

The Battle of Neighbourhoods

REPORT

Introduction

New York City has a population of over 8 million across five boroughs. The aim of this project is to determine which neighbourhood(s) across the entire city would be a suitable place for someone looking to set up a coffee shop in New York. Anyone interested in opening up a coffee shop somewhere in New York would benefit from this project as it will give them an idea of the current distribution of coffee shops all over the city, and recommend in which neighbourhoods they should consider opening their coffee shop in order to be more successful.

Data

The data used in this project will be Foursquare location data for New York, including the names of all of the Neighbourhoods, their latitude and longitude, the names of all the venues, venue latitude and longitude, and venue category. These can be written into a dataframe which will be used to eventually solve the overall problem of this project and plot a map showing the distribution of coffee shops as they stand currently, and guide the user on where they should consider opening their coffee shop.

Methodology

After obtaining the necessary Foursquare location data for New York, I wrote it into this dataframe:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
2	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
3	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
4	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop

Next, since this project only considers coffee shops, I created a new dataframe which only references the rows where the category is 'Coffee Shop'. As this process deletes all rows which do not apply to coffee shops, all of the neighbourhoods mentioned in the resulting dataframe must have at least one coffee shop.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
71	Kingsbridge	40.881687	-73.902818	Mon Amour Coffee & Wine	40.885009	-73.900332	Coffee Shop
132	Marble Hill	40.876551	-73.910660	Starbucks	40.877531	-73.905582	Coffee Shop
137	Marble Hill	40.876551	-73.910660	Starbucks	40.873755	-73.908613	Coffee Shop
179	Norwood	40.877224	-73.879391	Nicky's Coffee Shop	40.874933	-73.879404	Coffee Shop
242	Pelham Parkway	40.857413	-73.854756	Liberty Donut & Coffee Shop	40.855339	-73.855333	Coffee Shop
317	Bedford Park	40.870185	-73.885512	National Coffee Shop	40.872841	-73.889053	Coffee Shop
473	West Farms	40.839475	-73.877745	Prospect Coffee Shop	40.837577	-73.880839	Coffee Shop
550	Mott Haven	40.806239	-73.916100	Brook Lunch	40.807472	-73.919510	Coffee Shop
649	Throgs Neck	40.815109	-73.816350	The Miles Coffee Bar	40.819462	-73.817352	Coffee Shop
726	Van Nest	40.843608	-73.866299	Conti's Pastry Shoppe	40.845906	-73.862836	Coffee Shop
758	Morris Park	40.847549	-73.850402	La Casa Del Caffè	40.848675	-73.854973	Coffee Shop

At this stage, all the rows show information about coffee shop locations. Therefore the 'Venue Category' column is no longer useful. The final map showing distribution will plot the co-ordinates of neighbourhoods, not that of individual venues. So the venue longitude and latitude is not necessary. Furthermore the venue name is of no use, and so next I removed these columns from the dataframe:

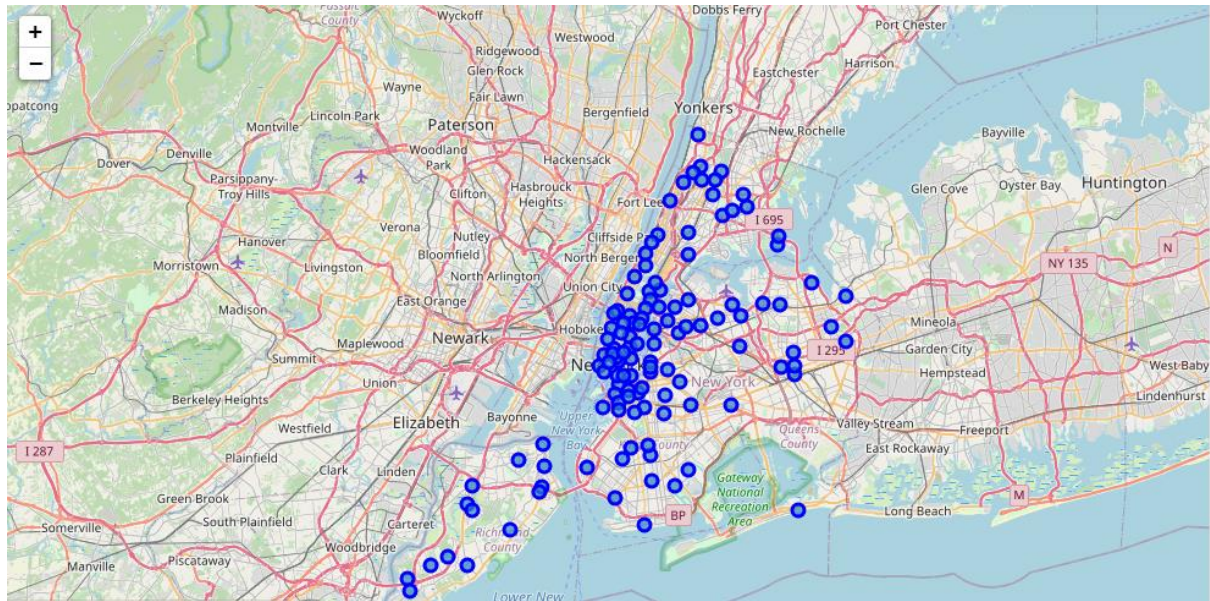
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude
71	Kingsbridge	40.881687	-73.902818
132	Marble Hill	40.876551	-73.910660
137	Marble Hill	40.876551	-73.910660
179	Norwood	40.877224	-73.879391
242	Pelham Parkway	40.857413	-73.854756
317	Bedford Park	40.870185	-73.885512
473	West Farms	40.839475	-73.877745
550	Mott Haven	40.806239	-73.916100
649	Throgs Neck	40.815109	-73.816350
726	Van Nest	40.843608	-73.866299
758	Morris Park	40.847549	-73.850402

In order to calculate the number of coffee shops in every neighbourhood in the dataframe, I created another column which shows this frequency by counting the number of times each unique location occurs in the dataframe:

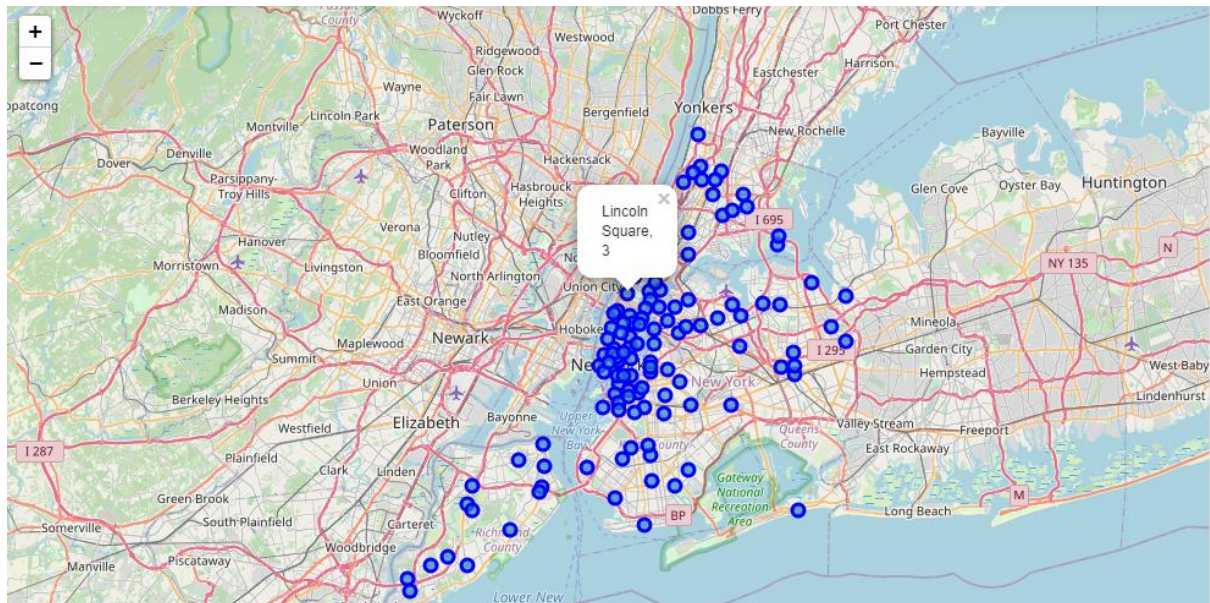
	Neighborhood	Latitude	Longitude	Number of Coffee Shops
71	Kingsbridge	40.881687	-73.902818	1
132	Marble Hill	40.876551	-73.910660	2
179	Norwood	40.877224	-73.879391	1
242	Pelham Parkway	40.857413	-73.854756	1
317	Bedford Park	40.870185	-73.885512	1
473	West Farms	40.839475	-73.877745	1
550	Mott Haven	40.806239	-73.916100	1
649	Throgs Neck	40.815109	-73.816350	1
726	Van Nest	40.843608	-73.866299	1
758	Morris Park	40.847549	-73.850402	1
828	Belmont	40.857277	-73.888452	2

(This resulting dataframe had 122 rows, meaning that 122 neighbourhoods out of 306 (as was calculated previously) have at least one coffee shop.)

