Battle of Neighborhoods for a Pharmacy Brooklyn

Sebin Thomas

14JAN2021

Introduction: Business Problem

During these unprecedented times with Pandemic. Medical supplies are utmost required in most populated places in New York City. In this project we are trying to find the optimal location for opening a Pharmacy in the borough of Brooklyn.

Since the borough has a lot of pharmacies already, we need to find a location with less density of pharmacies in a neighborhood. We will use data science techniques to generate the most promising neighborhoods. This will give the stakeholders boroughs with a smaller number of pharmacies so that they can shortlist and review other characteristics

Data

Based on definition of our problem, factors that will influence our decision will be:

- Number of pharmacies in the neighborhood.

Following data sources will be needed to extract/generate the required information:

- Extract NYC Neighborhoods from Wiki
  : https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City
- centers of candidate areas will be generated algorithmically.
- Number of Pharmacies and their type and location in every neighborhood will be obtained using **Foursquare API**

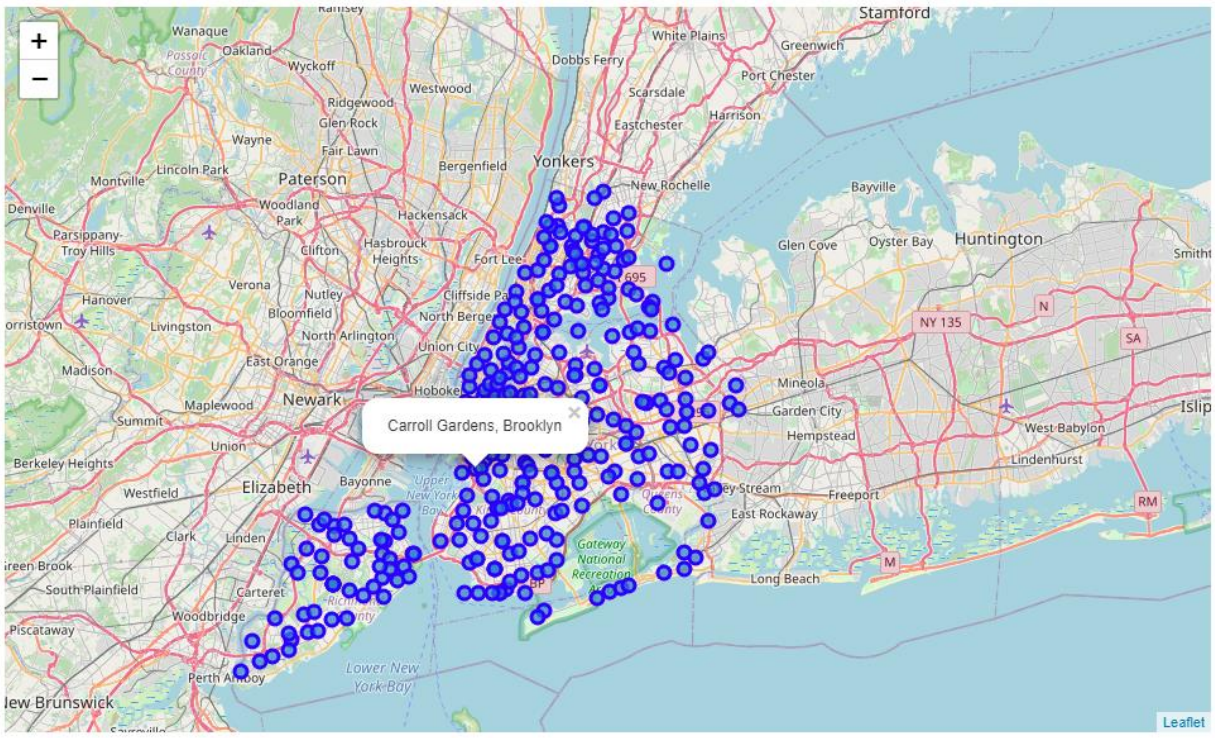
Data Cleansing and Enrichment

After extracting data about neighborhoods from Wikipedia. lets enrich with latitude and longitude.

