Introduction:-

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 global hub of business and commerce. The city is a major centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theatre, fashion and the arts in United States.

For this Capstone Week 5 project, I am creating a scenario which can help New York government to open new 'Pharmacy Stores' across neighbourhoods of New York City. The idea behind this project is that there may not be enough Pharmacy stores and it might present a great help to the Government and Health department.

As we all know about Covid-19 which is spreading at a higher rate across the world. There are total 358K confirmed cases as of now accompanied by 23,195 deaths in New York City. In future there can be possibility of another epidemic like Covid-19 so we need to be prepared for such situations. If there are enough pharmacy stores across all neighbourhoods of New York City then it would be of great help to the citizens.

My purpose in mind is of finding the locations across New York which lack in Pharmacy stores.

Business Problem:-

The objective of this capstone project is to find all the suitable locations where Pharmacy stores can be built. By using all the skills and techniques of machine learning and data science, this project aims at providing solutions to following business question: In which locations should Government and Health department focus in opening pharmacy stores?

As there are 5 Boroughs in New York City namely Bronx, Brooklyn, Manhattan, Queens and Staten Island, I am going to find all the suitable locations across all boroughs one by one in my project.

After this project we will be able to list out all the neighbourhoods of New York City which lack in Pharmacy stores in two groups:

- Neighbourhoods of Brooklyn and Manhattan where there is shortage of Pharmacy Stores.
- Neighbourhoods of Bronx, Queens and Staten Island where there is shortage of Pharmacy Stores.

Target Audience:

The Government and Health Department of New York City who wants to find all the possible locations where there is shortage of Pharmacy stores.

Data Section:

- City to be analysed in this project: New York

- The New York City has 5 boroughs and 306 neighbourhoods. In order to segment and explore them, we will essentially need a dataset contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood. The following dataset contains all this information:

  https://cocl.us/new_york_dataset

- By using Foursquare API we will get all the information about pharmacy stores in each neighbourhood of New York. By using this API we will get all the venues of neighbourhood then we have to filter these venues to get only Pharmacy stores.

Methodology:

Business Understanding:

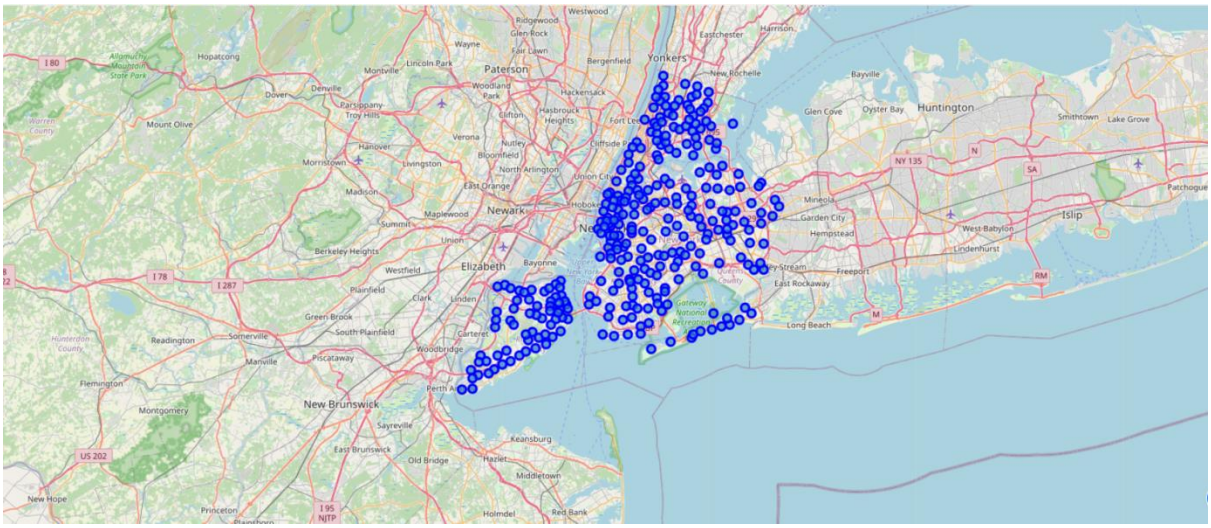
Our main goal is to get all the possible neighbourhoods of New York City where Pharmacy stores can be built for the convenience of citizens.

