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.

2. 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.

3. 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.

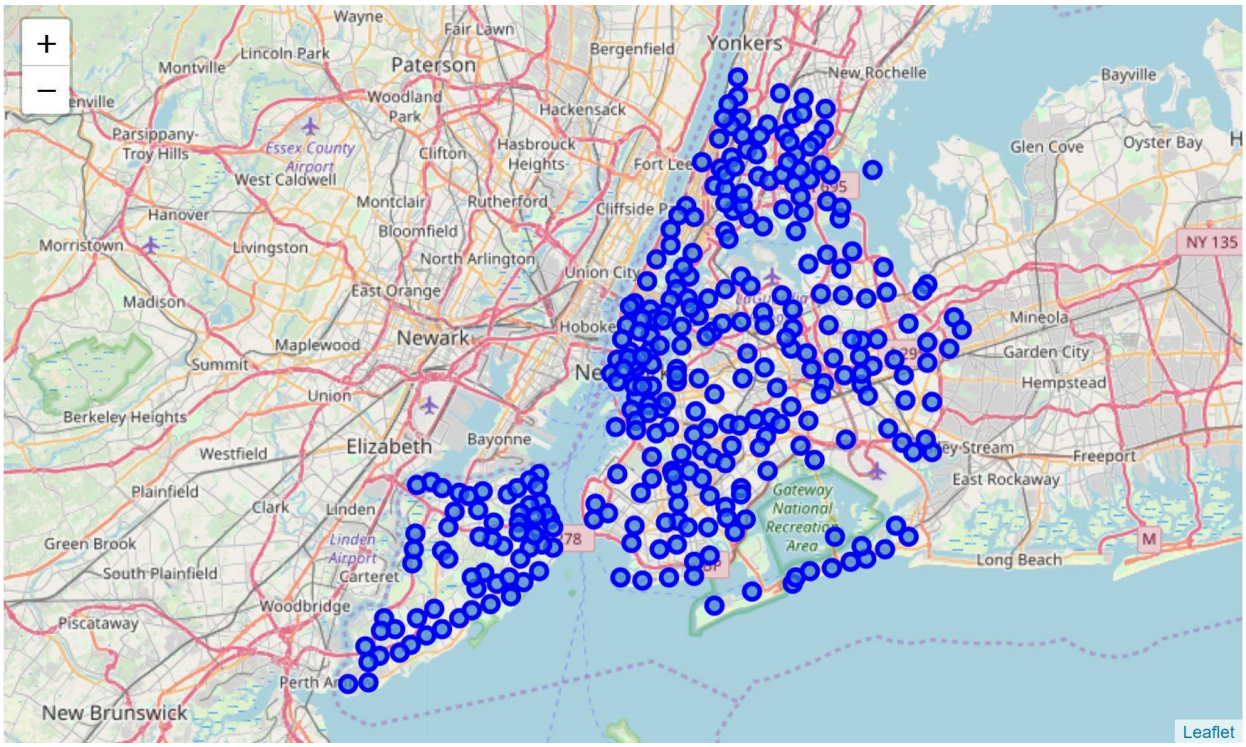

Figure 1. Map of New York City neighborhoods. There are 306 neighborhoods in total.

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.

