APPLY DATA ANALYTICS TO FIND THE BEST LOCATION
TO OPEN A COFFEE SHOP IN MANHATTAN

~ Hung Quach ~

May 28, 2020

1. Introduction

1.1. Background

Manhattan is the most densely populated of the five boroughs of New York City[1]. It has been described as the cultural, financial, media, and entertainment capital of the world. The borough hosts a number of the most critical organizations of the globe such as the United Nations and Wall Street in the Financial District. New York City is the **most economically powerful city** and the leading financial center of the world and Manhattan is home to the world's two largest stock exchanges NYSE and NASDAQ. Manhattan is unique and always be a dream place to open a business.

The population of Manhattan including local residents, commuting workers, and visitors is approximately 4 million[2]. There is a huge **demand for coffee** in the borough. Currently, there are over one hundred major coffee shops and hundreds of small ones within 23 square miles of Manhattan. That means the market has been relatively **competitive**.

1.2. Problem

Our investors would like to find the **best location** in Manhattan to **open a coffee shop** in the borough. Traditionally, entrepreneurs use common sense with simple data points to decide where to open a coffee shop. This time, we will use **data analytics** to try to **find an optimal location** to open a coffee shop in Manhattan, New York. Since there are lots of coffee shops in Manhattan, we will try to detect locations that are not already crowded with coffee shops and relatively no appearance of Starbucks in the vicinity. We would also prefer locations as close to the city center as possible.

1.3. Interest

By using data analytics, we can have a **brand new** and **logical view** about where to open a coffee shop. It will help us understand the big picture of the coffee business in Manhattan without getting lost in the maze of thousands of coffee shops in Manhattan.

---

[1] Wikipedia: https://en.wikipedia.org/wiki/Manhattan
[2] The Dynamic Population of Manhattan: https://wagner.nyu.edu/files/rudincenter/dynamic_pop_manhattan.pdf

2. Data

2.1. Data sources

We realized the factors that will influence our decision are:

- The number and the allocation of existing coffee shops in the neighborhood (any brands)

- The distance of each neighborhood to the center of Manhattan, which we decide is the Times Square

- The number and position of Starbucks coffee in the neighborhood, if any

Based on the data needed, we searched and collected data from different data sources. First, it is the **New York venue dataset** provided by Coursera in the Segmenting and Clustering Neighborhoods in New York assignment. Second, we use the **Foursquare API** to find the list and locations of coffee shops in every neighborhood. Third, we obtain coordinates of different locations including Times Square and Manhattan from **Google Maps API geocoding**.

2.2. Data Cleansing

Data downloaded or scraped from multiple sources were combined into tables. There were a lot of problems with the dataset. First, we have many data tables needed to be combined into one. For example, we have a data table from the New York venue dataset and a data table from Foursquare API for each neighborhood.

Second, the amount of data is large. We must filter the most impactful data to analyze and drop the rest. In this project, we choose the venue, neighborhood, neighborhood latitude, neighborhood longitude, venue latitude, and venue longitude of the coffee shop category data in the venue.

Third, there are many neighborhoods in Manhattan, so it is not easy to manage all the data. 37 neighborhoods will need to have coordinates defined and distance to Times Square calculated.

3. Methodology

3.1. Position each neighborhood in Manhattan

I used the New York venue dataset provided by Coursera to obtain the latitude and longitude of all 306 neighborhoods in New York.

```
neighborhoods.head(10)
```

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585
5	Bronx	Kingsbridge	40.881687	-73.902818
6	Manhattan	Marble Hill	40.876551	-73.910660
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446

Then I **filtered out 40 neighborhoods in Manhattan** because we are focusing on the Manhattan coffee market only.

```
manhattan_data = neighborhoods[neighborhoods['Borough'] == 'Manhattan'].reset_index(drop=True)
manhattan_data.head(10)
```

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688
5	Manhattan	Manhattanville	40.816934	-73.957385
6	Manhattan	Central Harlem	40.815976	-73.943211
7	Manhattan	East Harlem	40.792249	-73.944182
8	Manhattan	Upper East Side	40.775639	-73.960508
9	Manhattan	Yorkville	40.775930	-73.947118

