

The battle of neighbourhood

1. Introduction/Business Problem

New York City, the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over about 302.6 square miles (784 km²), New York is also the most densely populated major city in the United States. As one of the most popular cities around the globe. New York city attract tons of tourists annually. There are a different variety of cuisines for the tourists to choose.

This study is to help some small individual investors who would like to open up pharmacy shop at New York City. We will utilize the data to gather the information to identify which place would be the ideal place for a newly established pharmacy shop. The analysis will also include a bunch of parameters which the investors need to consider for decision-making.

2. Data

In this report, we will be using the following data:

New York City data that contains list Boroughs, Neighborhoods along with their latitude and longitude.

Data source : https://cocl.us/new_york_dataset

Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of new york city.

pharmacy shops in each neighborhood of new york city.

Data source : Fousquare API

Description : By using this api we will get all the venues in each neighborhood. We can filter these venues to get only coffee shops.

GeoSpace data

Data source : <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm>

Description : By using this geo space data we will get the New york Borough boundaries that will help us visualize choropleth map.

3. Methodology

BeautifulSoup to scrape information and parse data from the Wikipedia page, so that I can get an organized table of neighborhoods information of New York City.

In addition, Geopy is used to get the exact geological location of the neighbourhood.

Furthermore, we use Foursquare API to search venue information of the neighbourhood to gain more understanding of each neighbourhood.

Finally, we use Kmeans to cluster similar neighbourhood to reach our conclusion.

4. Analysis

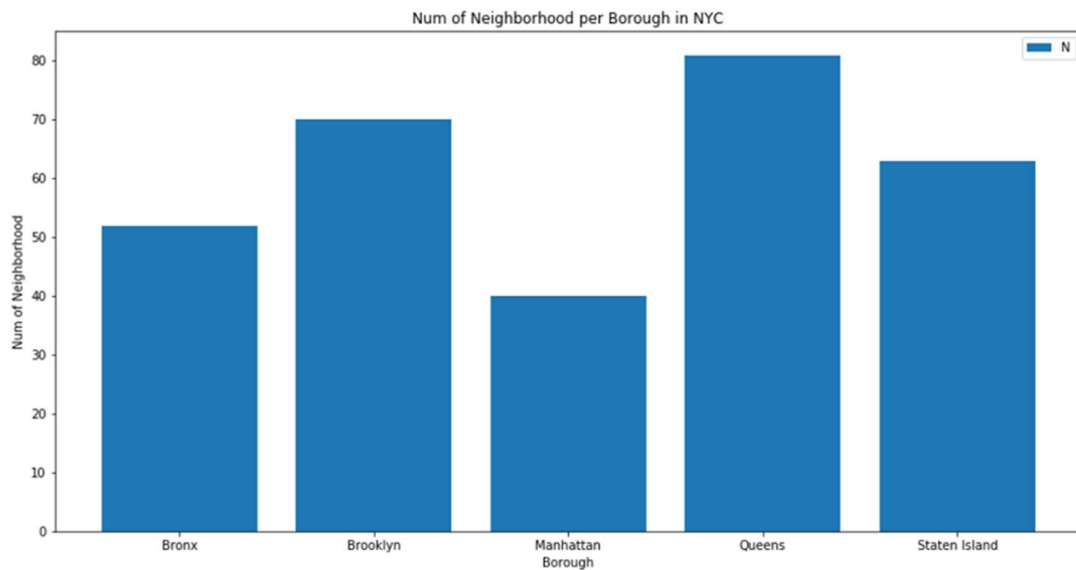
Firstly, we scrape the information from the Wikipedia to get the below dataframe.

	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
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488
13	Bronx	Bedford Park	40.870185	-73.885512
14	Bronx	University Heights	40.855727	-73.910416
15	Bronx	Morris Heights	40.847898	-73.919672
16	Bronx	Fordham	40.860997	-73.896427
17	Bronx	East Tremont	40.842696	-73.887356

After we calculate the sum of neighbourhoods, we found that Queens has the highest number of neighbourhoods.

