**Renting an apartment in New York City**

1. **Introduction**

**1.1 Background**

New York City is one of the most populated cities in the world. The population is 8,622,698, as of July 1, 2017. John is looking to rent an apartment in New York City for at least 1 year. New York City is one of the most popular places to live in the country. Considering this context John is asking our help in finding a good neighborhood to live in.

**1.2 Problem**

In order for John to find the correct neighborhood to buy an apartment, we will need to use the information about the localities and the different neighborhoods of New York City. This is based on a number of factors such as proximity to a number of cultural attractions and a number of restaurants and nightlife.

1. **Data**

The data used in the study is from four main parts:

1. List of neighborhoods in New York City.
2. List of boroughs in New York City
3. Latitude and longitude of each of the neighborhoods in New York City.
4. Once having the database of the neighborhoods in New York City, *Foursquare* API is used to explore the nearby venues.

The neighborhoods of New York City is scraped from the Cogitive Class which is from <https://cocl.us/new_york_dataset>. The data is then leveraged in order to determine which borough and neighborhood is the most appropriate for John to rent an apartment.

1. **Methodology**

The data collected for this study was scraped from <https://cocl.us/new_york_dataset>. This is from Cognitive Class. *Pandas* package was used to read in the data file and save them to the dataframe. The dataframe is called **NYC\_neighborhoods**. This contains the columns “Borough”, Neighborhood”, “Latitude” and “Longitude”. There are 306 neighborhoods. Then using the *Folium* package, I plotted all the neighborhoods from this dataframe in the New York City map as shown in Figure 1.

**![A picture containing text, map

Description automatically generated]()**

Figure 1. Map of New York City neighborhoods. There are 306 neighborhoods in total.

I then collected the Venue data for the New York City neighborhoods using *Foursquare* API. I named the dataframe **NYC\_venues**. I checked to see which borough has the most venues creating a dataframe called **number\_NYC\_venues.** I checked to see which borough has the most restaurants, cocktail bars, and museums creating a dataframes called **number\_restaurants\_borough**, **number\_cocktail\_bars\_borough**, and **number\_museums\_borough** respectively**.** Since Manhattan has the most venues overall and the most restaurants, cocktail bars and museums, I have decided to narrow the neighborhood search to the borough of Manhattan.

The dataframe used for the Manhattan neighborhoods is called **manhattan\_neighborhoods.** This contains the columns “Borough”, Neighborhood”, “Latitude” and “Longitude”. There are 40 neighborhoods. Then using the *Folium* package, I plotted all the neighborhoods from this dataframe in the New York City map as shown in Figure 2.

![A picture containing text, map

Description automatically generated]()

Figure 2. Map of Manhattan neighborhoods. There are 40 neighborhoods in total.

I then collected the Venue data for the Manhattan neighborhoods using *Foursquare* API. I named the dataframe **manhattan\_venues**. I checked to see which borough has the most venues creating a dataframe called **number\_manhattan\_venues.** I checked to see which borough has the most restaurants, cocktail bars, and museums creating a dataframes called **number\_restaurants\_borough**, **number\_cocktail\_bars\_borough**, and **number\_museums\_borough** respectively.

To explore the similarities and dissimilarities between neighborhoods in the borough of Manhattan, k-means clustering is used. K-means clustering aims to partition *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. In our case, we have 40 neighborhoods to cluster and a cluster number of 6 was used in the analysis. *scikit-learn* package is used to perform the machine learning tasks.

1. **Results**

We see that Manhattan has the most venues with 3,324 venues as shown in Figure 3.

![A screenshot of a cell phone

Description automatically generated]()

Figure 3. Number of venues for each borough

We also see that Manhattan has the most restaurants with 932 as shown in Figure 4.

![A screenshot of a cell phone

Description automatically generated]()

Figure 4. Number of restaurants for each borough

In addition, we see that Manhattan has the most cocktail bars with 57 as shown in Figure 5.

**![A screenshot of a cell phone

Description automatically generated]()**

Figure 5. Number of cocktail bars for each borough

In addition, we see that Manhattan has the most museums with 16 as shown in Figure 6.