We **visualize** those 40 neighborhoods in a map to see **how each neighborhood is allocated** across Manhattan.
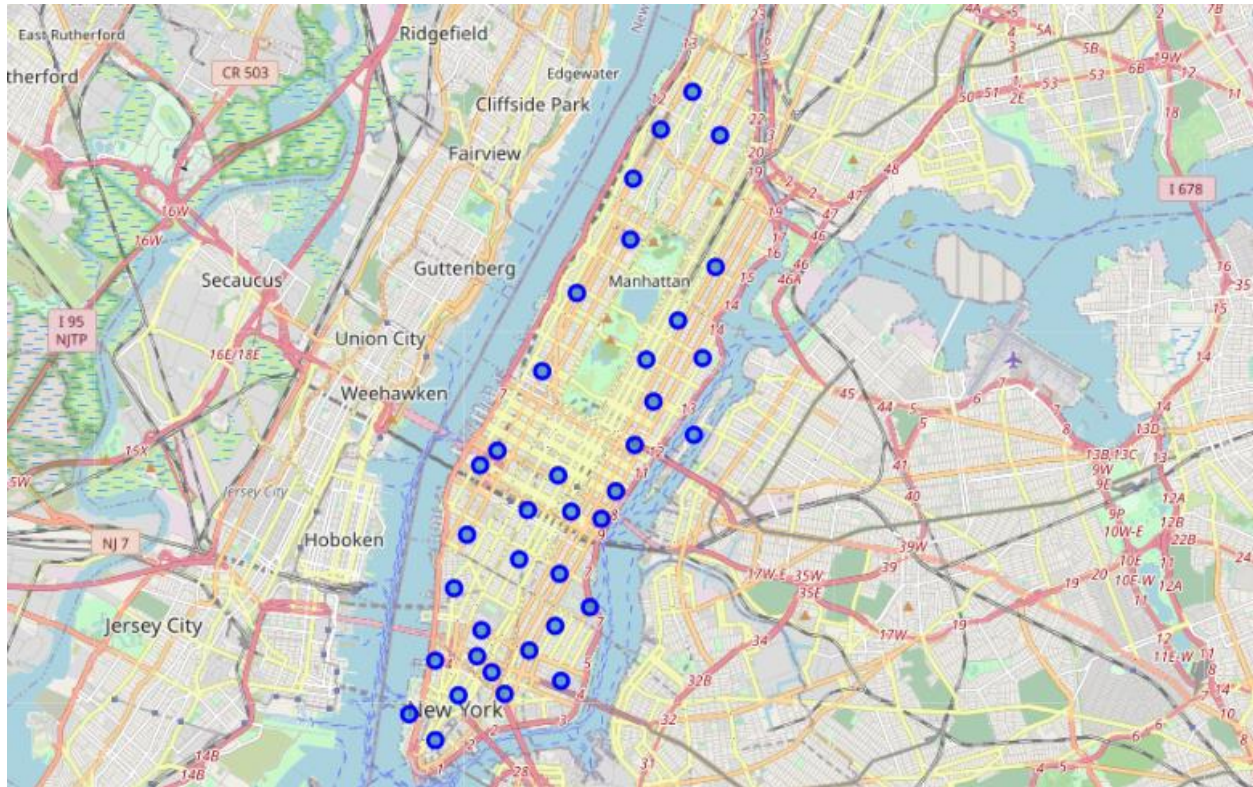
3.2. Position existing major coffee shops in Manhattan

We used Foursquare API to get a list of venues in each neighborhood with a maximum of one thousand venues per neighborhood to make sure we obtain all the major coffee shops. Then we had 3093 venues in Manhattan.

```
print(manhattan_venues.shape)
manhattan_venues.head(10)
```

(3093, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
5	Marble Hill	40.876551	-73.91066	Rite Aid	40.875467	-73.908906	Pharmacy
6	Marble Hill	40.876551	-73.91066	TCR The Club of Riverdale	40.878628	-73.914568	Tennis Stadium
7	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
8	Marble Hill	40.876551	-73.91066	Starbucks	40.873755	-73.908613	Coffee Shop
9	Marble Hill	40.876551	-73.91066	Astral Fitness & Wellness Center	40.876705	-73.906372	Gym

In those 3093 venues, there were 139 major coffee shops.

```
manhattan_coffee = manhattan_venues[manhattan_venues['Venue Category']=='Coffee Shop']
print(manhattan_coffee.shape)
manhattan_coffee.head(10)
```

(139, 7)

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
8	Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop
73	Chinatown	40.715618	-73.994279	Little Canal	40.714317	-73.990361	Coffee Shop
93	Chinatown	40.715618	-73.994279	Cafe Grumpy	40.715069	-73.989952	Coffee Shop
110	Chinatown	40.715618	-73.994279	Oliver Coffee	40.712986	-73.998106	Coffee Shop
128	Washington Heights	40.851903	-73.936900	Forever Coffee Bar	40.850433	-73.936607	Coffee Shop
161	Washington Heights	40.851903	-73.936900	Starbucks	40.850961	-73.938330	Coffee Shop
241	Inwood	40.867684	-73.921210	Kuro Kirin	40.866223	-73.925691	Coffee Shop
286	Hamilton Heights	40.823604	-73.949688	Monkey Cup	40.825694	-73.947234	Coffee Shop
292	Hamilton Heights	40.823604	-73.949688	Matto Espresso (Espresso Matto)	40.824958	-73.951759	Coffee Shop

And 16 Starbucks coffees

```
manhattan_starbucks = manhattan_venues[manhattan_venues['Venue']=='Starbucks']
print(manhattan_starbucks.shape)
manhattan_starbucks.head(10)
```

(16, 7)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
8	Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop
161	Washington Heights	40.851903	-73.936900	Starbucks	40.850961	-73.938330	Coffee Shop
354	Manhattanville	40.816934	-73.957385	Starbucks	40.815074	-73.958930	Coffee Shop
644	Yorkville	40.775930	-73.947118	Starbucks	40.772356	-73.949984	Coffee Shop
758	Roosevelt Island	40.762160	-73.949168	Starbucks	40.759360	-73.953153	Coffee Shop
1944	Manhattan Valley	40.797307	-73.964286	Starbucks	40.795369	-73.965589	Coffee Shop
1971	Manhattan Valley	40.797307	-73.964286	Starbucks	40.798880	-73.968370	Coffee Shop
2098	Gramercy	40.737210	-73.981376	Starbucks	40.733230	-73.980820	Coffee Shop
2238	Financial District	40.707107	-74.010665	Starbucks	40.705162	-74.011300	Coffee Shop