Analytic Approach:

New York City neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In this project first part is clustering of Manhattan and Brooklyn and find out the suitable neighbourhoods for our project and second part is clustering of Bronx, Queens and Staten Island and listing out the possible neighbourhoods of this part.

Exploratory Data Analysis:

- New York City Data
  In this we load and explore the data from newyork_data.json file. By scrapping the data we transform the required data of nested python dictionaries into a pandas dataframe. This data contains the geographical coordinates of New York city neighbourhoods. We used geopy and folium libraries to create a map of New York City.



In [9]:
```
neighborhoods.head(15)
```

Out[9]:

	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

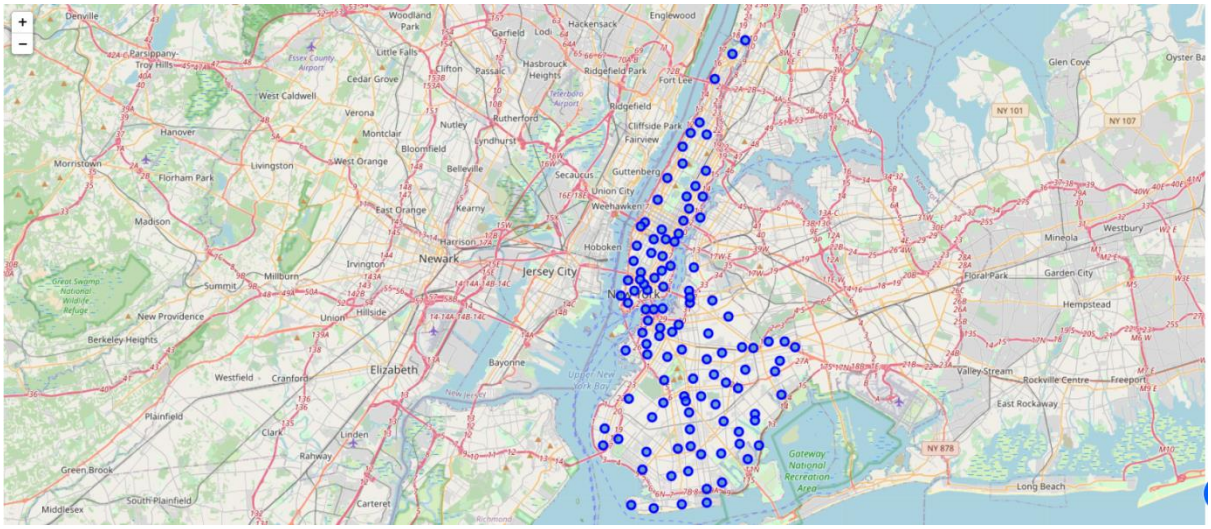
- Next we separate the neighbourhoods and their data of Brooklyn and Manhattan from the original data to perform clustering and listing the possible neighbourhoods for our project.

```
bm_data = neighborhoods.loc[(neighborhoods['Borough'] == 'Brooklyn')|(neighborhoods['Borough'] == 'Manhattan')]
bm_data = bm_data.reset_index(drop=True)
bm_data.head(15)
```

]:

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Brooklyn	Bay Ridge	40.625801	-74.030621
2	Brooklyn	Bensonhurst	40.611009	-73.995180
3	Brooklyn	Sunset Park	40.645103	-74.010316
4	Brooklyn	Greenpoint	40.730201	-73.954241
5	Brooklyn	Gravesend	40.595260	-73.973471
6	Brooklyn	Brighton Beach	40.576825	-73.965094
7	Brooklyn	Sheepshead Bay	40.586890	-73.943186
8	Brooklyn	Manhattan Terrace	40.614433	-73.957438
9	Brooklyn	Flatbush	40.636326	-73.958401
10	Brooklyn	Crown Heights	40.670829	-73.943291
11	Brooklyn	East Flatbush	40.641718	-73.936103
12	Brooklyn	Kensington	40.642382	-73.980421
13	Brooklyn	Windsor Terrace	40.656946	-73.980073
14	Brooklyn	Prospect Heights	40.676822	-73.964859

The following map shows the visualization of neighbourhoods of Brooklyn and Manhattan.



