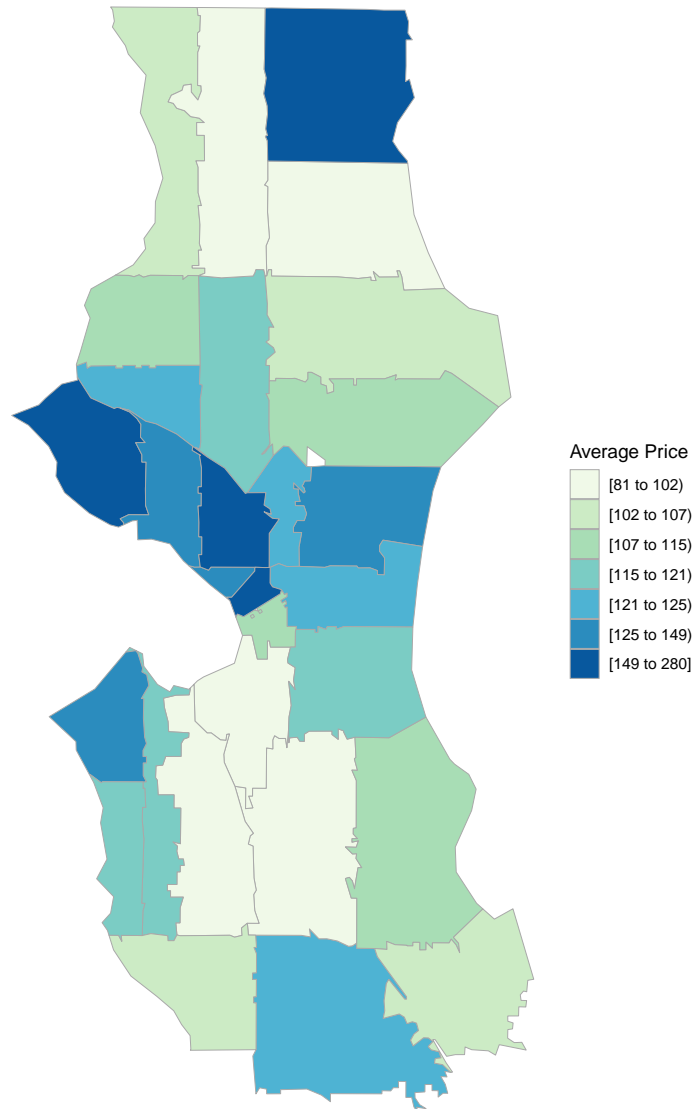
Exploring Data Relationships

Relation Between Superhost

Comparing points and violin plots, there is no apparent relation between being a Super Host and the price per night of the AirBnB. The violin plots appear to have mostly even means, the point plot also shows Super Hosts are interspersed with non Super Hosts when grouped by neighborhood.

Price by Type of Room per Grouped Location


Price vs Neighborhood Group
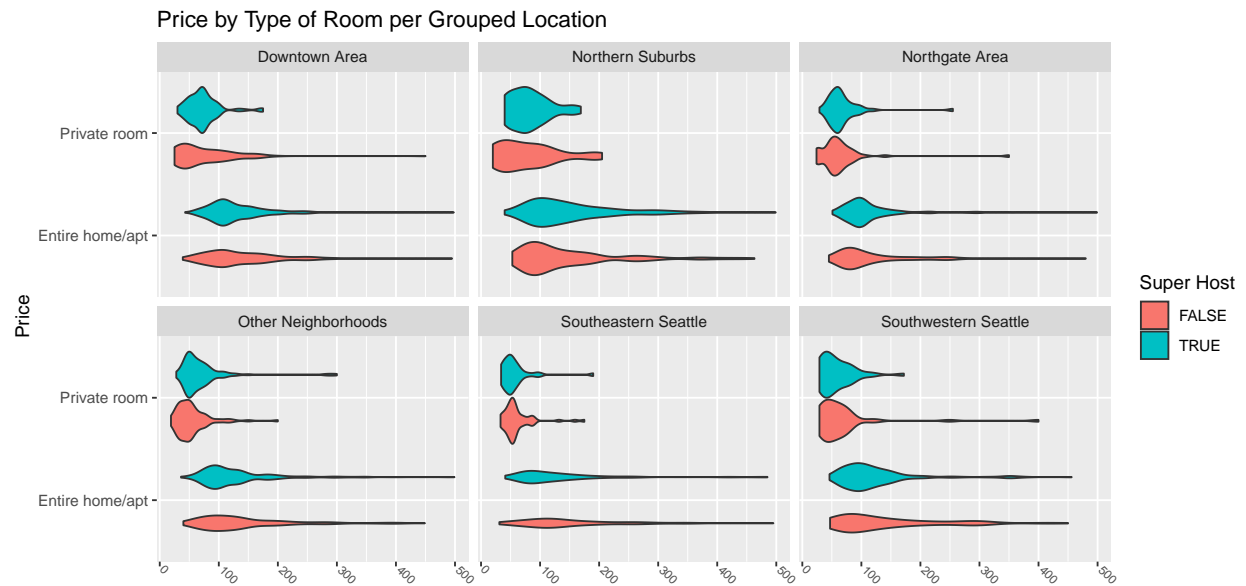
Model Fitting

When fitting a model, the log was taken of the price as a variance stabilizing measure. I chose an additive model which observed neighborhood group and room type as predictors. The fit had moderate residuals and an RMSE of 80 indicating a moderately good fit for how noisy the original data was.

