Price Prediction for AirBnB using Linear Regression

Bradley Ellisor

Objective

To entice landowners to list their property on AirBnB by giving them a model to estimate their price per night
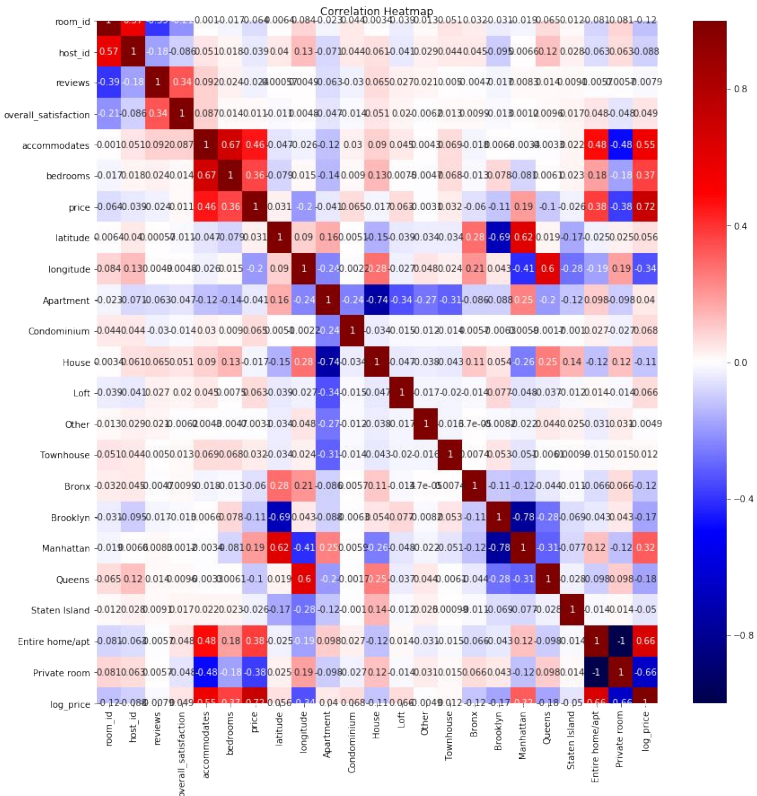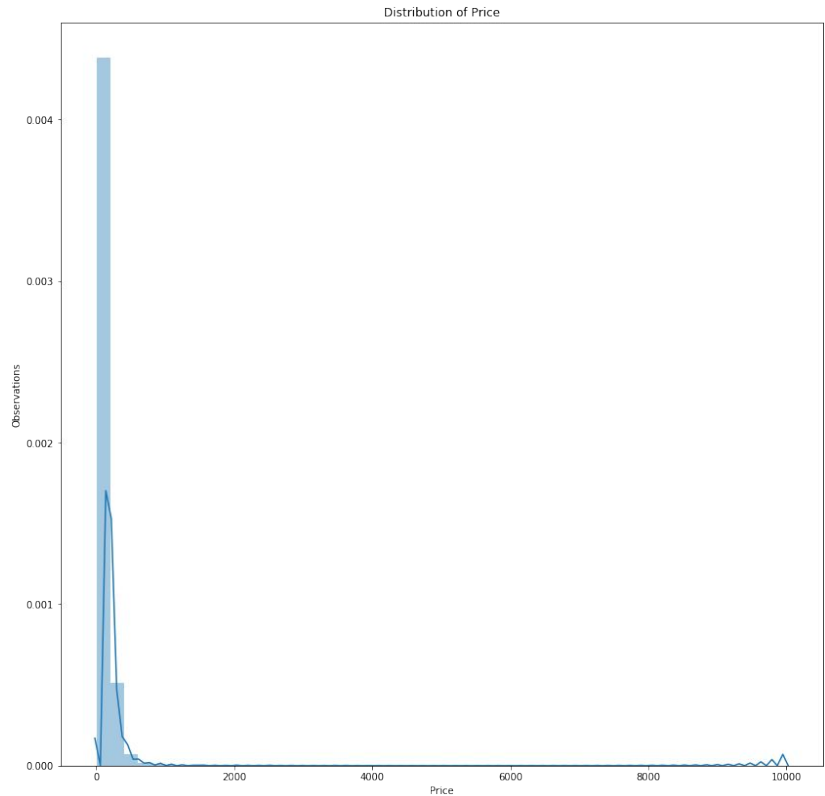
Methodology