- Then using foursquare API we list out all the venues and category of venues of neighbourhoods of Brooklyn and Manhattan.

```
[25]: bm_venues.head(20)
```

Out[25]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
5	Marble Hill	40.876551	-73.91066	El Malecon	40.879338	-73.904457	Caribbean Restaurant
6	Marble Hill	40.876551	-73.91066	Baker Athletic Complex	40.872061	-73.914876	Athletics & Sports
7	Marble Hill	40.876551	-73.91066	The Bronx Public	40.878377	-73.903481	Pub
8	Marble Hill	40.876551	-73.91066	Estrellita Poblana V	40.879687	-73.906257	Mexican Restaurant

- Next we filter out the data to list to list out the details of Pharmacy stores with onehot encoding and apply the Cluster Labels to all the neighbourhoods.
  To find out the optimal number of 'k' for clustering we use Silhouette coefficient method.
  We get k = 9 as the optimal value.

	Neighborhood	Pharmacy	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bath Beach	0.010417	3	40.599519	-73.998752	Bay Parkway Water Front	40.595941	-74.000917	Surf Spot
0	Bath Beach	0.010417	3	40.599519	-73.998752	Bensonhurst Park	40.597065	-73.998340	Park
0	Bath Beach	0.010417	3	40.599519	-73.998752	Caesar's Bay Shopping Center	40.594946	-73.999706	Shopping Plaza
0	Bath Beach	0.010417	3	40.599519	-73.998752	Five Guys	40.595236	-74.000225	Burger Joint
0	Bath Beach	0.010417	3	40.599519	-73.998752	Starbucks	40.595227	-74.000017	Coffee Shop
0	Bath Beach	0.010417	3	40.599519	-73.998752	King's Kitchen	40.603844	-73.996960	Cantonese Restaurant
0	Bath Beach	0.010417	3	40.599519	-73.998752	Grotta Azzurra	40.603611	-73.995381	Pizza Place
0	Bath Beach	0.010417	3	40.599519	-73.998752	Delacqua	40.604216	-73.997452	Spa
0	Bath Beach	0.010417	3	40.599519	-73.998752	Istanbul Turkish Fast Food & Restaurant	40.601771	-73.993856	Turkish Restaurant
0	Bath Beach	0.010417	3	40.599519	-73.998752	German Chocolate Cake	40.596284	-73.997543	German Restaurant
0	Bath Beach	0.010417	3	40.599519	-73.998752	Ichi Sushi	40.601774	-73.993869	Sushi Restaurant
0	Bath Beach	0.010417	3	40.599519	-73.998752	Vivi Bubble Tea	40.602312	-73.994312	Bubble Tea Shop
0	Bath Beach	0.010417	3	40.599519	-73.998752	Carvel	40.598733	-73.997670	Ice Cream Shop
0	Bath Beach	0.010417	3	40.599519	-73.998752	Lutzina Bar&Lounge	40.600807	-74.000578	Hookah Bar
0	Bath Beach	0.010417	3	40.599519	-73.998752	Cherry Hill Gourmet Market	40.600723	-73.991912	Gourmet Shop
0	Bath Beach	0.010417	3	40.599519	-73.998752	Waterwalk along Bath Beach	40.597457	-74.004417	Beach

In the following map we can see the different types of clusters created by using k-means for Brooklyn and Manhattan.

- Through clustering we get that cluster – 0 contains the list of all neighbourhoods where there is shortage of Pharmacy Stores.