	Neighborhood	Latitude	Longitude
Borough			
Bronx	52	52	52
Brooklyn	70	70	70
Manhattan	40	40	40
Queens	81	81	81
Staten Island	63	63	63

Below is a bar graph visual.



After that, we interact the Foursquare API to gain more geographic data as below.

	venue.name	venue.categories	venue.location.lat	venue.location.in
0	Lollipops Gelato	[{"id": "4bf58dd8d48988d1d0941735", "name": "D..."}]	40.894123	-73.84589
1	Ripe Kitchen & Bar	[{"id": "4bf58dd8d48988d144941735", "name": "C..."}]	40.898152	-73.83887
2	Ali's Roti Shop	[{"id": "4bf58dd8d48988d144941735", "name": "C..."}]	40.894036	-73.85693
3	Carvel Ice Cream	[{"id": "4bf58dd8d48988d1c9941735", "name": "I..."}]	40.890487	-73.84856
4	Jimbo's	[{"id": "4bf58dd8d48988d16c941735", "name": "B..."}]	40.891740	-73.85822
5	Rite Aid	[{"id": "4bf58dd8d48988d10f951735", "name": "P..."}]	40.896649	-73.84484
6	Jackie's West Indian Bakery	[{"id": "4bf58dd8d48988d144941735", "name": "C..."}]	40.889283	-73.84331
7	Walgreens	[{"id": "4bf58dd8d48988d10f951735", "name": "P..."}]	40.896687	-73.84485
8	Dunkin'	[{"id": "4bf58dd8d48988d148941735", "name": "D..."}]	40.890459	-73.84908
9	Rite Aid	[{"id": "4bf58dd8d48988d10f951735", "name": "P..."}]	40.889062	-73.84299
10	Cooler Runnings Jamaican Restaurant Inc	[{"id": "4bf58dd8d48988d144941735", "name": "C..."}]	40.898083	-73.85025
11	SUBWAY	[{"id": "4bf58dd8d48988d1c5941735", "name": "S..."}]	40.890468	-73.84915

After cleaning the data for a little bit and grouping data. We can now see how much pharmacy shops are within 1000 m of each neighbourhood.

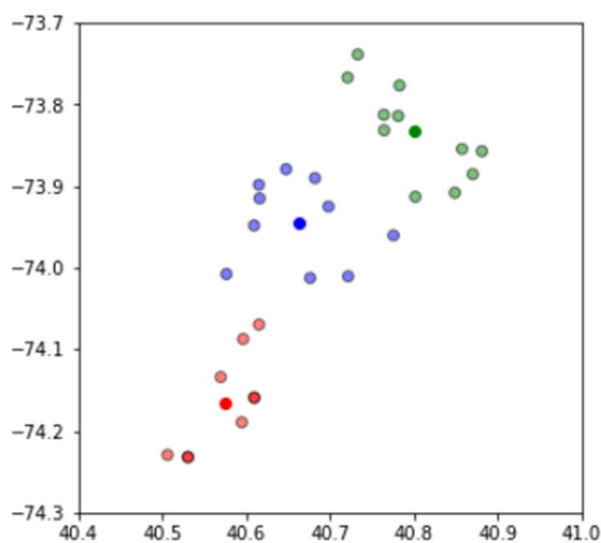
We took a random sample of 30 and found that Bronx has the highest number of pharmacy shops.

	Latitude	Longitude	Pharmacy No
Borough			
Bronx	204.259143	-369.419235	12
Brooklyn	325.120834	-591.477731	12
Manhattan	81.497161	-147.971191	0
Queens	244.546685	-442.741574	6
Staten Island	365.162294	-667.493572	8

We can also distribute the sample on the map.



By using Kmeans clustering, we can cluster the 30 neighbourhood into below scatter plot.



5. Conclusion

In the random sample, we found that Bronx and Brooklyn have the highest number of pharmacy shops. However, in the scatter plot, Bronx and Brooklyn belong to two different clusters. As Brooklyn has a higher number of neighbourhoods, we recommend that investors may prefer to open a pharmacy shop in Brooklyn.