Data Source

- Tom Slee Blog
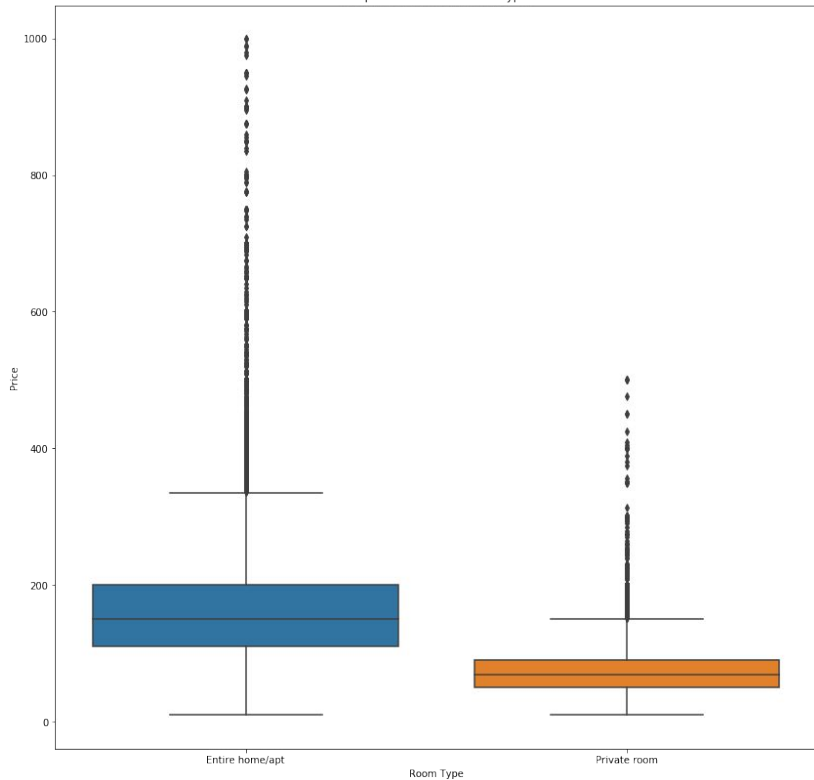
Primary Tools

- Pandas, StatsModels, Scikit-learn

