**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. Our goal is to analyze the Airbnb housing price, quality and availability in each of New York City’s boroughs and neighborhoods, using the company’s released dataset in September 2017. Our intent is to provide clear visualization and meaningful interpretation of the data for convenience of visitors and all interested in the service of Airbnb.

To complete this analysis, we will be utilizing Airbnb’s dataset from Jeff Goldsmith’s website: <http://jeffgoldsmith.com/DSI/dataset_airbnb.html>

Below are the planned analyses and visualization aids to be provided:

1. Ridge regression or Lasso regression for interpretation between price and predictors
2. PLS and continuous polynomial regression
3. Generalized Additive Models for prediction of score and price with predictors, evaluated with both natural and smoothing splines
4. Graphs showing how MSE changes with each model, optimizing degrees of freedom and showing best line of fit of the overall data for each method
5. Classification tree method (pruned tree model, boosting model, rainforest model, bagging model): Use scores, room type, price and availability as predictors to classify which borough does it belong to.
6. Linear Discriminant Analysis or Quadratic Discriminant Analysis by boroughs
7. K-Nearest Neighbor by boroughs
8. Barplot comparing Price v Neighborhood

Estimated Timeline:

Role Assignments: 04/24/2018

Clean and Tidy data: 04/25/2018

Visualization: 04/27/2018

Analyses: 04/28/2018

Write Up: 05/01/2018 – 05/06/2018