Group 5 Proposal
Data Mining
10/29/2019

Airbnb is the most demanding online marketplace where people seek and offer lodging as homestay. There are plenty of criteria to list a property with price variation like neighborhood, amenities, bedrooms, bathrooms, pictures etc. For our final project, we will work upon Florence, Toscana, Italy Airbnb open dataset for 2019 which is comprised of necessary listing activities and metrics. Our focus of work will be on predicting important features impacting prices of Airbnb listings. We will work upon the dataset for exploring new dimensions, feature engineering, visualization and making predictions regarding pricing of Airbnb rentals. The dataset needs to be prepared for modeling as there are some missing values. For statistical analysis, we will perform standard regression models like linear regression (to analyze predicted price and actual price), or classification models' like decision trees, random forest, and support vector machine from the Scikit-learn library to compare different models. We may also perform hyper parameter tuning to improve the model performance. The code will be implemented using python libraries in PyCharm. We will analyze the rmse, r-squared score, absolute error and accuracy score for measuring the performance of models. As a group project, we will work in collaboration and have planned to work on the following timetable: data preprocessing completion by November 8, modeling by November 15, GUI and Final Report by November 27. This timeline will ensure that we can complete our work on time

Link to the Data Set:
http://insideairbnb.com/get-the-data.html

Reference links:
https://www.dataquest.io/blog/machine-learning-tutorial/
https://medium.com/datadriveninvestor/making-models-airbnb-price-prediction-data-analysis-15b9af87c9d8
https://towardsdatascience.com/airbnb-rental-listings-dataset-mining-f972ed08ddec
https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74