Cluster 0

```
#Cluster 0
cluster0 = bm_merged.loc[bm_merged['Cluster Labels'] == 0]
cluster0
```

]:

	Neighborhood	Pharmacy	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
1	Battery Park City	0.0	0	40.711932	-74.016869	Battery Park City Esplanade	40.711622	-74.017907
1	Battery Park City	0.0	0	40.711932	-74.016869	Institute of Culinary Education	40.712399	-74.015971
1	Battery Park City	0.0	0	40.711932	-74.016869	Hudson Eats	40.712666	-74.015901
1	Battery Park City	0.0	0	40.711932	-74.016869	Waterfront Plaza, Brookfield Place	40.713241	-74.016241
1	Battery Park City	0.0	0	40.711932	-74.016869	Equinox Brookfield Place	40.712704	-74.014995
1	Battery Park City	0.0	0	40.711932	-74.016869	Brookfield Place (BFPL)	40.713240	-74.015193

The neighbourhoods are:
Battery Park City, Bedford Stuyvesant, Boerum Hill, Bergen Beach, Broadway Junction, Brooklyn Heights, Bushwick, Carnegie Hill, Carroll Gardens, Central Harlem, Chelsea, Chinatown, Civic Center, Clinton, Clinton Hill, Cobble Hill, Cypress Hills, Ditmas Park, Downtown, Dumbo, East Village, East Williamsburg, Financial District, Fort Greene, Fulton Ferry, Gowanus, Gramercy, Greenpoint, Greenwich Village, Hamilton Heights, Hudson Yards, Lenox Hill, Lincoln Square, Little Italy, Lower East Side, Manhattan Beach, Mill Island, Murray Hill, Noho, North Side, Paerdegat Basin, Park Slope, Prospect Heights, Prospect Park South, Red Hook, Soho, South Side, Stuyvesant Town, Tribeca, Tudor City, Upper East Side, Vinegar Hill, Washington Heights, West Village, Williamsburg, Windsor Terrace, Yorkville.

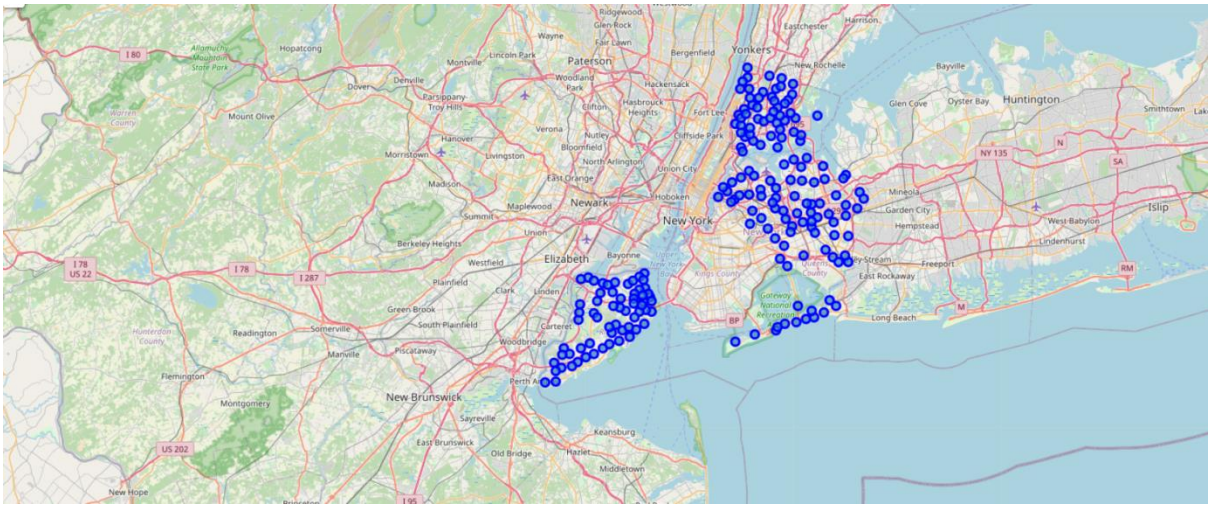
- Now we begin our second part i.e. for Bronx, Queens and Staten Island.
  We separate the neighbourhoods and their data of Bronx, Queens and Staten Island from the original data to perform clustering and listing the possible neighbourhoods for our project.

```
bqs_data = neighborhoods.loc[(neighborhoods['Borough'] == 'Bronx')|(neighborhoods['Borough'] == 'Que
bqs_data = bqs_data.reset_index(drop=True)
bqs_data.head(30)
```

