

Springboard--Data Science Career Track
Airbnb price in New York City
Capstone Project 1 Proposal
By Kin Sun
January 2020

- **What is the problem you want to solve?**

With so many Airbnb available in NYC, some people may wonder if they are charging too little or too much for the places they rent. The 40,000-samples dataset I am planning to use--which I will obtain from Kaggle¹-- contains some features that can be used to predict the price per night of a listing.

- **Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?**

Prospective clients are any individuals, property management companies, existing Airbnb host, etc—in other words, anyone who just wants to use their properties for rental purposes on Airbnb. These clients would be able to use the results of this project to find out how much they can reasonably charge per night of stay.

- **What data are you using? How will you acquire the data?**

For this dataset, I am looking into Airbnb price for properties listed in the New York City area. This dataset will be acquired from Kaggle as I mentioned above.

- **Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.**

My current vision is that I will build multiple regression models and compare their performance to determine what would be the most appropriate models that could be used to predict the price of a listing as a function of features to be identified.

There are 5-6 features that I anticipate I will use from the dataset, such as neighborhood/boroughs, room type, minimum nights (stay), number of reviews, reviews per month and availability in a year. Here is a brief description of how I think these features might be helpful to build regression models:

1. Neighborhood/Borough – This is a categorical variable and the location of the property.

¹ <https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

2. Room type – This is also a categorical variable. A whole entire house/apartment, a private room and shared room.
3. Minimum nights – This is the minimum night a guest need to stay/pay.
4. Number of reviews – This is the number of reviews of this property
5. Reviews per month – This one is similar with “Number of reviews.” However, longer tenure hosts tend to have higher number of “Number of reviews.”
6. Availability in a year – Out of 365 days, this shows how many days this property is up for listing

- **What are your deliverables? Typically, this includes code, a paper, or a slide deck.**

For deliverables, as required, I will post the following in my GitHub repository: all Jupyter notebooks I will develop, a final report, and a presentation slide deck.