Then we visualized all those major coffee shops and Starbucks coffee on the Manhattan map, which were in blue and green, respectively.
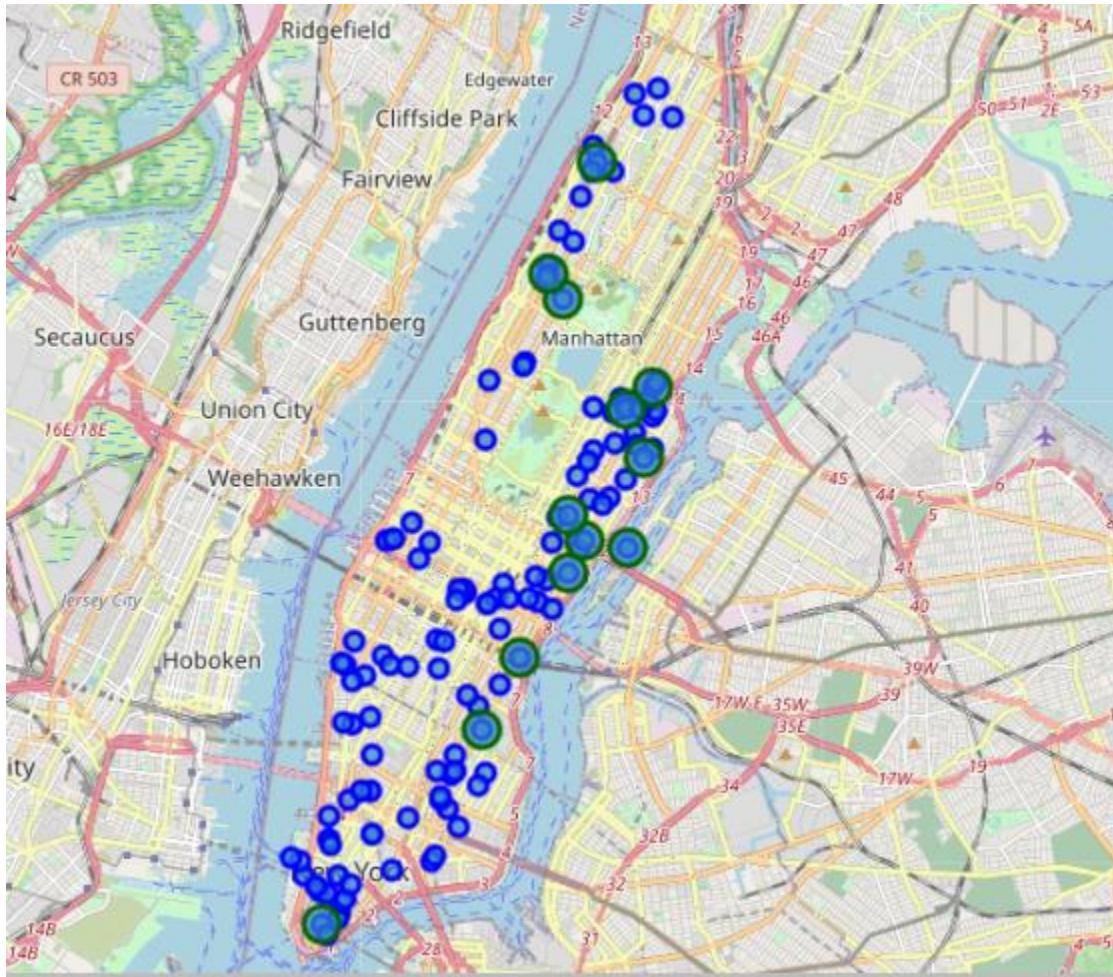
We realized that coffee shops in general and Starbucks in particular, were very crowded on the east side of Manhattan, facing the East River. Starbucks and some other coffee shops also enjoyed some river view, just about 20 percent, in the north of Manhattan, facing the Hudson River. The remaining coffee shops are allocated somewhere in between Midtown and Lower Manhattan, but the trend was not clear as coffee shops near the East River, facing Queens.

3.3. Clustering coffee shops in Manhattan

We analyzed each neighborhood in Manhattan and list down every coffee shop to group the frequency of each coffee shop in each neighborhood before using k-means to cluster them.

Let's group rows by neighborhood and by taking the mean of the frequency of occurrence of each category

```
manhattan_coffee_grouped = manhattan_coffee_onehot.groupby('Neighborhood').mean().reset_index()
print(manhattan_coffee_grouped.shape)
manhattan_coffee_grouped.head(10)
```

(37, 96)

	Neighborhood	787 Coffee	About Coffee	Abraço	Birch Coffee	Bird & Branch	Black Fox Coffee Co.	Blank Slate Coffee + Kitchen	Blue Bottle Cart	Blue Bottle Coffee	Blue Spoon Coffee Co.	Bluestone Lane	C & B Cafe	Cafe Grumpy	Cafe Jax	Caffe Lavazza @ Eataly Fidi	Caffé Vergnano	Café Grumpy
0	Battery Park City	0.0	0.0	0.00	0.000	0.0	0.000	0.0	0.000000	0.200000	0.000	0.0	0.00	0.000000	0.0	0.2	0.0	0.000000
1	Carnegie Hill	0.0	0.0	0.00	0.125	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.000000	0.0	0.0	0.0	0.000000
2	Chelsea	0.0	0.0	0.00	0.000	0.0	0.000	0.0	0.142857	0.142857	0.000	0.0	0.00	0.000000	0.0	0.0	0.0	0.142857
3	Chinatown	0.0	0.0	0.00	0.000	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.333333	0.0	0.0	0.0	0.000000
4	Civic Center	0.0	0.0	0.00	0.200	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.000000	0.0	0.0	0.0	0.000000
5	Clinton	0.0	0.0	0.00	0.000	0.2	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.000000	0.0	0.0	0.0	0.000000
6	East Village	0.0	0.0	0.25	0.000	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.25	0.000000	0.0	0.0	0.0	0.000000
7	Financial District	0.0	0.0	0.00	0.000	0.0	0.125	0.0	0.000000	0.125000	0.125	0.0	0.00	0.000000	0.0	0.0	0.0	0.000000
8	Flatiron	0.0	0.0	0.00	0.000	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.000000	0.0	0.0	0.5	0.000000
9	Gramercy	0.0	0.0	0.00	0.000	0.0	0.000	0.0	0.000000	0.000000	0.000	0.0	0.00	0.000000	0.0	0.0	0.0	0.000000