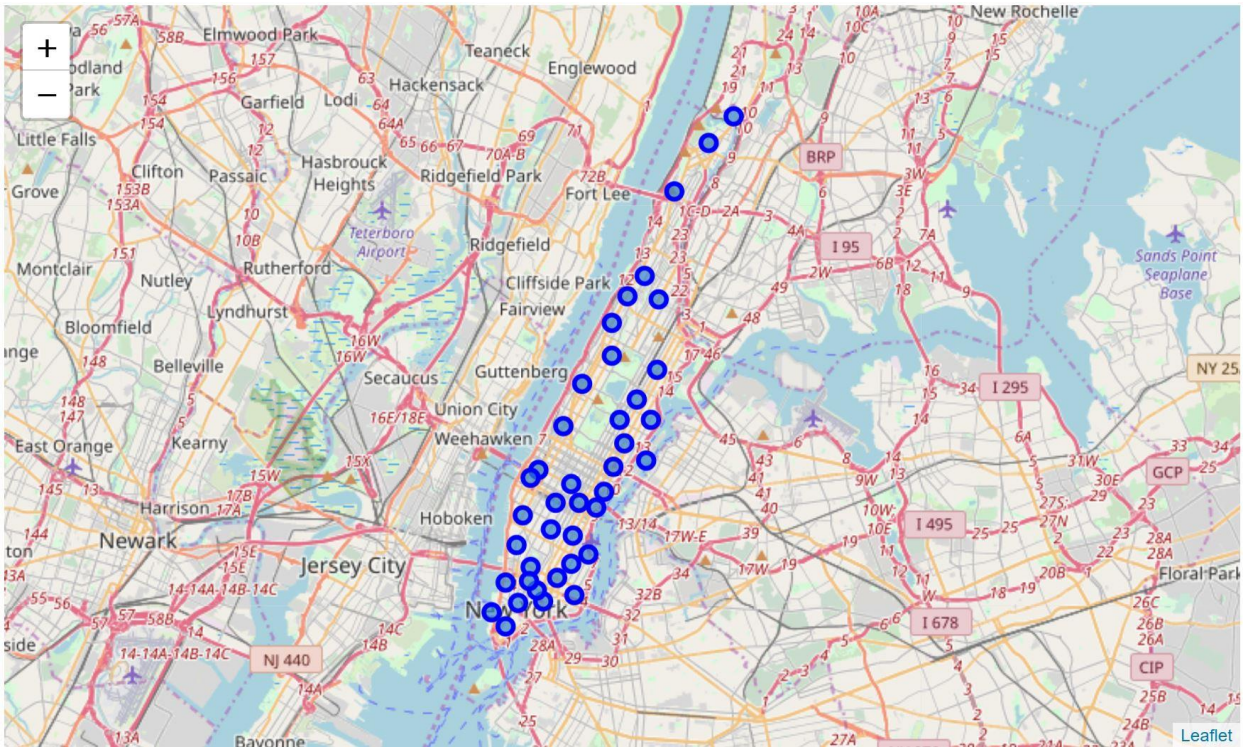

Figure 2. Map of Manhattan neighborhoods. There are 40 neighborhoods in total.

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.

4. Results

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

	Number of Venues
Borough	
Manhattan	3324
Brooklyn	2812
Queens	2129
Bronx	1243
Staten Island	798

Figure 3. Number of venues for each borough

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

	Number of Restaurants
Borough	
Manhattan	932
Brooklyn	661
Queens	582
Bronx	248
Staten Island	161

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.

	Number of Cocktail Bars
Borough	
Manhattan	57
Brooklyn	38
Queens	4
Staten Island	1

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.

	Number of Museums
Borough	
Manhattan	16
Brooklyn	9
Staten Island	4
Bronx	3
Queens	3

Figure 6. Number of museums for each borough

We check to see how many venues there are in each neighborhood in Manhattan. 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, you can see this in Figures 7a and 7b.

	Number of Venues
Neighborhood	
Battery Park City	100
Lenox Hill	100
West Village	100
Upper West Side	100
Upper East Side	100
Turtle Bay	100
Tribeca	100
Sutton Place	100
Soho	100
Noho	100
Murray Hill	100
Midtown South	100
Midtown	100
Carnegie Hill	100
Little Italy	100
Lincoln Square	100
Yorkville	100
Clinton	100
Chelsea	100
Chinatown	100
Greenwich Village	100
Gramercy	100
Civic Center	100
Flatiron	100

Figure 7a. Number of venues in Manhattan

Financial District	100
East Village	100
Washington Heights	86
Tudor City	81
Hudson Yards	75
Inwood	60
Lower East Side	60
Manhattan Valley	59
Hamilton Heights	59
Central Harlem	47
East Harlem	45
Manhattanville	44
Morningside Heights	38
Roosevelt Island	26
Marble Hill	25
Stuyvesant Town	19

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.

	Number of Restaurants
Neighborhood	
Chinatown	43
Greenwich Village	42
Turtle Bay	39
East Village	38
Upper West Side	36
Murray Hill	35
West Village	34
Noho	31
Midtown South	31
Tudor City	30
Sutton Place	30
Yorkville	28
Lenox Hill	28
Flatiron	28
Little Italy	27
Gramercy	26
Chelsea	25
Tribeca	25
Carnegie Hill	24
Manhattan Valley	24
Civic Center	24
Upper East Side	23
Financial District	23

Figure 8a. Number of restaurants in Manhattan

Financial District	23
Midtown	22
Washington Heights	22
Soho	21
Clinton	21
Hamilton Heights	19
Hudson Yards	17
Manhattanville	17
Lower East Side	16
Central Harlem	16
Inwood	16
Lincoln Square	15
East Harlem	13
Battery Park City	9
Morningside Heights	8
Roosevelt Island	3
Marble Hill	2
Stuyvesant Town	1

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.