Scope

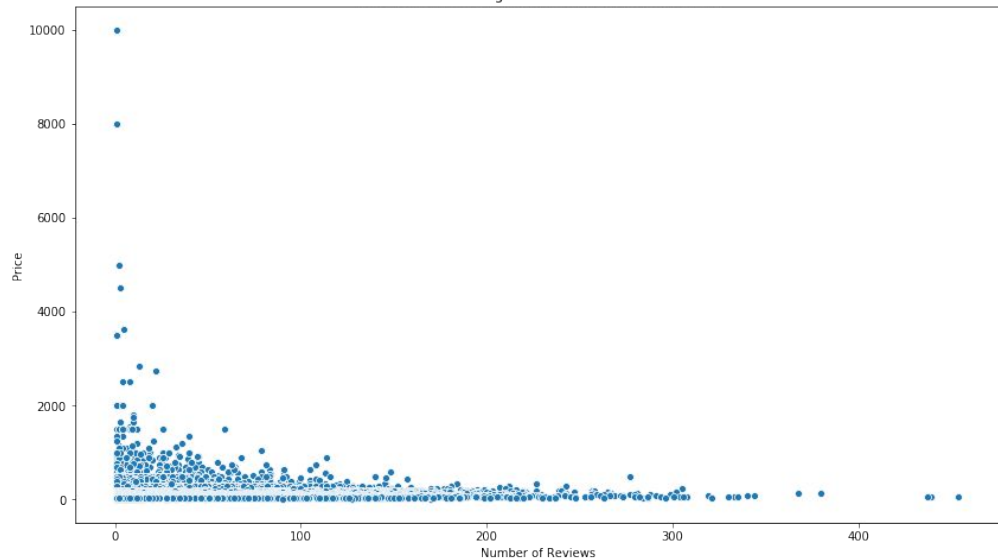- 2017 Survey Data of NYC AirBnB

Early Data Exploration

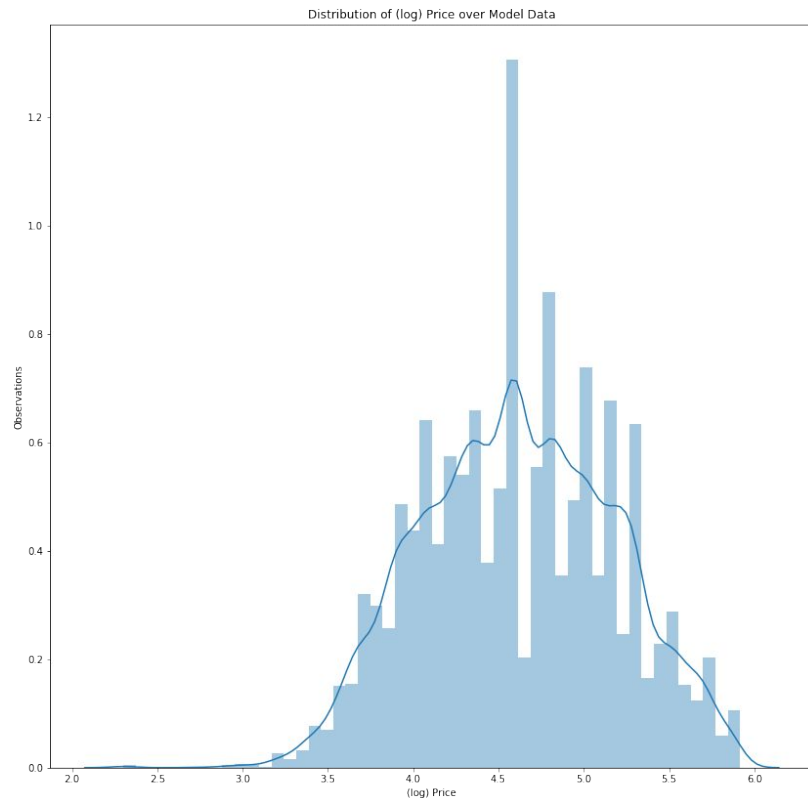