

Airbnb has become one of the most popular methods to earn extra money, since anyone who has some spare room can become a host. However, it is not clear how best to invest one's money in Airbnb: do you want to focus on having good amenities for your guests? or having frequent communication and fast response? or is location the most important factor? The goal of this project is to analyze the importance of all of these factors using the InsideAirbnb dataset.

The data comes from a public website called InsideAirbnb, which compiles a multitude of data on every Airbnb listing in 16 major cities in the US (totaling around 1.7GB). The data includes vital statistics on the listings like location, description, size, price, amenities, and reviews. Another factor which I suspect might be important is proximity to public transportation. So, I chose to supplement the InsideAirbnb data by querying the Google Maps API to get walking directions/durations between every listing in Boston and the nearest subway station.

How do we determine which Airbnb listings are the most successful? I used the number of reviews per month as a proxy for number of occupants, that way I could identify the most successful listings in each city. After that, I can apply Machine Learning techniques to determine what makes those successful listings so successful. One approach I have tried is to use Natural Language Processing (NLP) to analyze the text of each listing's description and reviews from guests, so I can examine the intrinsic differences between the texts associated with the most popular listings compared to all the other listings. Furthermore, I can apply modeling techniques (like regression or random forest) to use the information available for each listing as a feature (location, price, size, amenities ...) to predict the popularity of that listing. After training my model, I can assess the feature importance to determine where Airbnb hosts should invest their money/time.

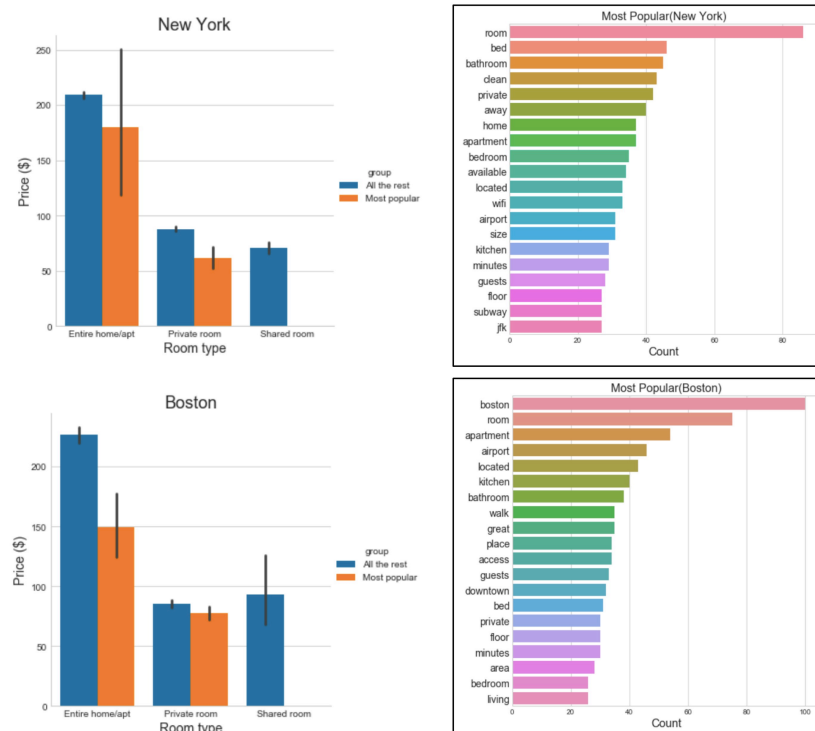


Figure 1.

After using NLP to analyze the most common words in the description of the 50 most popular listings in New York and Boston, some interesting results can be found (see Figure 1, right):

- “Airport” is one of the most common terms in these popular listings. Interestingly, when I repeat this analysis on the rest of the listings, “airport” is not one of the most common terms anymore.
- In New York, “wifi”, “subway”, and “private” ranks very high on the most popular listings, which indicates that guests in the city care about convenience and privacy.

Also, the left side of Figure 1 shows a comparison of the average price of Airbnb listings in New York and Boston, separated into categories based on type of apartments. The two colored bars show prices of the most popular listings in orange and all the other listings in blue. This Figure highlights the fact that the most popular listings in Boston are cheaper than all the rest, while there is no significant difference in price range between the two categories in New York.

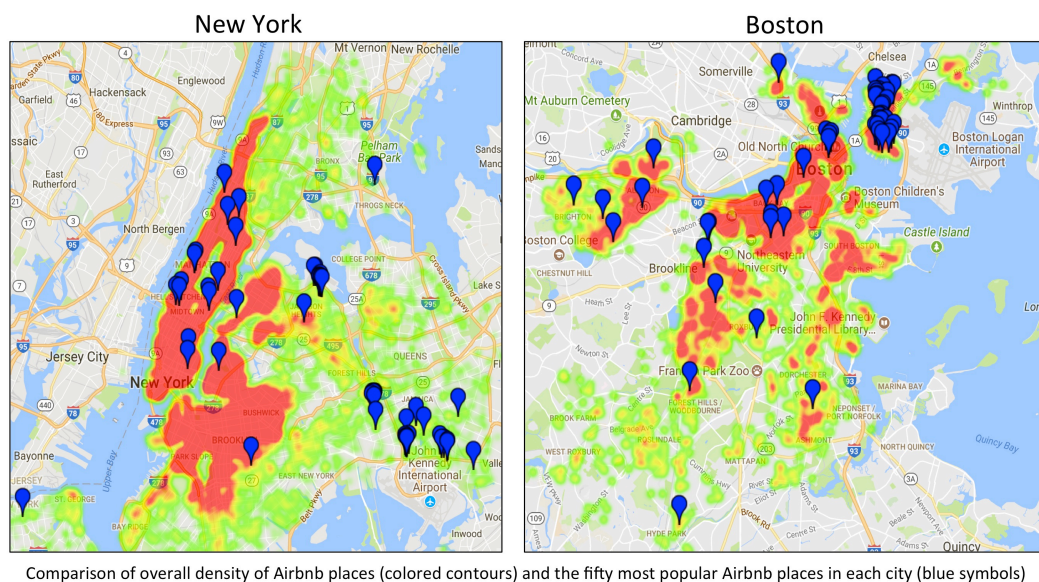


Figure 2.

The importance of location can be directly examined by making a spatial plot of the most popular Airbnb listings in a city, which is shown for New York and Boston in Figure 2. Recall that “airport” is one of the most common words in a description for a popular listing in these two cities, therefore it is not too surprising that I find a significant cluster of the most popular listings centered around the airport (JFK around New York and Logan around Boston).

This project will provide valuable insights for anyone who wishes to maximize their occupancy as an Airbnb host in one of the 16 cities included in the dataset. The completed analysis will rank the relative importance of the following factors: location of apartments, description of apartments, furniture/amenities, communication with guests, and more.