	Number of Cocktail Bars
Neighborhood	
Noho	5
Chinatown	4
East Village	4
Gramercy	4
Midtown	4
Civic Center	3
Upper East Side	3
Midtown South	3
Lenox Hill	2
Little Italy	2
Tribeca	2
Murray Hill	2
Lower East Side	2
West Village	2
Hamilton Heights	2
Greenwich Village	2
Financial District	2
Clinton	2
Chelsea	1
Hudson Yards	1
Stuyvesant Town	1
East Harlem	1
Turtle Bay	1
Washington Heights	1
Carnegie Hill	1

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.

	Number of Museums
Neighborhood	
Carnegie Hill	2
Soho	2
Upper East Side	2
Chinatown	1
Civic Center	1
Financial District	1
Inwood	1
Little Italy	1
Manhattanville	1
Murray Hill	1
Sutton Place	1
Turtle Bay	1
Upper West Side	1

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
8	Upper East Side	Exhibit	Italian Restaurant	Coffee Shop	Gym / Fitness Center	Juice Bar
9	Yorkville	Italian Restaurant	Bar	Coffee Shop	Gym	Sushi Restaurant
10	Lenox Hill	Sushi Restaurant	Italian Restaurant	Coffee Shop	Pizza Place	Sporting Goods Shop
12	Upper West Side	Italian Restaurant	Wine Bar	Bar	Coffee Shop	Mediterranean Restaurant
13	Lincoln Square	Gym / Fitness Center	Theater	Café	Concert Hall	Plaza
14	Clinton	Theater	Gym / Fitness Center	Italian Restaurant	American Restaurant	Hotel
15	Midtown	Hotel	Coffee Shop	Cocktail Bar	Clothing Store	Theater
16	Murray Hill	Hotel	Coffee Shop	Sandwich Place	Japanese Restaurant	Gym
17	Chelsea	Coffee Shop	Italian Restaurant	Ice Cream Shop	Bakery	Nightclub

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
7	East Harlem	Mexican Restaurant	Bakery	Deli / Bodega	Spa	Thai Restaurant
26	Morningside Heights	Park	American Restaurant	Coffee Shop	Bookstore	Burger Joint
36	Tudor City	Park	Mexican Restaurant	Café	Greek Restaurant	Asian Restaurant

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
37	Stuyvesant Town	Park	Bar	Playground	Boat or Ferry	Pet Service

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Chinatown	Chinese Restaurant	American Restaurant	Cocktail Bar	Dumpling Restaurant	Bubble Tea Shop
19	East Village	Bar	Wine Bar	Mexican Restaurant	Ice Cream Shop	Chinese Restaurant

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Marble Hill	Discount Store	Sandwich Place	Coffee Shop	Yoga Studio	Pizza Place
2	Washington Heights	Café	Mobile Phone Shop	Bakery	Mexican Restaurant	Pizza Place
3	Inwood	Mexican Restaurant	Café	Lounge	Bakery	Pizza Place
4	Hamilton Heights	Pizza Place	Café	Mexican Restaurant	Coffee Shop	Yoga Studio
5	Manhattanville	Chinese Restaurant	Coffee Shop	Park	Mexican Restaurant	Seafood Restaurant
6	Central Harlem	African Restaurant	Chinese Restaurant	Public Art	Fried Chicken Joint	Seafood Restaurant
20	Lower East Side	Coffee Shop	Ramen Restaurant	Café	Pizza Place	Cocktail Bar
25	Manhattan Valley	Indian Restaurant	Coffee Shop	Pizza Place	Yoga Studio	Mexican Restaurant
30	Carnegie Hill	Coffee Shop	Pizza Place	Café	Yoga Studio	Bookstore

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

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
11	Roosevelt Island	Park	Coffee Shop	Sandwich Place	Playground	Bus Stop

5. 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.

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.

6. 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.