Springboard Capstone Project 1

NYC Airbnb Average Listing Price Prediction

Problem to Solve

Serving as a marketplace for house sharing, Airbnb provides a platform for hosts and renters to make a connection. Hosts with spare properties from a sofa bed in the living room to a beach house located along sea shore now are able to rent out all these spaces via Airbnb to travelers looking for a place where they would like to spend their nights at.

Moreover, what does it make people to choose to stay at a stranger's property over at a hotel room? The reason behind could be really different to every individuals. For example, people who are into exploring a new city might see staying at a local's place as an priceless opportunity to indulge in the local culture and attain exotic experience. Or, some might simply not be able to afford staying at a hotel.

For the economical perspective, staying at an Airbnb might cost you less money due to less complimentary services and amenities that hotels offer. However, how to determine if the listing price of a property is reasonable compared to others similar properties in the area becomes a question that potential renters can think about to see whether the price is truly worth it while doing Airbnb research.

Given variables like the number of rooms / bathrooms, locations / neighborhoods and historical listing prices, I want to build out an average pricing prediction model to help renters to gauge whether they are getting a good deal or they are actually overcharged by the hosts?

Client to Serve

Renters, looking for a local property as their short term stay destination, will be beneficial to the result of the prediction model. By comparing the current listing price online with the outcome, which is a price predicted by giving variables like the number of rooms / bathrooms, locations / neighborhoods and historical listing prices, renters will be able to have a sense of historical average price of similar property in the area and to take the predicted price as a gauge to see whether the hosts are over-charging or not.

Data to Use

For this project, I will acquire and use listing level data from Inside Airbnb for New York City by running through all available zip files that the site extracted and collected from Airbnb site for public use.

Method to Approach

This analysis will be a supervised regression problem by using independent variables related to each listing such as the number of bedrooms, the number of people being able to be accommodated and the location of the property, etc. to predict an average price for renters as a reference.

In terms of methods that will be implemented to this analysis, I am currently unsure which method could be the best fit. But I am thinking that method such as K-Nearest Neighbor (Knn) may be a good starting point for this prediction project.

Output to Deliver

The final deliverable for this project will be packaged as followed:

- Full Python code available on GitHub
- Delivering final prediction model along with EDA findings in PowerPoint