3]:

	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	Bronx	Woodlawn	40.898273	-73.867315
7	Bronx	Norwood	40.877224	-73.879391
8	Bronx	Williamsbridge	40.881039	-73.857446
9	Bronx	Baychester	40.866858	-73.835798
10	Bronx	Pelham Parkway	40.857413	-73.854756
11	Bronx	City Island	40.847247	-73.786488
12	Bronx	Bedford Park	40.870185	-73.885512
13	Bronx	University Heights	40.855727	-73.910416
14	Bronx	Morris Heights	40.847898	-73.919672
15	Bronx	Fordham	40.860997	-73.896427
16	Bronx	East Tremont	40.842696	-73.887356

The following map gives a visualization of neighbourhoods of Bronx, Queens and Staten Island.



- Then using foursquare API we list out all the venues and category of venues of neighbourhoods of Bronx, Queens and Staten Island.

```
bqs_venues.head(25)
```

]:

	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	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
5	Wakefield	40.894705	-73.847201	Jimbo's	40.891740	-73.858226	Burger Joint
6	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
7	Wakefield	40.894705	-73.847201	Rite Aid	40.889062	-73.842993	Pharmacy
8	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
9	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
10	Wakefield	40.894705	-73.847201	Subway	40.890468	-73.849152	Sandwich Place
11	Wakefield	40.894705	-73.847201	Shell	40.894187	-73.845862	Gas Station
12	Wakefield	40.894705	-73.847201	E&L Bakery	40.893564	-73.856997	Bakery

- Next we filter out the data to list to list out the details of Pharmacy stores with onehot encoding and apply the Cluster Labels to all the neighbourhoods.
  To find out the optimal number of 'k' for clustering we use Silhouette coefficient method.
  We get k = 9 as optimal value.

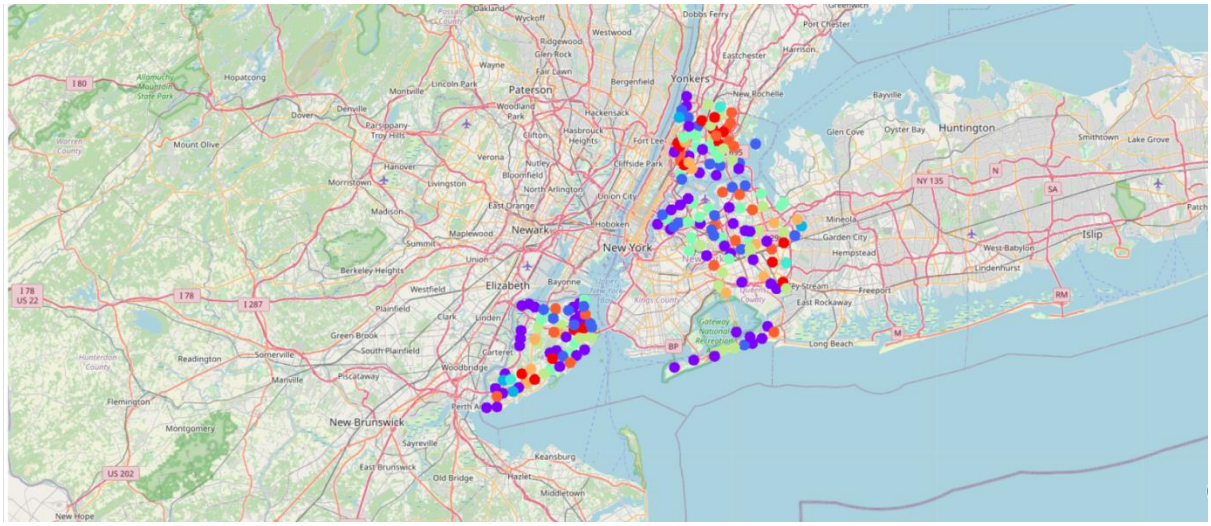
```
print(bqs_merged.shape)
bqs_merged.head(25)
```

```
(10912, 9)
```