**![A screenshot of a cell phone

Description automatically generated]()**

Figure 6. Number of museums for each borough

We check to see how many venues there are in each neighborhood in Manhattan. Therefore, it is difficult to tell which has the most venues as shown in Figures 7a and 7b.

![A screenshot of a cell phone

Description automatically generated]()

Figure 7a. Number of venues in Manhattan

![A screenshot of a cell phone

Description automatically generated]()

Figure 7b. Number of venues in Manhattan (continued)

We check to see how many restaurants there are in each neighborhood in Manhattan as shown in Figures 8a and 8b. We see that Chinatown has 43, followed by Greenwich Village with 42 followed by Turtle Bay with 39.

![A screenshot of a cell phone

Description automatically generated]()

Figure 8a. Number of restaurants in Manhattan

**![A screenshot of a cell phone

Description automatically generated]()**

Figure 8b. Number of restaurants in Manhattan (continued)

We check to see how many cocktail bars there are in each neighborhood in Manhattan as shown in Figure 9. We see that Noho has the most with 5.

![A screenshot of a cell phone

Description automatically generated]()

Figure 9. Number of cocktail bars in Manhattan

We now check to see how many museums there are in each neighborhood in Manhattan as shown in Figure 10. We see that Carnegie Hill, Soho and the Upper East Side are tied with having only 2 museums.

**![A screenshot of a cell phone

Description automatically generated]()**

Figure 10. Number of museums in Manhattan

We see that the most common venues in the top 5 for each neighborhood in Manhattan are food places shown for each of the Clusters.

Table 1. Sample Neighborhoods in Cluster 0 for top 5 common venues

**![A screenshot of a cell phone

Description automatically generated]()**

Table 2. Sample Neighborhoods in Cluster 1 for top 5 common venues

**![A screenshot of a cell phone

Description automatically generated]()**

Table 3. Sample Neighborhoods in Cluster 2 for top 5 common venues

**![A screenshot of a cell phone

Description automatically generated]()**

Table 4. Sample Neighborhoods in Cluster 3 for top 5 common venues

**![A screenshot of a cell phone

Description automatically generated]()**

Table 5. Sample Neighborhoods in Cluster 4 for top 5 common venues

![A screenshot of a cell phone

Description automatically generated]()

Table 6. Sample Neighborhoods in Cluster 5 for top 5 common venues

**![A screenshot of a cell phone

Description automatically generated]()**

1. **Discussion**

New York City is a large city with 306 neighborhoods. This is difficult to limit the number of neighborhoods that are available. When I ran the clusters I used the top 10 venues. However there is not enough room on the page of this report to fit the entire table. Therefore, I only used the top 5 venues in this report. If you want to see the top 10 venues you can check it out in my notebook at [https://github.com/pbrot0827/Coursera\_Capstone/blob/master/Battle%20of%20the%20Neighborhoods%20(Week%202).ipynb](https://github.com/pbrot0827/Coursera_Capstone/blob/master/Battle%20of%20the%20Neighborhoods%20%28Week%202%29.ipynb).

Due to the limitation of *Foursquare* API for each neighborhood we can only get a maximum of 100 venues nearby. Therefore, when determining the number of venues in each neighborhood, it is difficult to tell which has the most because it may not reflect the accurate number of venues in each of the neighborhoods for those that are at least 100. When searching for the restaurants for each borough and neighborhood, this only includes the venues with the word “restaurant” and does not include fast food places, diners, coffee shops, ice cream places nor sandwich shops. I used in searching for the venues using “Cocktail Bars” rather than “Bars” because using “Bars” can also include “Juice Bars” which are not nightlife places. Please keep in mind that when searching for museums this does not include art galleries nor exhibits which some people might consider to be museums.

Based on my research and data analysis, I would recommend that John rent an apartment on the Upper West Side neighborhood in the borough of Manhattan. Since in the top 5 venues includes restaurants and bars in that neighborhood.

1. **Conclusion**

Based on the results, we can see most of the neighborhoods in the borough of Manhattan is where food places are very popular. Bars are also very popular in these neighborhoods as well. By devoting more time to this, there is a way to improve the analysis and that is if there is no limitation to from *Foursquare* API, the clustering of neighborhoods could be improved. Ways to collect more data from *Foursquare* would need to be studied.