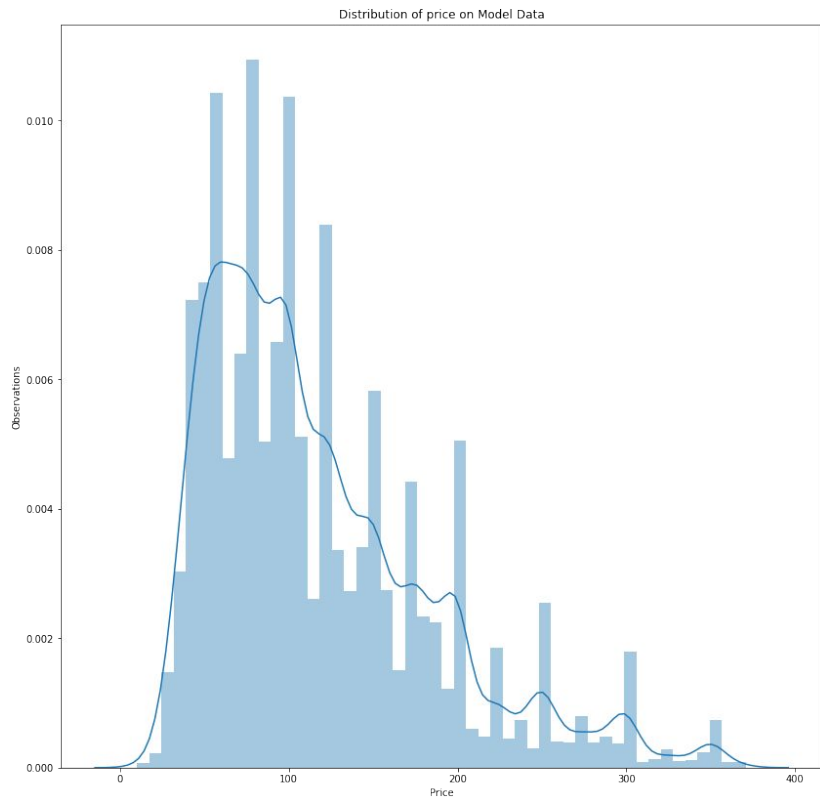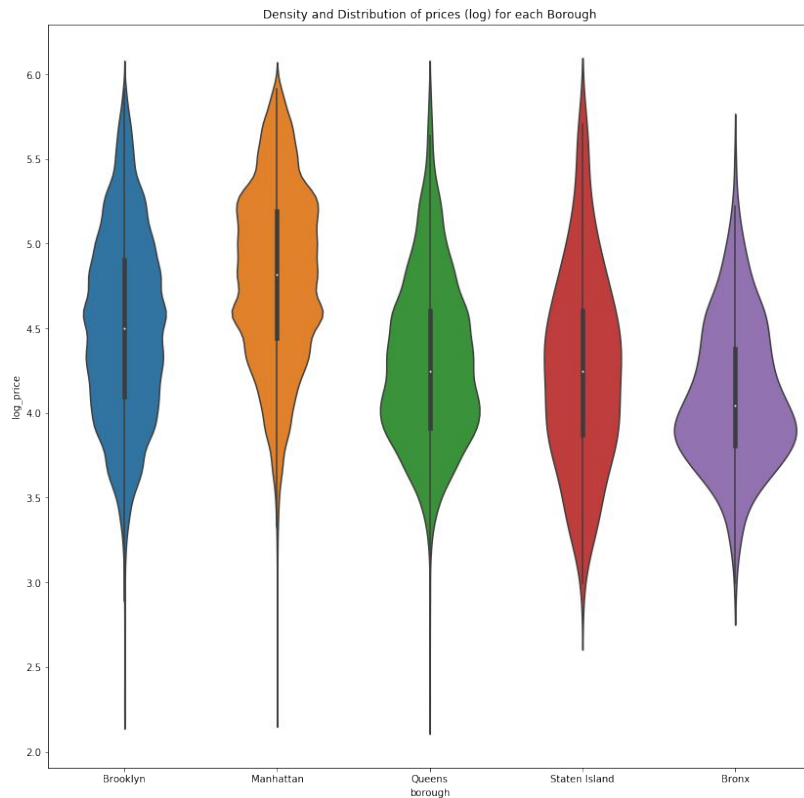# Early Data Exploration

