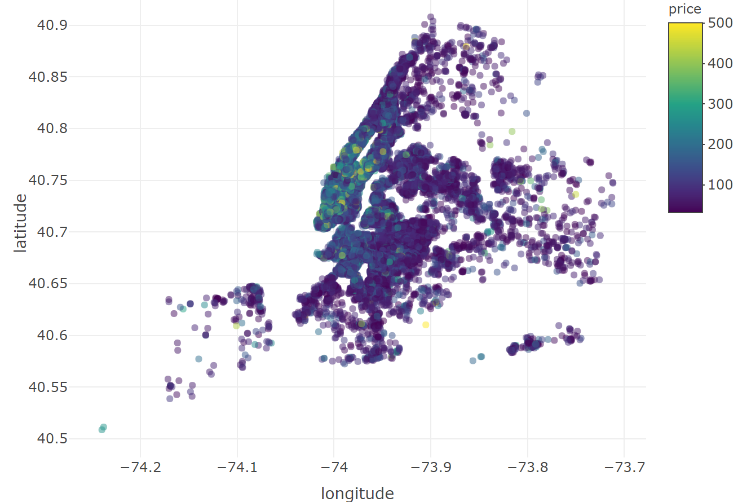
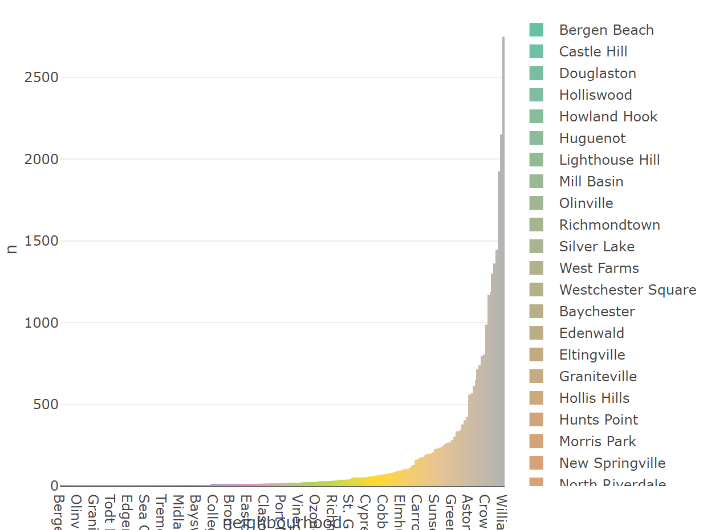
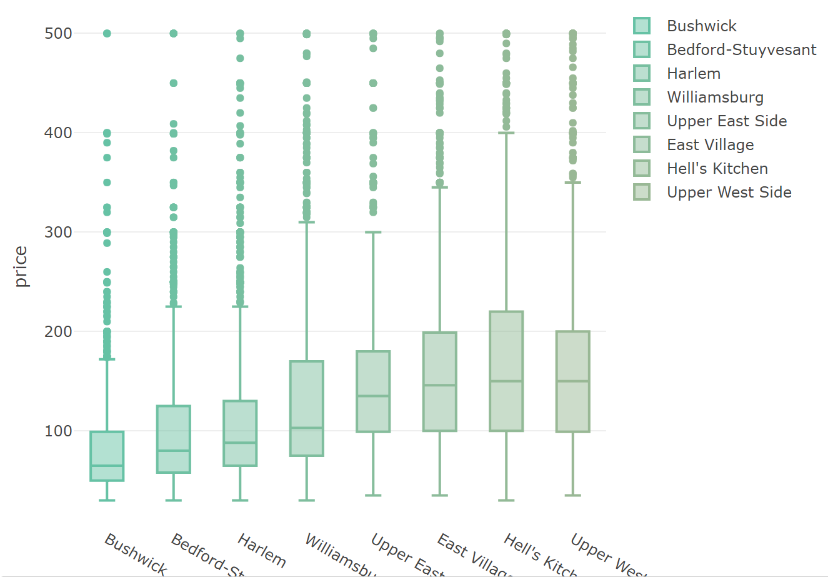
**Price Analysis and Borough Prediction of Airbnb housing in New York City**

Kwangmin Kim (kk3225), Leiyu Yue (ly2428), and Kathryn Addabbo (kma2162)

Countless tourists and students visit to explore all the attractions NYC has to offer as one of the largest cities in the world. Being a dynamic environment, the overall population of the city fluctuates daily. To handle the influx of individuals, information on available housing is extremely important. One of the most well-known housing companies is Airbnb. In this report, we analyze the Airbnb housing affordability and availability in each of New York City’s boroughs using the company’s released dataset in September 2017. The aim of this report is to provide clear visualization and meaningful interpretation of the data for convenience of visitors and all interested in the service of Airbnb.

One of the covariates of interest (price) has values ranging from $10 to $10,000 per night. We restricted the analysis to values between $30 and $500 because the average consumer does not have the appropriate funds to spend thousands of dollars a night on a room for a night, and we do not want too much variation in our analysis. The rating of each property was on a scale of 1 to 10, we changed the rating to a 5 – star scale and assigned zeros to the properties with no ranking. We also assigned zeros

Housing is widely available through all five of the boroughs, however generally the closer to downtown Manhattan the property is, the more expensive the price per night as seen in Figure 1. Of the properties, the majority are in Williamsburg and Bedford – Stuyvesant, both of which are in Brooklyn, which are displayed in the bar plot of Figure 2. To explore further, Figure 3 displays the price range of the top eight most populated neighborhoods, which range of average from $50 to $150 and maximums of $500 and presumably over. To evaluate the relationship between the variables of the dataset, a correlation plot is created, which shows little important correlation between variables, which is good for modeling. Two methods are used to investigate the relationship between price and the other covariate: lasso and regression trees.

**Figure 2: Property Count**

**Figure 1: Price and Location in NYC**

**Figure 3: Top Neighborhood Prices**

The other investigation being completed with this dataset is whether we can predict the borough the property is located using the other covariates of the dataset. Methods used for this analysis include LDA, QDA, classification tree, bagging, random forest and boosting.

Figures:

