The Battle of the Neighborhoods - Report

1. **Introduction & Business Problem:**

   The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. This means that many people use various transport. The most cheap and safe transport is bus. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analyzed carefully.

   Business Problem

- There are many transports and high-density population in New York City, NY
- It is not easy to find bast bus stop to improve public transport

2. **Data**

   Data 1: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segments the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is: https://geo.nyu.edu/catalog/nyu_2451_34572

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Tibbett & W 230Th St Mta Bus Stop	40.879627	-73.909990	Bus Stop
1	Marble Hill	40.876551	-73.91066	MTA New York City Bus & Bus Company - Bx10/Bx2...	40.879444	-73.912384	Bus Stop
2	Marble Hill	40.876551	-73.91066	MTA Bus - M100/Bx7/Bx20/BxM1 @ Broadway & W. 2...	40.872234	-73.912636	Bus Stop
3	Marble Hill	40.876551	-73.91066	mta bus stop #100705 (Independance AVE &Henry ...	40.880829	-73.918971	Bus Stop
4	Marble Hill	40.876551	-73.91066	MTA Bus - Bx1/Bx1LTD/Bx7/Bx10 @ W. 231st St & ...	40.878803	-73.904570	Bus Stop

   Data2: Newyork city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

   In addition, Bus Stop category Id 52f2ab2ebcbc57f1066b8b4f is used for retrieving data from Foursquare API.

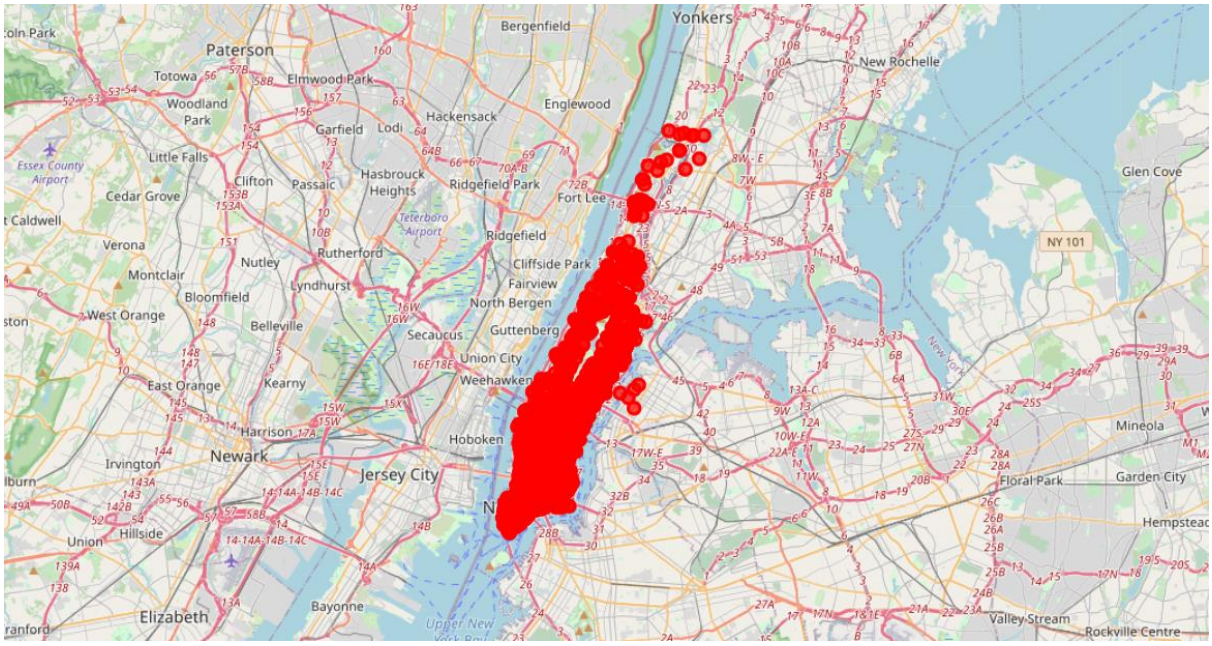
3. Methodology

In this project, I will use the basic methodology as taught in Week 3 lab.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Bus Stop	General Travel	Bus Station	Bus Line
1	Manhattan	Chinatown	40.715618	-73.994279	0	Bus Stop	Bus Line	General Travel	Bus Station
2	Manhattan	Washington Heights	40.851903	-73.936900	2	Bus Stop	General Travel	Bus Station	Bus Line
3	Manhattan	Inwood	40.867684	-73.921210	2	Bus Stop	General Travel	Bus Station	Bus Line
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Bus Stop	General Travel	Bus Station	Bus Line

Above, I have done convert addresses into their equivalent latitude and longitude values. Then we will use the Foursquare API to explore neighborhoods in Manhattan, New York. After that, explore function to get bus stop categories in each neighborhood.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Tibbett & W 230Th St Mta Bus Stop	40.879627	-73.909990	Bus Stop
1	Marble Hill	40.876551	-73.91066	MTA New York City Bus & Bus Company - Bx10/Bx2...	40.879444	-73.912384	Bus Stop
2	Marble Hill	40.876551	-73.91066	MTA Bus - M100/Bx7/Bx20/BxM1 @ Broadway & W. 2...	40.872234	-73.912636	Bus Stop
3	Marble Hill	40.876551	-73.91066	mta bus stop #100705 (Independance AVE &Henry ...	40.880829	-73.918971	Bus Stop
4	Marble Hill	40.876551	-73.91066	MTA Bus - Bx1/Bx1LTD/Bx7/Bx10 @ W. 231st St & ...	40.878803	-73.904570	Bus Stop

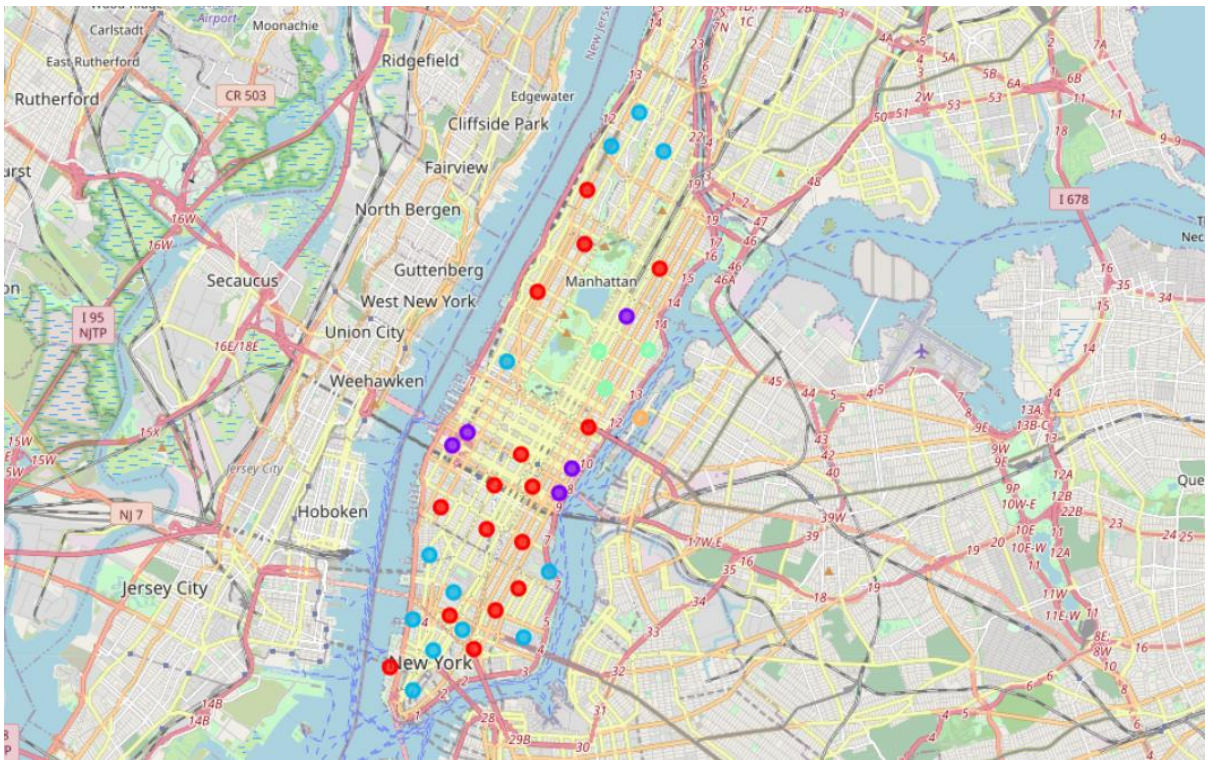


Bus Stop in NY

	Neighborhood	Asian Restaurant	Bakery	Chinese Restaurant	Cocktail Bar	Deli / Bodega	Grocery Store	Hawaiian Restaurant	Japanese Restaurant	Noodle House	Poke Place	Ramen Restaurant	Restaurant	Sake Bar	Sandwich Place	R
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Chinatown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Chinatown	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
3	Chinatown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Chinatown	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Then use this feature to group the neighborhoods into clusters K-means clustering algorithm will be use to complete this task. And also, the Folium library to visualize the neighborhoods in NY and its emerging clusters.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Manhattan	Marble Hill	40.876551	-73.910660	2	Bus Stop	General Travel	Bus Station	Bus Line
1	Manhattan	Chinatown	40.715618	-73.994279	0	Bus Stop	Bus Line	General Travel	Bus Station
2	Manhattan	Washington Heights	40.851903	-73.936900	2	Bus Stop	General Travel	Bus Station	Bus Line
3	Manhattan	Inwood	40.867684	-73.921210	2	Bus Stop	General Travel	Bus Station	Bus Line
4	Manhattan	Hamilton Heights	40.823604	-73.949688	2	Bus Stop	General Travel	Bus Station	Bus Line



4. Results

K-mean Cluster Using K-mean to clustering data area with less number of Bus Stop

Cluster 1

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
14	Clinton	Bus Stop	Bus Line	General Travel	Bus Station
30	Carnegie Hill	Bus Stop	Bus Line	Bus Station	General Travel
35	Turtle Bay	Bus Stop	Bus Line	General Travel	Bus Station