']:

	Neighborhood	Pharmacy	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Allerton	0.045455	0	40.865788	-73.859319	White Castle	40.866065	-73.862307	Fast Food Restaurant
0	Allerton	0.045455	0	40.865788	-73.859319	Domenick's Pizzeria	40.865576	-73.858124	Pizza Place
0	Allerton	0.045455	0	40.865788	-73.859319	Bronx Martial Arts Academy	40.865721	-73.857529	Martial Arts Dojo
0	Allerton	0.045455	0	40.865788	-73.859319	Sal & Doms Bakery	40.865377	-73.855236	Dessert Shop
0	Allerton	0.045455	0	40.865788	-73.859319	Dunkin'	40.865204	-73.859007	Donut Shop
0	Allerton	0.045455	0	40.865788	-73.859319	IHOP	40.865728	-73.862460	Breakfast Spot
0	Allerton	0.045455	0	40.865788	-73.859319	La Estrellita Poblana	40.867077	-73.867595	Mexican Restaurant
0	Allerton	0.045455	0	40.865788	-73.859319	Gun Hill Brewing Co.	40.872139	-73.855698	Brewery

In the following map we can see the different types of clusters created by using k-means for Bronx, Queens and Staten Island.



- Through clustering we get that cluster – 1 contains the list of all neighbourhoods where there is shortage of Pharmacy Stores.

Cluster 1

```
0]: #Cluster 1
    cluster1 = bqs_merged.loc[bqs_merged['Cluster Labels'] == 1]
    cluster1
```

t[80]:

	Neighborhood	Pharmacy	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	
3	Arlington	0.0	1	40.635325	-74.165104	Kohl's	40.626935	-74.164104	De
3	Arlington	0.0	1	40.635325	-74.165104	Comic Book Jones	40.626547	-74.163384	Co
3	Arlington	0.0	1	40.635325	-74.165104	The Home Depot	40.628964	-74.172358	
3	Arlington	0.0	1	40.635325	-74.165104	7-Eleven	40.626979	-74.165870	Cor
3	Arlington	0.0	1	40.635325	-74.165104	Lowe's	40.627646	-74.162136	
3	Arlington	0.0	1	40.635325	-74.165104	P And E Custom Kitchen Inc	40.634743	-74.165784	Hom

The neighbourhoods are:
Arlington, Arverne, Astoria, Blissville, Bloomfield, Broad Channel, Brookville, Buffer Manor, Charleston, Claremont Village, Egbertville, Elmhurst, Fieldston, Flushing, Forest Hills Garden, Fox Hills, Glendale, Grymes Hills, High Bridge, Holliswood, Howland Hook, Huguenot, Hunters Point, Hunts Point, Kew Gardens, Kingsbridge Heights, Lighthouse Hill, Midland Beach, Neoponsit, New Brighton, North Riverdale, Oakwood, Pleasant Plains, Pomonok, Port Ivory, Queensboro Hill, Randall Manor, Ravenswood, Rossville, Roxbury, Silver Lake, Somerville, South Beach, South Jamaica, South Ozone Park, Springfield Gardens, Steinway, Throgs Neck, Todt Hill, Tottenville, Travis, Utopia, West Farms.

Limitations and Suggestions for future research:

In this project, I only take into consideration of one factor: the occurrence / existence of Pharmacy Stores in each neighbourhood. There are many factors that can be taken into consideration such as population density, income of residents, rent that could influence the decision to open new Pharmacy Stores. However, to put all these data into this project is not possible for this capstone project. Future research can take into consideration of these factors.

Conclusion:

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. In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the Government.