BUAN 6341.002 - APPLIED MACHINE LEARNING - PROJECT PROPOSAL

Harish Sathiyanarayanan – HXS210073

Introduction:
Airbnb is a popular platform for individuals and groups to book short-term accommodations. One of the most important factors for both hosts and guests on the platform is the price of the listing. For hosts, pricing their listings competitively can attract more guests and generate higher revenue. For guests, finding a listing at an affordable price is critical for their travel plans. Furthermore, predicting Airbnb prices can also help policymakers and urban planners understand the effects of short-term rentals on the housing market and the economy. Accurate price predictions can inform decisions around regulating the industry and managing the impacts of short-term rentals on local communities.

Data Description:
The "Airbnb Price Prediction" dataset by STEVEZHENG on Kaggle contains information on Airbnb listings in various cities around USA. The dataset comprises over **70k (74112) observations and 28 independent variables** that can be used to **predict the target variable – Price** using **independent variables** like location, property type, room type, occupancy, amenities, reviews, and more. The independent variables include categorical and numerical variables. Categorical variables include the cleaning fee, host identity verification. Numerical variables include the number of bedrooms and bathrooms, number of reviews, etc. **The source of the data is from publicly available Airbnb listings on their platform which is available in Kaggle**. Overall, this dataset provides a rich set of features for predicting Airbnb listing prices and analyzing the factors that contribute to listing popularity or demand.

Dataset Link:
https://www.kaggle.com/datasets/stevezhenghp/airbnb-price-prediction/versions/1

High Level Overview of the process:

- Data Preprocessing: Clean and preprocess the data to ensure it is ready for analysis. This includes dealing with missing values, removing duplicates, and converting categorical variables to numerical values.

- Feature Engineering: Create new features from the existing data that could be helpful in predicting prices. For example, you might create a feature that represents the distance from a listing to popular tourist attractions or the average review score of similar listings in the same neighborhood.

- Data Splitting: Split the data into training and testing sets to evaluate the accuracy of the model.

- Model Selection: Choose an appropriate machine learning algorithm for the problem at hand. In the case of predicting Airbnb prices, common algorithms include linear regression, decision trees, and random forests.

- Model Training: Train the chosen machine learning algorithm on the training data.

- Model Evaluation: Evaluate the performance of the model on the testing data. This can be done using metrics such as mean absolute error, mean squared error, and root mean squared error.