Box plot of Different Room Types

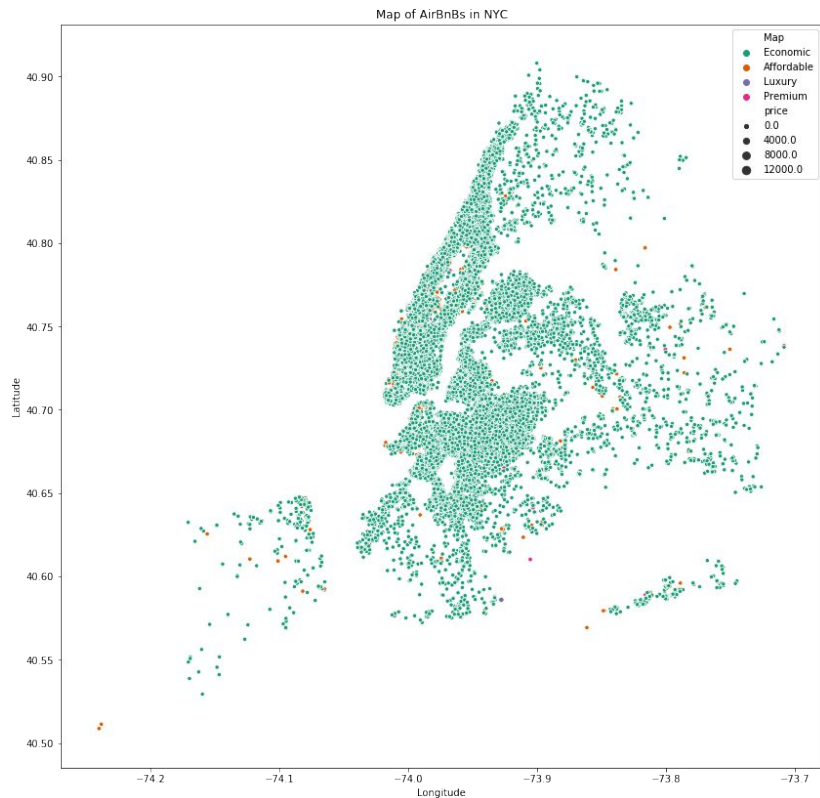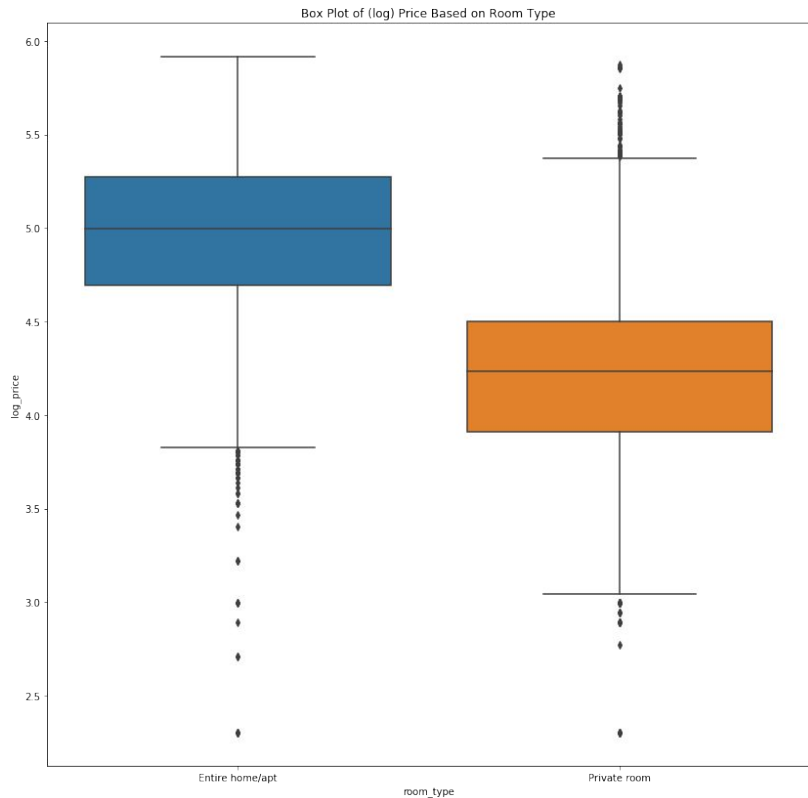Scatter Plot showing Price and Number of Reviews