Using 5 clusters, we clustered coffee shops in Manhattan and visualize it on the map.
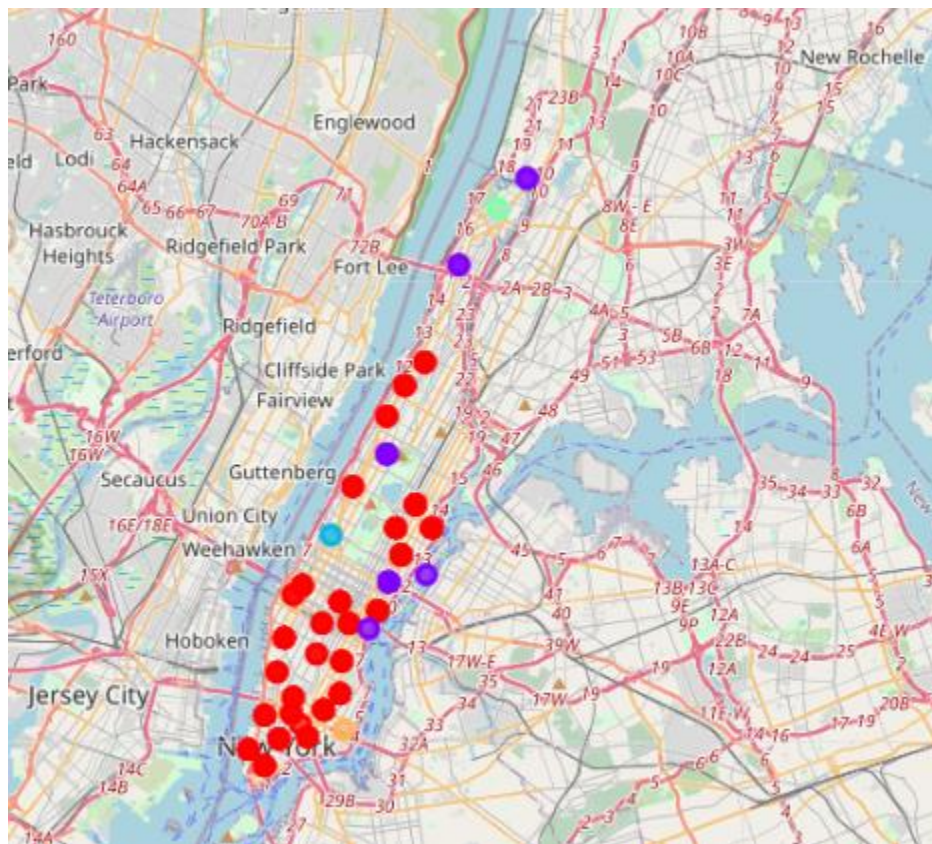

As expected, Starbucks was grouped in a cluster, which was purple points on the map. Starbucks has already occupied the best viewed and most demanded locations in Manhattan, from the east side facing the East River to the north side facing the Hudson River without being interrupted by busy ports.

3.4. Position the Times Square, or the heart of Manhattan

Visualize the Times Square on Manhattan Map to see how far it is from each coffee shops to the heart of Manhattan. Then we realized there are two potential areas to open a coffee shop.

1- Riverside near the Ports, where there was only 1 coffee shop in Cluster 2 in light blue color. There may be some existing problems, but we can solve those problems and exploit the blue ocean market.
2- In the north of Manhattan, where Starbucks has 2 coffee shops but the competition with other coffee shops is minimal. The view must be spectacular there.

