

Project Milestone 1

I'm a man who likes travelling, so the dataset that I'm interested in is New York City Airbnb open data from Kaggle, which is also abbreviated as "ABNYC2019". Several reasons make it attractive to me. First, this data is from Airbnb, a source which is popular and recognized by the whole world. Not only Airbnb, but also many other companies are trying to dig in and discover the secret hidden in the data. In other words, it is much easier to find up-to-date information and technique to enable me to evaluate problems in diverse dimensions. Second, the patterns or relationships from the data analysis may help me to better understand business decisions, figure out customers' and hosts' potential behavior and predict the trend of hotel markets in the future. Finally, investigating the data could give guidance to my choices of airbnb from this point forward. The choices will be made more wisely and data-based.

Let's go into the data in detail. Each column is an attribute of a room and each row represents the record of it. This data has several attributes, some of them are about hosts, like hosts name and id. Furthermore, the basic geography and price information about rooms. For example, latitude, longitude, neighbourhood(names of blocks) and price per night. Also, the behaviour of travelers, such as minimum nights stayed and number of reviews. Plentiful attributes imply the numerous ways of visualization during the analysis. For instance, a bar chart can be created with top N attractive rooms. When it comes to a more complex one, a map can be adapted to a heat map. First of all, we can find a New York City map that ranges from min(latitude,longitude) and max(latitude,longitude), then put all the points on the map by using their geography attributes. Next, different kinds of colors can be implemented to show the price of listings.

To sum up, my goal and task is to find relations between the features by using this dataset and obtain potential trends and predictions. Which kind of rooms is the most popular? Which area in NYC is possible to be more profitable? What makes travelers stay more nights? These are all the questions that I'd like to answer by visualization.

