Neighborhood Clustering - Manhattan, NYC

If someone is looking to open a coffeeshop, where would you recommend that they open it?

Gulnaz Khabibullina

June 10, 2021

1. Introduction

1.1 Background

New York is the most populous city in the US state of New York with over 8 million residents. New York is also very diverse as its residents come from various backgrounds and ethnicities. Diverse population and busy lifestyle provide many business opportunities to New York residents including the coffee shop industry.

1.2 Business statement

Opening a new coffeeshop in a city such as New York takes many factors into consideration: financing, right location, catchy name, unique environment, etc. New York consists of five boroughs: Manhattan, The Bronx, Brooklyn, Queens, and Staten Island. Among these five boroughs, Manhattan definitely stands out as it is often described as the cultural and financial center of the world. (1) Therefore, in this report I will focus on coffee shop industry business opportunities that Manhattan has to offer.

In this report, I will identify if the area has a potential success of opening a coffee shop based on the following criteria:

- number of coffee shops represented in the particular neighborhood of Manhattan
- identify which neighborhood of Manhattan has the smallest number of coffeeshops
- identify potential success of new coffee shop in the particular neighborhood of Manhattan

1.3. Target Audience

This research will be helpful for businessmen and businesswomen thinking of opening a new coffee shop in Manhattan, NYC. The objective is to recommend the best neighborhood to start such a business.

2. Data

In order to do the research in this report, data will be gathered through the following sources:

- For this analysis I will be using a New York City dataset that has data on all five boroughs of New York City as well as longitude and latitude of each neighborhood. The dataset allows me to explore the neighborhoods in Manhattan, NYC. Below is the data sample:

	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

- In order to identify all the coffee shops in Manhattan neighborhoods I will be using Foursquare location dataset. Below is the data sample:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
6	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
66	Chinatown	40.715618	-73.994279	Little Canal	40.714317	-73.990361	Coffee Shop
117	Chinatown	40.715618	-73.994279	Oliver Coffee	40.712986	-73.998106	Coffee Shop
134	Washington Heights	40.851903	-73.936900	Forever Coffee Bar	40.850433	-73.936607	Coffee Shop
183	Washington Heights	40.851903	-73.936900	Starbucks	40.850961	-73.938330	Coffee Shop

- The Foursquare API will be used to collect information regarding neighborhoods with a smallest number of existing coffee shops. Below is the data sample:

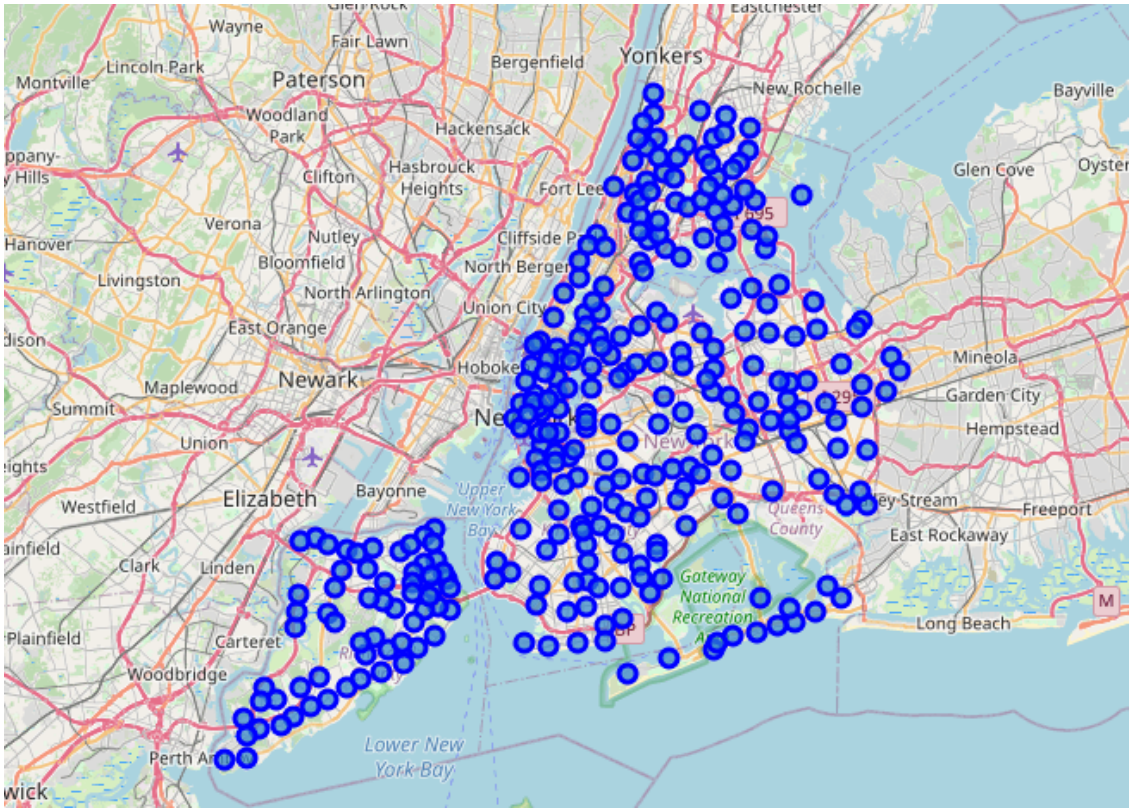
	Neighborhood	Venue Category
13	Inwood	1
19	Marble Hill	1
30	Tudor City	1
27	Stuyvesant Town	1
15	Lincoln Square	2
25	Roosevelt Island	2
28	Sutton Place	2
29	Tribeca	2
34	Washington Heights	2

3. Methodology

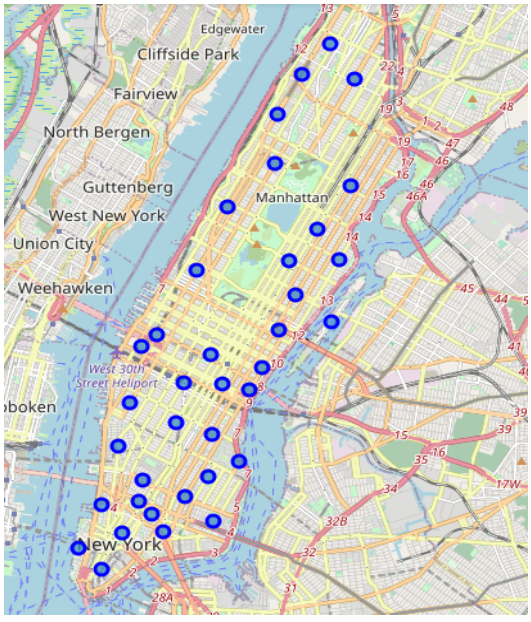
The objective of this project is to find which neighborhood of Manhattan, NYC is a good choice for a new coffee shop business to open. For data analysis I will be using a New York City dataset that has data on the neighborhoods in Manhattan as well as longitude and latitude of each neighborhood. Manhattan, NYC has 53 neighborhoods and the goal of my report is to cluster these neighborhoods based on the presented coffee shops that each of them has.

Exploratory Data Analysis

- Load New York City dataset that has data on all five boroughs of New York City as well as longitude and latitude of each neighborhood. The dataset allows me to explore the neighborhoods in Manhattan, NYC and create a map of New York with neighborhoods using folium Python library.



- Slice the original New York City dataframe and create a new dataframe of the Manhattan data and Manhattan neighborhoods. Visualize Manhattan neighborhoods using folium Python library.



- Utilize the Foursquare API to explore the Manhattan neighborhoods and identify all presented venues.
- Filter Manhattan dataset's Venue Category to identify Coffee Shops in each presented neighborhood.
- Use data to calculate the number of coffee shops presented in each neighborhood of Manhattan.
- Identify which neighborhoods of Manhattan have the smallest number of presented coffee shops.
- Cluster coffee shops and neighborhoods using the k-means clustering algorithm.