In this project, we prioritize locations that are closer to Times Square. So option 1 would be preferred.
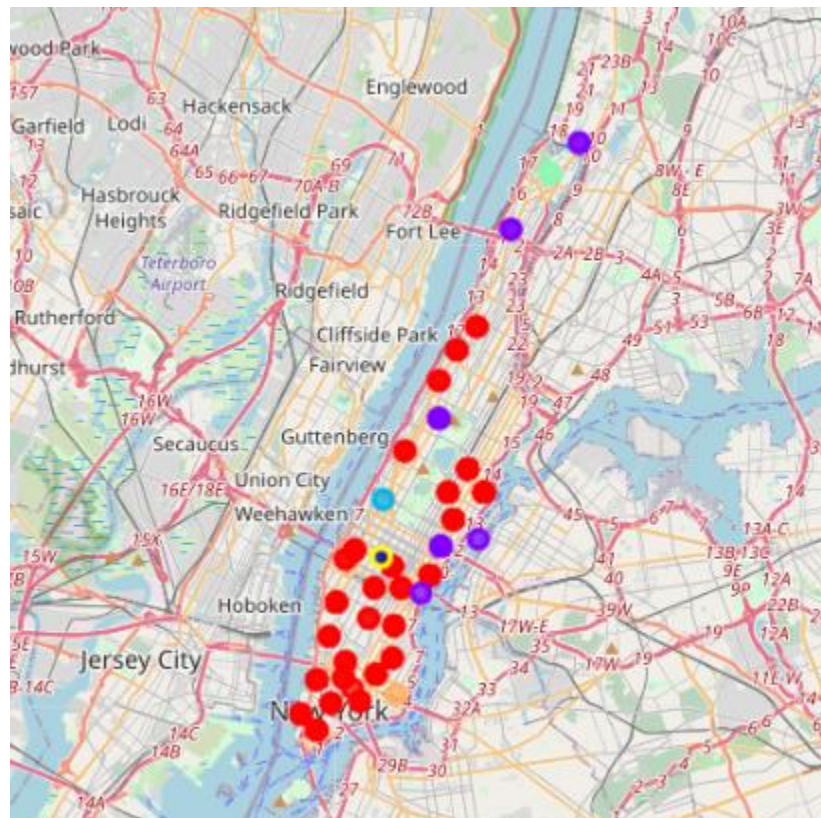
3.5. Review and examine neighborhoods

We talk a look again at the location of each neighborhood, especially the neighborhood that we are targeting for option 1:
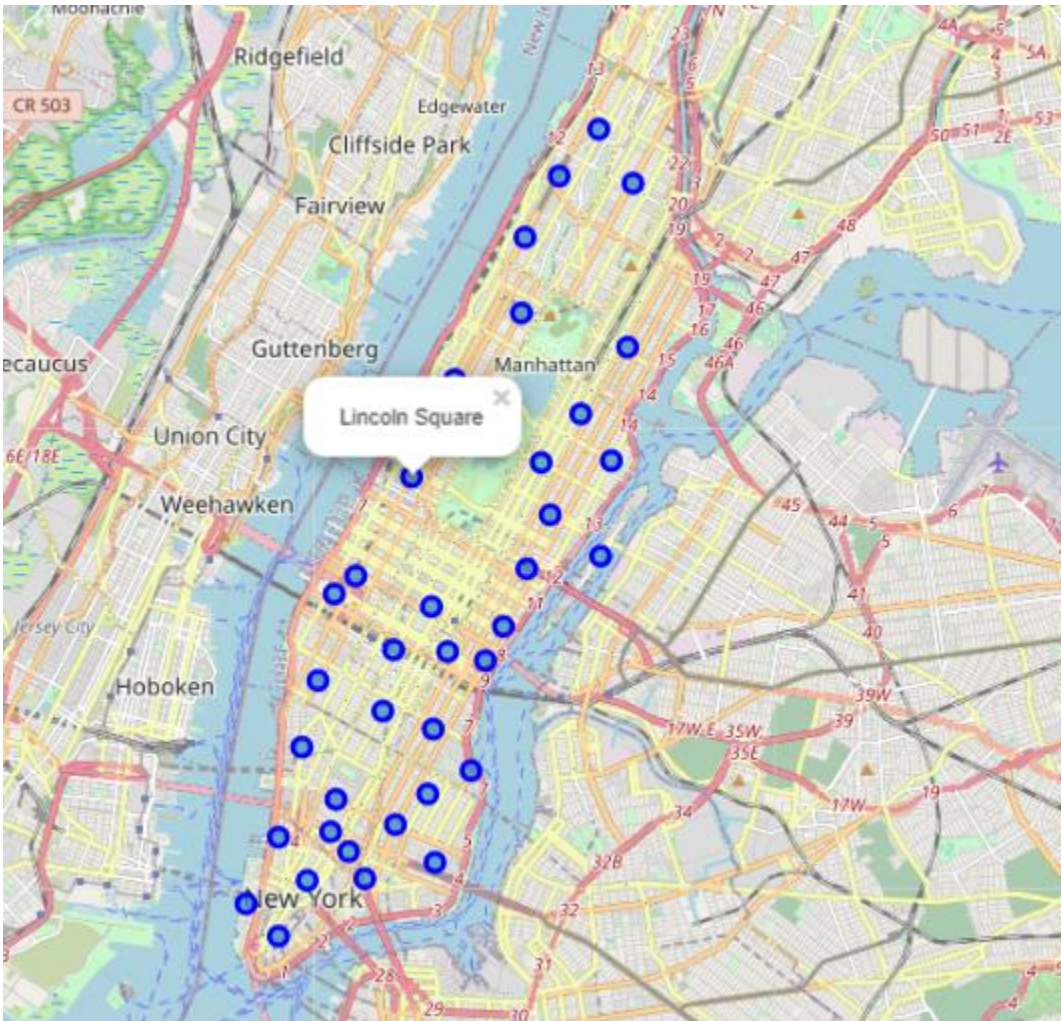
Then we reviewed and list down the top 3 coffee shops in each neighborhood to see the top performers, especially Lincoln Square, where we are aiming for opening a new coffee shop.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Battery Park City	Laughing Man Coffee & Tea	For Five Coffee Roasters	Blue Bottle Coffee
1	Carnegie Hill	Starbucks	Variety Coffee Roasters	DTUT
2	Chelsea	Kobrick Coffee Co.	Blue Bottle Cart	Blue Bottle Coffee
3	Chinatown	Cafe Grumpy	Little Canal	Oliver Coffee
4	Civic Center	Birch Coffee	The Wooly Daily	Laughing Man Coffee & Tea
5	Clinton	Empire Coffee & Tea	Bird & Branch	The Jolly Goat Coffee Bar