Transformation

The Five Boroughs
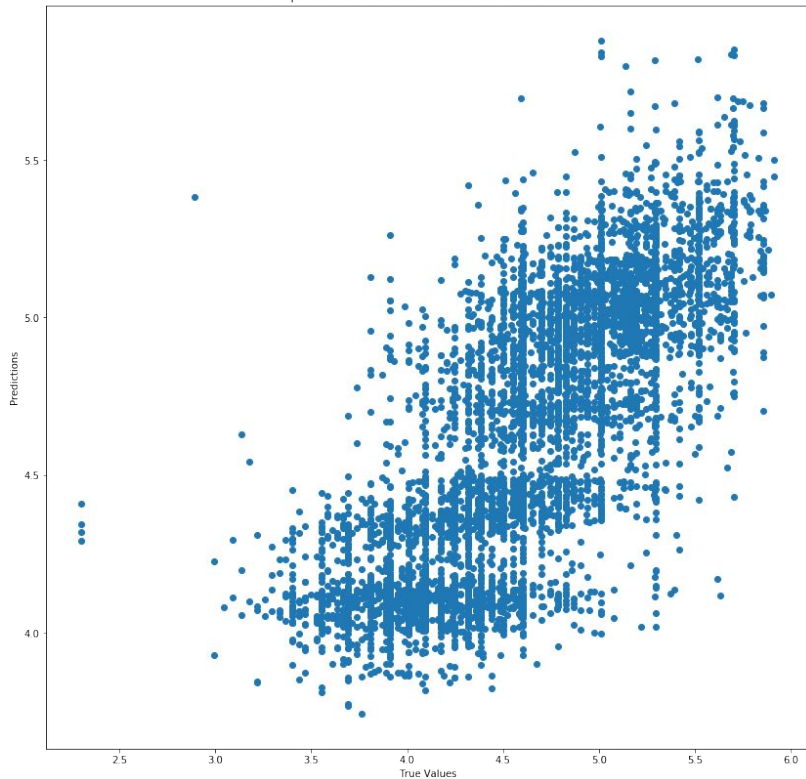
Type of Room

Box Plot of (log) Price Based on Room Type

Extremely low P-Value showing difference between Room Type Accommodations

Model

Comparison between True Values and Predicted Values

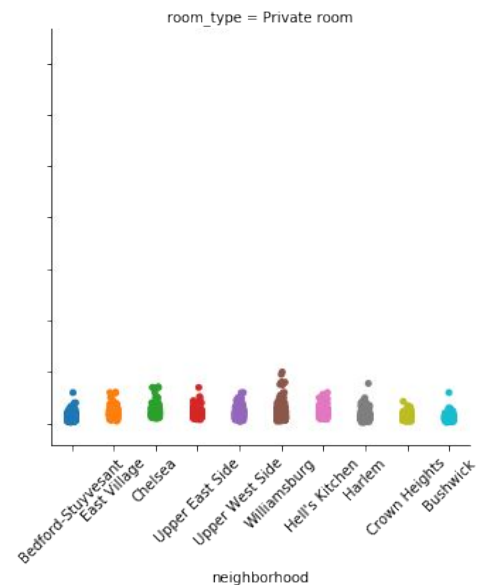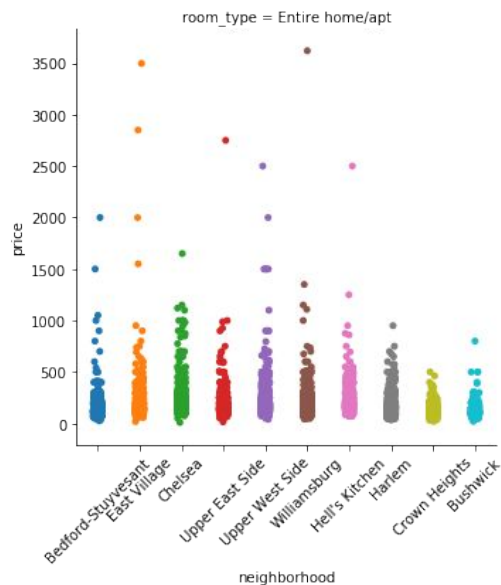R Score   0.579

KFold Cross Validation Mean 0.571

Recommendation

- Accomodation Size
- Rent Entire House
- Brooklyn and Manhattan

Most Popular Neighborhoods

Williamsburg
Bedford-Stuyvesant
Harlem
Bushwick
Upper West Side
East Village
Hell's Kitchen
Upper East Side
Crown Heights
Chelsea

Further Study

- Examine the Clusters in the model
- Include Proximity to Attractions, Entertainment, Groceries etc...

Questions?