Project Proposal for Econ 445

Group Members: Madelyn Elise Caufield, Xiaohan Xiao, Xinyue Wu, Yidi Hou

# I. Question

For this project, we are focusing on predicting whether the rental price of Airbnb is undervalued based on the inside data provided by Airbnb. We intend to build a proper predicting model and then compare the current housing price with the prediction output to decide whether the housing price is undervalued. To be more accurate, variables on different aspects such as house structure, neighborhood, and past customers’ review and multiple machine learning methods would be taken into consideration.

# II. Data

The data was collected from Inside Airbnb (<http://insideairbnb.com/get-the-data.html>), a website that reports and visualizes scraped data from Airbnb so that various metrics and characteristics on listings can be obtained in a concise format. The specific data we will be using consists of listings in Los Angeles and has information on pricing, neighborhood, ratings, number of amenities/rooms, and other important variables. The data is all pulled at a single point in time, so we have updated data as of 01/08/22.

# III. Analyzing Method

​We plan to use several machine learning models to predict the airbnb house price in different areas. The NLP aims to process the comments and review of one house; The linear regression, neural network or Random Forest aims to forecast the house price in Airbnb using many potential factors that we gained from the Airbnb website. We will compare different kinds of models and choose the best one with the smallest MSPE.

# IV. Significance

The problem we analyzed is not only an economic problem, but contains many other subjects as well, such as data science, machine learning. The project can help us grasp the knowledge learned from class comprehensively and can help us analyze similar problems in the future.