MSAI 349 - Machine Learning
Group 2 Members - Rhett D'Souza, Albert Guo, Lukas Justen, Keith Pallo

Airbnb - Efficient Markets Case Study

Our group will try and address the task of understanding Airbnb market dynamics. There are several different ways to accomplish this goal, but more specifically, we will try and predict the best price for any Airbnb given standard measures such as the location of the listing, the date of pricing, and the features that particular Airbnb offers. This can take several different forms as our project unfolds. Our most promising idea is to try to recommend a price for an individual looking to put a listing on Airbnb for the first time. However, we know that reviews play a strong part of customer perception, so we would also like to explore if any given Airbnb is undervalued or overvalued given their historical performance.

This is an interesting project as it deals with many elements of behavioral economics, game theory, and tech industry disruption. As the Airbnb concept increasingly becomes a reality for the hotel industry, individual Airbnb listings and competition are increasing. Hence, it can be very difficult to know, from a host's point of view, if each person is maximizing their revenue potential or if they are cannibalizing their own performance by listing their prices too high. Also, it can be very unclear if there may be a substantial benefit for the host to invest in an additional amenity, such as having more bedrooms available or something as simple as providing toiletries. This is something very important for Airbnb as a firm to understand, since their success relies heavily on creating a marketplace that is simple and easy to navigate for current and new hosts. The inspiration for this project came from one of our team members staying at a Brooklyn Airbnb during the summer. The hosts asked the team member what they thought about the price and description of the listing - and eventually, the conversation shifted towards the difficulty of creating the posting, and how it was one of the biggest deterrents from actually setting up shop.

Fortunately, there are already several publicly available datasets from both Kaggle and data.world. We will plan to first focus on the Seattle and Boston datasets from Kaggle as they are fairly comprehensive in the features/attributes provided. However, we would also like to explore potentially enriching each of these datasets, either by performing transforms on the features already available, generating new features from the given data points (for example deep learning based sentiment analysis and/or keyword extraction from raw text data) or adding new data points entirely from outside sources such as hotel locations in the area. The attributes that are most available for additional analysis are reviews and images of the property.

As previously mentioned, the features and attributes we will use for this task will be the standard information provided by hosts at each posting. These include things such as number of guests that can be accommodated, the experiences available, and the amenities available for that location. At first, we would like to exclude host variables such as minimum night stay, but we can add these in if necessary.

Our initial approach will be to cluster Airbnb listings based on similar attributes and offerings provided to consumers within a certain area, and use regression to predict listing price. Essentially, we would like to group "like" Airbnb's together in terms their raw offerings to determine price. We would also like to examine the price discrepancies that exist within clusters, as well as the success differences between them. However, as mentioned before, we understand that review's can play a huge role, so we would also like to cluster based on this attribute if sample size allows. As our project progresses, we plan to dig into the landscape of the market, and learn how these standard attributes affect performance.