Summer 2023
Introduction to Data Mining

Topic Proposal

Team 10: Ambar Pathak, Henry Hirsch, Nena Beecham, Robert Williams

Founded in 2007, AirBnB has become one of the world's leading companies for short-term rentals. The company boasts an "excess of six million listings, covering more than 100,000 cities and towns." Among the cities hosting an array of AirBnB listings is Washington, DC. In 2022 alone, hosts in the DC area "made 29% more than the typical host in the U.S."

Despite the relative success of DC owners, hosts around the country have been troubled by declining revenue. Even more worrisome, almost 50% of hosts depend on the extra income to stay in their homes. Given the current threats to host revenue and the essentiality of the short-term rental market to the city of Washington, DC, we developed the following SMART question:

> *What are the most important features for predicting an AirBnB's rating in the city of Washington, DC from 2013 to 2023?*

We intend to use a variety of regression models to predict a listing's rating. The source of our data set is Inside AirBnB, where quarterly data is shared. Through our research, we hope to gauge how a listing's rating is influenced by a number of factors including the structural characteristics of the property, the location of the property, the host's behavior, the ratings of previous renters, and the year in which it was rented. The dataset contains 6542 rows of data and 75 columns including the id column.

Repository Link:
https://github.com/ambar3497/DATS-6103-End-Term-Project

Intended Modeling Methods:
Multiple Linear Regression, Decision Tree Regression, and Nearest Neighbors Regression.