...

15	Lincoln Square	Joe: The Art of Coffee	Yorkafe	Handcraft Coffee
16	Little Italy	Café Integral	Yorkafe	Dear Mama Coffee
17	Lower East Side	Blue Bottle Coffee	Yorkafe	Handcraft Coffee

<div align="center">…</div>

34	Washington Heights	Forever Coffee Bar	Starbucks	Yorkafe
35	West Village	Partners Coffee Roasters	Oslo Coffee Roasters	The ELK
36	Yorkville	Yorkafe	Stella & Fly	Oren's Daily Roast

4. Result

We realized there are two potential areas to open a coffee shop.

1- Riverside near the Ports, where there was only 1 coffee shop in Cluster 2 in light blue color. The target neighborhood is Lincoln Square.
2- In the north of Manhattan, where Starbucks has 2 coffee shops but the competition with other coffee shops is minimal. The view must be spectacular there. However, it may be too far from the heart of Manhattan

Because we are looking for a location is closer to Times Square, option 1 would be preferred. The coffee competition in the Lincoln Square neighborhood is much less than in other neighborhoods in Manhattan. It is a great chance to open a coffee shop there, where is close to Times Square and Central Park, and also have a great view of the Hudson River.

5. Discussion

Manhattan is a beautiful and densely populated borough and is the heart of New York City. With a total population of approximately 4 million. There is a huge demand for coffee in the borough. Currently, there are over one hundred major coffee shops and hundreds of small ones within 23 square miles of Manhattan. Therefore, it is not easy to find the best location to open your own coffee shop.

I used the K means algorithm as part of this clustering study. I found out Starbucks are clustered for their own group. It seems that they have already searched for the best locations to open their coffee shops. To open our own coffee shop, we can stay away from them and also learn from them. Where they located their business is the best locations with the highest demand. We can follow the coffee giant to be the second biggest coffee provider as well.

There is some information that has not been discussed in the project, for example, the income in each neighborhood, the view, and the air quality of each neighborhood. There are always some factors that we have not touched yet. But geographically, the location we have decided to choose is the best option.

6. Conclusion

We will open our coffee shop in the Lincoln Square neighborhood.