

## IEOR E4650-002 Project Proposal

### Listing Rating in NYC Airbnb

Submitted by: Ji Yin Chen (jc5498), Xinyi Lin (xl3024), Gaole Lyu (gl2704), Yifan Xia (yx2610), and Yifei Zhu (yz3919)

#### Problem statement

Airbnb, a vacation rental online platform, allows both travellers and guests to explore more travelling possibilities and engage to a more personalized and unique experience. How to ensure the customers successfully find the most satisfying accommodation while maintaining a high quality of listing hosts is, therefore, an important issue for the company to focus on. When evaluating a current listing property and host, reviewing score is the most direct way to reflect customers' experience. Therefore, our group intends to analyze listing ratings and figure out what factors would lead to a higher rating for Airbnbs in NYC which could help the company to have a better understanding of customer satisfaction and needs in NYC specifically.

#### Why it is interesting

Review scores rating not only have the power to influence consumer decisions but also can strengthen a company's credibility. Customer interaction ultimately leads to improved profits for businesses. Our program aims to find out which factors may impact the final review score rating of a listing house. Learning what factors that may influence the guests most when choosing an accommodation would help Airbnb to have a clearer idea of customers needs. Therefore, the company and the hosts can use this information to adjust its strategy to attract more customers.

#### What data we will use

The datasets on Airbnb Open Data's web page, i.e., <http://insideairbnb.com/>, draw our attention. Here Airbnb records rental data to keep track of hosts' and renters' information, as well as the general and specific ratings that renters provide, and due to the relative familiarity we unanimously have of New York City and the detailedness of the NYC dataset, we decide to choose the "Detailed Review Data for listings in New York City" as the very dataset to be used. The data can be accessed as "reviews.csv.gz" under the category "New York City, New York, United States", with 05 October, 2020 as "Date Compiled", through the link: <http://insideairbnb.com/get-the-data.html>.

#### Plan of Attack

To tackle this task, we first need some data cleansing to remove columns that are irrelevant to our analysis or unquantifiable as well as rows that are statistically inadequate. Then we will compare correlations between each factor and the response variable to filter out uncorrelated and weakly correlated ones. Lastly we apply a variety of statistical methods such as ridge/lasso regression and cross-validation on the remaining explanatory variables and response variable to build the most accurate review score prediction model.