Final Project Proposal

Machine Learning Fall 2021 - Nahee Kim

Research Question:

Airbnb is a fast-growing company for people to book places for stay. Airbnb competes with traditional hotels, and now they have more than 20% market share in the vacation rental industry. As the company's market share grows, the demand to become a host for Airbnb also increases.

So, how can we advise future NYC Airbnb hosts to select the price range and location that's most profitable?

Summary:

This project is for people who want to start Airbnb hosting business in New York City. Using the 'New York City Airbnb open dataset' from kaggle.com, we will create a machine learning model that advises future Airbnb hosts that what's the appropriate price range in which area to be selected by long-term stay users to minimize vacancy of the place. We don't have a dataset of the price of properties compared to a price per night. So there is a limitation of how much the host would make, but understanding the most selected price range depending on location would be beneficial. Especially reducing the chance of long vacancy of the property would give an idea to set a price range depending on the location of the New york city.

Problem:

Hypothetically, I want to start an Airbnb host business in New York City and don't have any background in the vacation rental market industry. Therefore, I want to find out the most profitable price range and neighborhoods to invest in. So I can maximize the booking stay nights of the place that I am hosting. Given its data, we can't provide how much the host would profit by the place, but we can tell which price range per neighborhood would be most preferred by users, especially long-term stay users. This machine learning model would help the host make the decision on how much to invest in the property, depending on the neighborhood.

Possible Solutions:

The focal point of this data is to find the relationship between price, number of stays, and neighborhood. The first thing I will do to create a machine learning model explores the data and filter the unnecessary data. And I am also going to process the data for a regression algorithm or decision tree algorithm depending on the data structure. And create a graph that shows you what's the appropriate price range is in different neighborhoods.

Reference:

https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data