

CMSC 320 - Final Project

1. GitHub Repository Link

https://github.com/BigMasterC/CMSC320_Final_Project

2. Dataset Choice

We have chosen the following dataset(s):

Dataset Name:

Airbnb Dataset (@Kaggle)

Source/Citation:

Jangra, A. (2024, May 28). Airbnb dataset. Kaggle.
<https://www.kaggle.com/datasets/ashishjangra27/airbnb-dataset>

Brief Description:

This dataset is a comprehensive collection of Airbnb listings that captures key details about said listings. From characteristics and host profiles to pricing, location data, and guest feedback. One major benefit of this dataset is that it's structured to support analytical projects (like our Final Project), predicating rental prices, analyzing customer reviews, and mapping geographic trends in the short-term rental market.

3. Justification for Dataset Choice

We chose this dataset, because we felt it would be a topic that people would actually be interested in reviewing. People can use our analysis of this data to compare prices, locations, and amenities in order to make informed travel decisions.

Another reason we chose this dataset is because it contains almost 13k rows of data. We predict that it'll provide us with a rich source of information with being overwhelmingly large for initial processing and testing on our end. Additionally, in its "Overview" description, the dataset is listed as being specifically good for data analysis and machine learning tasks such as price prediction and sentiment analysis, thereby making it an ideal dataset for our project!