The coefficients show both locations and room type are significant when pricing a room; however, my first model included Super Host status as a predictor - which was removed in favour of a better model fit and because it was found to be insignificant.

In an analysis of the coefficients, with a baseline intercept of $\exp(4.8)$ - the average price of a full house/apt in the Downtown Area. The other areas are generally less expensive, with the exception of the northern suburbs which are approximately the same, as seen by the summary, additionally private rooms are approximately $\exp(0.68)$ less than their whole house/apt equivalents. Bootstrapping was done to test the coefficients and yielded that the true coefficients are close to the approximate. The full table of coefficients can be found in the appendix (table 1). These coefficients show the log relationship between our baseline Downtown full home/apt the other respective groupings.
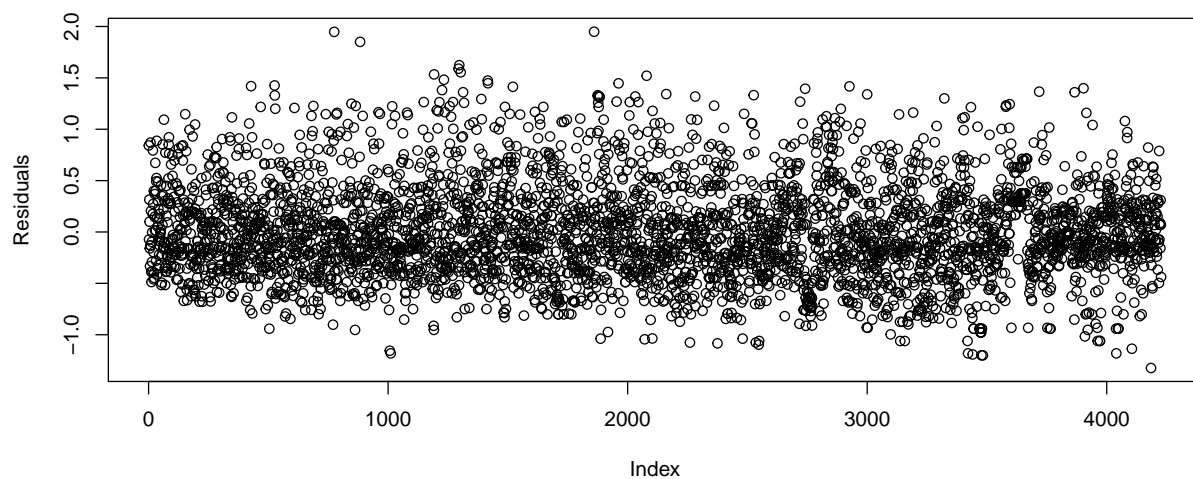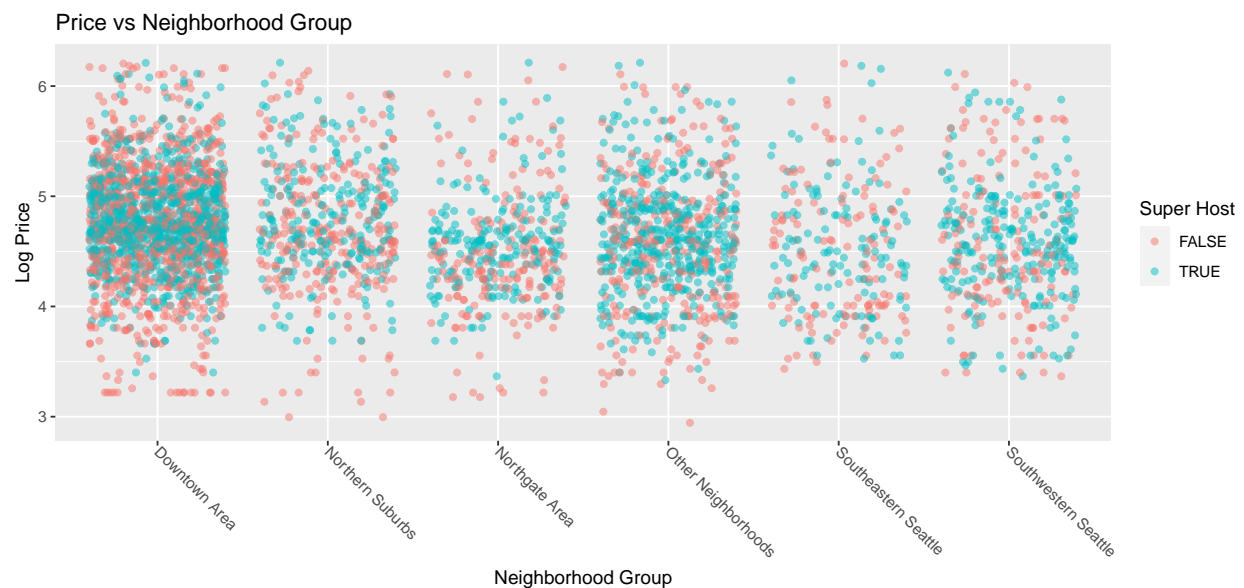
Appendix

Table 1

```
## 
## | Predictor | Estimate | Std. Error |
## |-----------|----------|------------|
## |Intercept  | 4.844    | 0.011      |
## |Nor Suburb | 0.036    | 0.023      |
## |Northgate  | -0.153   | 0.025      |
## |Other      | -0.079   | 0.019      |
## |Southeast  | -0.087   | 0.030      |
## |Southwest  | -0.117   | 0.025      |
## |Private Rm | -0.684   | 0.018      |
```