According to Foursquare, there are 115 coffee shops in Manhattan. Below is the total number of coffee shops in each neighborhood of Manhattan:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Battery Park City	4	4	4	4	4	4
Carnegie Hill	5	5	5	5	5	5
Chelsea	5	5	5	5	5	5
Chinatown	2	2	2	2	2	2
Civic Center	6	6	6	6	6	6
Clinton	6	6	6	6	6	6
East Village	3	3	3	3	3	3
Financial District	9	9	9	9	9	9
Flatiron	3	3	3	3	3	3
Gramercy	3	3	3	3	3	3
Greenwich Village	3	3	3	3	3	3

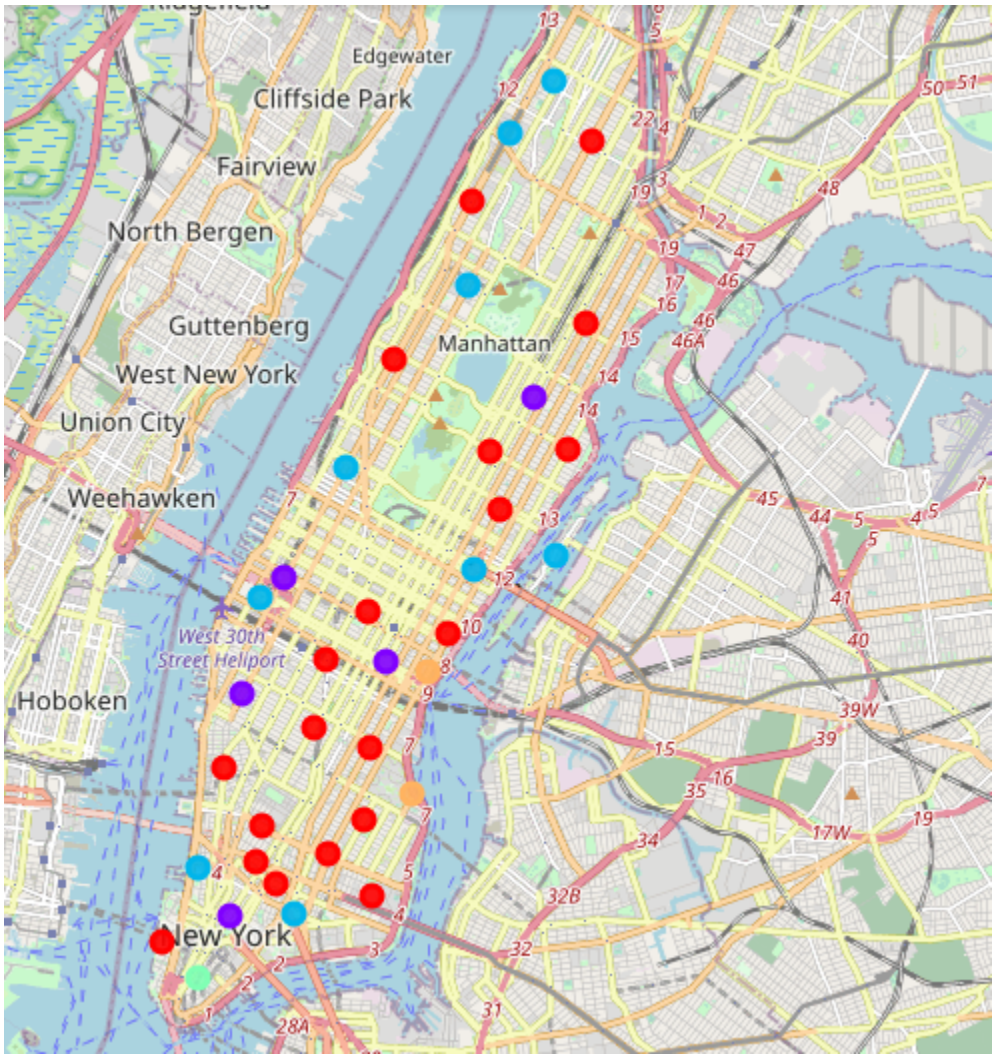
Neighborhoods with smallest number of coffeeshops:

	Neighborhood	Venue Category
13	Inwood	1
27	Stuyvesant Town	1
19	Marble Hill	1
30	Tudor City	1
28	Sutton Place	2
29	Tribeca	2
17	Manhattan Valley	2
15	Lincoln Square	2
25	Roosevelt Island	2
12	Hudson Yards	2
11	Hamilton Heights	2
18	Manhattanville	2
3	Chinatown	2
34	Washington Heights	2

4. Results.

Cluster the coffee shops and neighborhoods into 5 clusters using the k-means clustering algorithm. Using the Foursquare location data, I can extract the neighborhood's location and the number of coffee shops in each neighborhood. Applying the k-means clustering algorithm on the number of coffee shops in each neighborhood the dataset is partitioned into 5 clusters. The 5 clusters are partitioned based on a similar number of coffee shops that belong to neighborhoods.

Clustered data representation with the use of folium Python library:



Examining the Clusters:

Now, I can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Cluster 1:

	Neighborhood	Venue Category
0	Marble Hill	1.0
1	Chinatown	2.0
2	Washington Heights	2.0
3	Inwood	1.0
4	Hamilton Heights	2.0
5	Manhattanville	2.0
6	Central Harlem	0.0
7	East Harlem	0.0
11	Roosevelt Island	2.0
13	Lincoln Square	2.0
20	Lower East Side	0.0
21	Tribeca	2.0
25	Manhattan Valley	2.0

Cluster 2:

	Neighborhood	Venue Category
29	Financial District	9.0

Cluster 3:

	Neighborhood	Venue Category
14	Clinton	6.0
16	Murray Hill	5.0
17	Chelsea	5.0
30	Carnegie Hill	5.0
32	Civic Center	6.0

Cluster 4:

	Neighborhood	Venue Category
8	Upper East Side	3.0
10	Lenox Hill	3.0
12	Upper West Side	3.0
18	Greenwich Village	3.0
19	East Village	3.0
22	Little Italy	3.0
23	Soho	3.0
24	West Village	3.0
26	Morningside Heights	3.0
27	Gramercy	3.0
33	Midtown South	3.0
35	Turtle Bay	3.0
38	Flatiron	3.0

Cluster 5:

	Neighborhood	Venue Category
9	Yorkville	4.0
15	Midtown	4.0
28	Battery Park City	4.0
31	Noho	4.0

5.Discussions

After performed analysis, the following recommendations can be made:

- By observing the cluster representation on a map, I can say that cluster 1 (red) has the smallest number of represented coffee shops and there is a good scope to open a new coffee shop in these neighborhoods.
- According to the dataset, neighborhoods such as Central Harlem, East Harlem, and Lower East Side have 0 represented coffee shops. Other neighborhoods in cluster 1 have 1 or 2 represented coffee shops which shows a good potential to open a new coffee shop in one of these neighborhoods.
- According to the results, neighborhoods such as Financial District, Chelsey, or Carnegie Hall with a large number of represented coffee shops are not recommended for consideration for opening coffee shop business.

6.Conclusion

New York City is a big city with its residents coming from various backgrounds. It is also a busy city that never sleeps and opening a new coffee shop in a city like New York sounds like a good idea. However, many factors have to be taken into consideration such as location, demand, etc. In this analysis, I focused mainly on one of the boroughs of New York City which is Manhattan and locations with the smallest number of coffee shops. However, the analysis can be expanded by further analysis of types of coffee shops, cafes, and other bakeries in specific areas. For instance, whether it is a traditional coffee shop or shop with unique brewing styles, does it offer tea or pastry, etc. Based on the current results it can be observed that the North and Northeast sides of Manhattan with its neighborhoods have big potential for successful coffee shop business.

7. References:

https://en.wikipedia.org/wiki/New_York_City