36	Tudor City	Bus Stop	Bus Line	General Travel	Bus Station
39	Hudson Yards	Bus Stop	Bus Line	General Travel	Bus Station

Cluster 2

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	Marble Hill	Bus Stop	General Travel	Bus Station	Bus Line
2	Washington Heights	Bus Stop	General Travel	Bus Station	Bus Line
3	Inwood	Bus Stop	General Travel	Bus Station	Bus Line
4	Hamilton Heights	Bus Stop	General Travel	Bus Station	Bus Line
5	Manhattanville	Bus Stop	General Travel	Bus Station	Bus Line
6	Central Harlem	Bus Stop	General Travel	Bus Station	Bus Line
13	Lincoln Square	Bus Stop	General Travel	Bus Station	Bus Line
18	Greenwich Village	Bus Stop	General Travel	Bus Station	Bus Line
20	Lower East Side	Bus Stop	General Travel	Bus Station	Bus Line
21	Tribeca	Bus Stop	General Travel	Bus Station	Bus Line
22	Little Italy	Bus Stop	General Travel	Bus Station	Bus Line
24	West Village	Bus Stop	General Travel	Bus Station	Bus Line
29	Financial District	Bus Stop	General Travel	Bus Station	Bus Line
32	Civic Center	Bus Stop	General Travel	Bus Station	Bus Line
37	Stuyvesant Town	Bus Stop	General Travel	Bus Station	Bus Line

Cluster 3

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
8	Upper East Side	Bus Stop	Bus Line	Bus Station	General Travel
9	Yorkville	Bus Stop	Bus Station	Bus Line	General Travel
10	Lenox Hill	Bus Stop	Bus Station	Bus Line	General Travel

Cluster 4

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
11	Roosevelt Island	Bus Stop	Bus Line	General Travel	Bus Station

Based on dataframe analysis above Cluster 3 (Upper East Side) and Cluster 4 (Roosevelt Island) areas are the best places to open a new bus stop.

5. Discussion

In this section, I would be discussing the observations I have noted and the recommendation that I can make based on the results.

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results.

- Upper East Side has also potential where closes to Linox Hill area.
- It can be done more detailed analysis by adding other factors such as transportation, demographics of inhabitants.

Finally, FourSquare proved to be a good source of data but frustrating at times. Despite having a Developer account I regularly exceeded my hourly limit locking me out for the day.

6. Conclusion

Although all of the goals of this project were met there is definitely room for further improvement and development as noted below. However, the goals of the project were met and, with some more work, could easily be developed optimal locations for bus stops.