Next, using this dataframe I created a map using folium which visualises the neighbourhoods with the number of coffee shops in each one.



Below you can see that Lincoln Square has three coffee shops:



In order to be more useful, I decided to categorise the data I had into three groups, separated by the number of coffee shops they contained.

Group A had three or less coffee shops:

	Neighborhood	Latitude	Longitude	Number of Coffee Shops
71	Kingsbridge	40.881687	-73.902818	1
132	Marble Hill	40.876551	-73.910660	2
179	Norwood	40.877224	-73.879391	1
242	Pelham Parkway	40.857413	-73.854756	1
317	Bedford Park	40.870185	-73.885512	1
473	West Farms	40.839475	-73.877745	1
550	Mott Haven	40.806239	-73.916100	1
649	Throgs Neck	40.815109	-73.816350	1
726	Van Nest	40.843608	-73.866299	1
758	Morris Park	40.847549	-73.850402	1
828	Belmont	40.857277	-73.888452	2

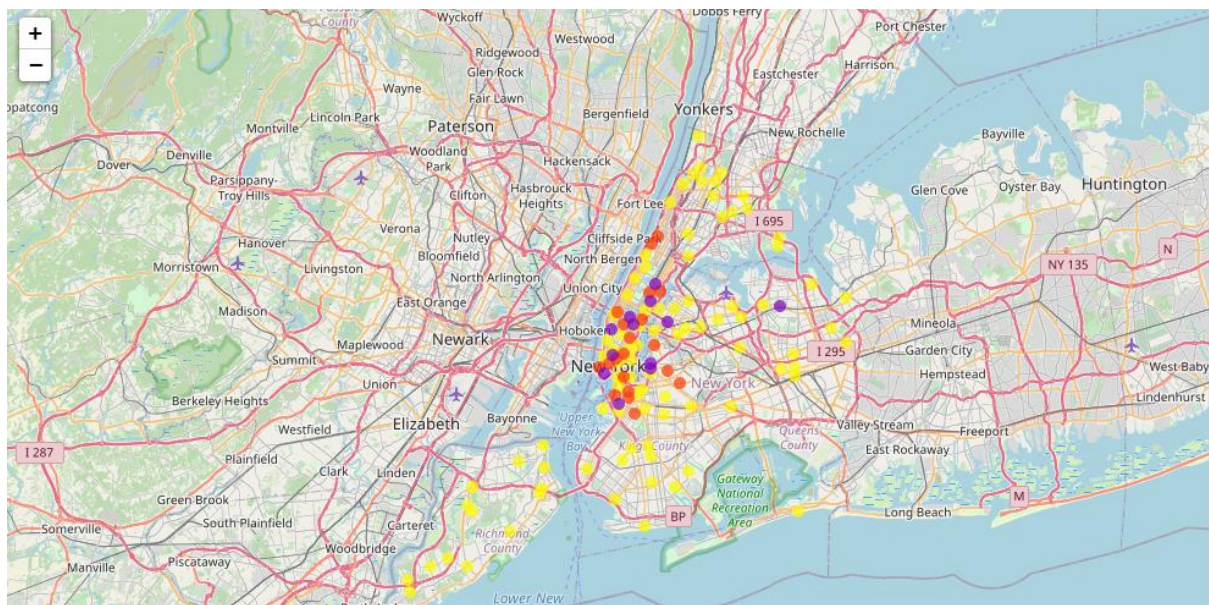
Group B had between four and six coffee shops:

	Neighborhood	Latitude	Longitude	Number of Coffee Shops
1235	Greenpoint	40.730201	-73.954241	6
1712	Bushwick	40.698116	-73.925258	5
1931	Cobble Hill	40.687920	-73.998561	4
2283	Park Slope	40.672321	-73.977050	5
2654	Downtown	40.690844	-73.983463	5
2748	Boerum Hill	40.685683	-73.983748	5
3054	East Williamsburg	40.708492	-73.938858	4
3659	Hamilton Heights	40.823604	-73.949688	4
3719	Manhattanville	40.816934	-73.957385	4
3839	Upper East Side	40.775639	-73.960508	6
3934	Yorkville	40.775930	-73.947118	5
4361	Clinton	40.759101	-73.996119	4
5082	Little Italy	40.719324	-73.997305	4
5454	Gramercy	40.737210	-73.981376	4
5568	Battery Park City	40.711932	-74.016869	4
8416	Noho	40.723259	-73.988434	6
8550	Civic Center	40.715229	-74.005415	6
8606	Midtown South	40.748510	-73.988713	4

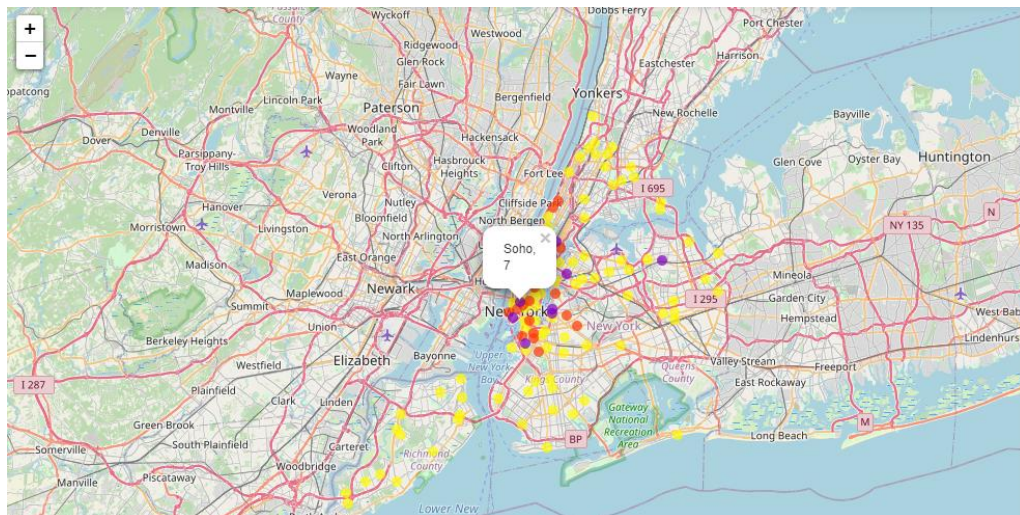
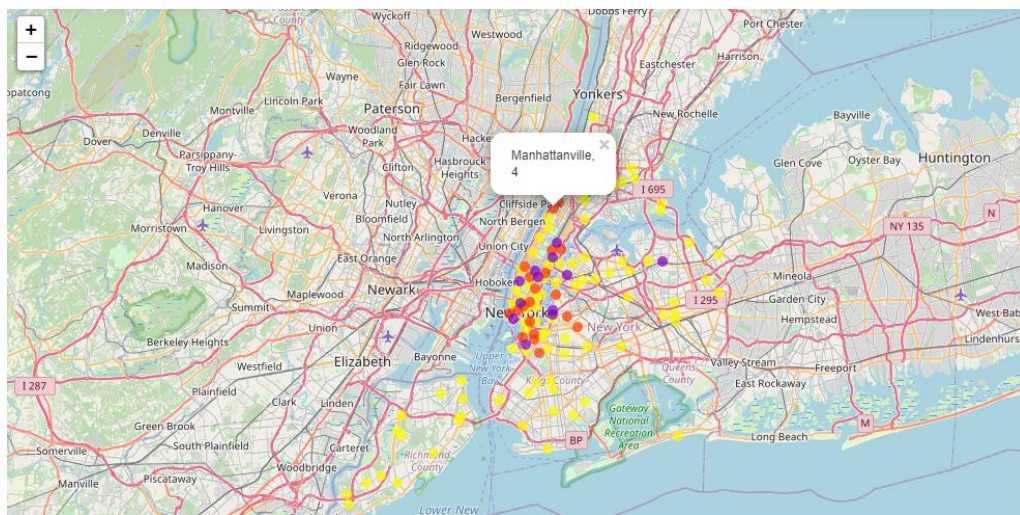
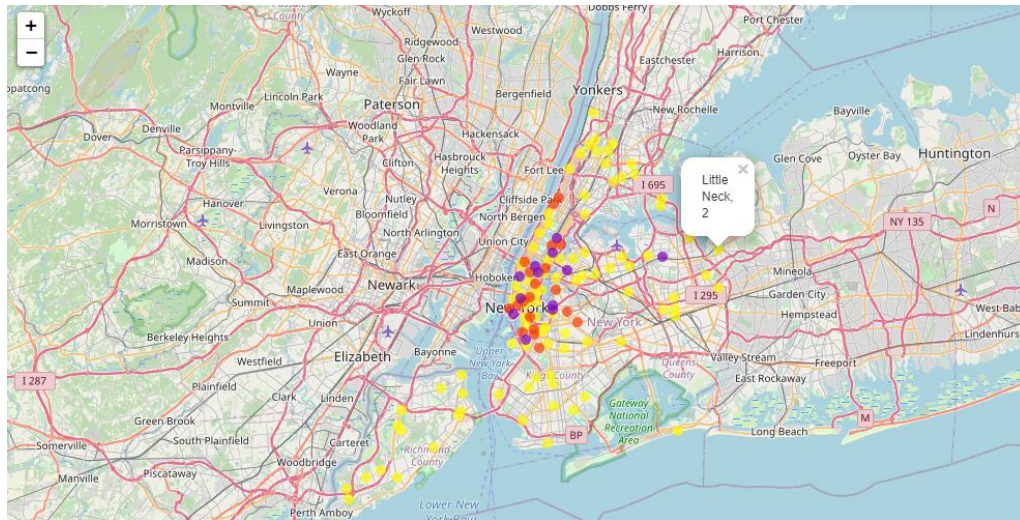
Group C had seven or more coffee shops:

	Neighborhood	Latitude	Longitude	Number of Coffee Shops
1997	Carroll Gardens	40.680540	-73.994654	7
3127	North Side	40.714823	-73.958809	9
3225	South Side	40.710861	-73.958001	7
4072	Lenox Hill	40.768113	-73.958860	7
4460	Midtown	40.754691	-73.981669	8
4542	Murray Hill	40.748303	-73.978332	7
4647	Chelsea	40.744035	-74.003116	9
5207	Soho	40.722184	-74.000657	7
5614	Financial District	40.707107	-74.010665	8
6234	Long Island City	40.750217	-73.939202	8
7236	Murray Hill	40.764126	-73.812763	7
8314	Carnegie Hill	40.782683	-73.953256	7

Finally, I used these new categories to create the map again, this time showing the different groups using different colours:



As an example, you can see that Little Neck has two coffee shops, Manhattanville has four and Soho has seven:



Results

The results of this project show how coffee shops are distributed across New York City. It is recommended that someone wanting to open a coffee shop should choose a location which has between four to six coffee shops already, shown on the map as orange dots, for example Manhattanville from the image above.

This is because the purple dots, showing neighbourhoods with at least seven coffee shops, would provide more competition for a newcomer. The neighbourhoods represented with a yellow dot have small numbers of coffee shops, suggesting there is not as much demand for more coffee shops compared to the other neighbourhoods. Therefore it is suggested that a new coffee shop may benefit from being located in a neighbourhood shown in orange on the map.

Also note that there seems to be a higher density of coffee shops towards the centre of New York which indicates more competition in those areas whilst simultaneously suggesting higher demand for coffee shops.

Discussion

(The recommendations for users from this project are given above, in the *Results* section)

This project shows how neighbourhoods across New York City are distributed in relation to the number of coffee shops, and categorises them based on this.

The user can make a decision based on the colour of the dot of the neighbourhood they are considering, as well as its location in relation to other neighbourhoods and densities of coffee shops in surrounding neighbourhoods.

This project only considers the number of coffee shops in neighbourhoods, whereas in reality the demand and competition in an area would depend on various other factors such as population of the neighbourhood and its size. These are factors which I would take into consideration if completing this project again, as it would provide additional perspectives and allow the user to make a more informed decision.

Conclusion

To summarise, this project is useful for someone who wants to open a coffee shop in New York. It will give them an idea of how coffee shops are currently distributed across the city, including categorising the neighbourhoods depending on the numbers of coffee shops they already contain, helping them in making their decision. This project can be further improved by taking into account other factors such as neighbourhood population, however this project provides a good initial view and will benefit the user in choosing a suitable location for their coffee shop.