	Borough	CommunityBoard	Area	Census	Neighborhood	Latitude	Longitude
0						43.1562	-75.845
1	Bronx	Bronx CB 1	7.17	91,497	Melrose	40.8257	-73.9152
2	Bronx	Bronx CB 1	7.17	91,497	Mott Haven	40.809	-73.9229
3	Bronx	Bronx CB 1	7.17	91,497	Port Morris	40.8015	-73.9096
4	Bronx	Bronx CB 2	5.54	52,246	Hunts Point	40.8126	-73.884
...	...	...	...	...	...	...	...
326	Staten Island	Staten Island CB 3	58.97	152,908	Prince's Bay	40.529	-74.1976
327	Staten Island	Staten Island CB 3	58.97	152,908	Richmond Valley	40.5201	-74.2293
328	Staten Island	Staten Island CB 3	58.97	152,908	Rossville	40.5556	-74.2129
329	Staten Island	Staten Island CB 3	58.97	152,908	Tottenville	40.5112	-74.2493
330	Staten Island	Staten Island CB 3	58.97	152,908	Woodrow	40.5434	-74.1976

331 rows × 7 columns

Visualize the dataset



Methodology

In this project we will direct our efforts on detecting areas of Brooklyn that have low density of pharmacies, particularly in the borough of Brooklyn.

Once we do the exploratory analysis and then we will create k-means clustering algorithm to group it into clusters for further analysis.

Analysis

Since the area we interested is Brooklyn. We filter the data set.

	Borough	CommunityBoard	Area	Census	Neighborhood	Latitude	Longitude
61	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.7237	-73.951
62	Brooklyn	Brooklyn CB 1	12.82	160,338	Williamsburg	40.7146	-73.9535
63	Brooklyn	Brooklyn CB 1	12.82	160,338	Williamsburg Houses	40.7096	-73.9419
64	Brooklyn	Brooklyn CB 2	7.72	98,620	Boerum Hill	40.6856	-73.9842
65	Brooklyn	Brooklyn CB 2	7.72	98,620	Brooklyn Heights	40.6961	-73.995

Foursquare

Now that we have our location candidates, let's use Foursquare API to get info on pharmacies in each neighborhood.

:

	Borough	CommunityBoard	Area	Census	Neighborhood	Latitude	Longitude	Pharmacy_Name	Pharma_Latitude	Pharma_Longitude
0	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.723713	-73.950971	JAG-ONE Physical Therapy	40.725820	-73.951291
1	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.723713	-73.950971	Polska Przychodnia	40.724210	-73.949222
2	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.723713	-73.950971	Greenpoint Footcare	40.725489	-73.950907
3	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.723713	-73.950971	Tribeca Pediatrics - Greenpoint	40.722475	-73.949241
4	Brooklyn	Brooklyn CB 1	12.82	160,338	Greenpoint	40.723713	-73.950971	The Smilist Dental	40.724378	-73.948829
...	...	...	...	...	...	...	...	...	...	...
577	Brooklyn	Brooklyn CB 18	24.68	194,653	Mill Island	40.650104	-73.949582	Diamond Braces Church Ave	40.650555	-73.949753
578	Brooklyn	Brooklyn CB 18	24.68	194,653	Mill Island	40.650104	-73.949582	Advanced Dental Center	40.650555	-73.949753
579	Brooklyn	Brooklyn CB 18	24.68	194,653	Mill Island	40.650104	-73.949582	Diamond Braces	40.650555	-73.949753
580	Brooklyn	Brooklyn CB 18	24.68	194,653	Mill Island	40.650104	-73.949582	Advanced Oral Surgery of Brooklyn, PLLC	40.650426	-73.949818
581	Brooklyn	Brooklyn CB 18	24.68	194,653	Mill Island	40.650104	-73.949582	Church Ave Orthadontics	40.650444	-73.949796

582 rows × 10 columns

Explore the data set to find out the nearby pharmacies

.

:

Neighborhood	Borough	CommunityBoard	Area	Census	Latitude	Longitude	Pharmacy_Name	Pharma_Latitude	Pharma_Longitude
Bensonhurst	10	10	10	10	10	10	10	10	10
Brighton Beach	20	20	20	20	20	20	20	20	20
Brooklyn Heights	21	21	21	21	21	21	21	21	21
Canarsie	12	12	12	12	12	12	12	12	12
Clinton Hill	6	6	6	6	6	6	6	6	6
Cobble Hill	9	9	9	9	9	9	9	9	9
Coney Island	2	2	2	2	2	2	2	2	2
Cypress Hills	1	1	1	1	1	1	1	1	1
Dumbo	5	5	5	5	5	5	5	5	5
East Flatbush	4	4	4	4	4	4	4	4	4

Cluster the dataset

Looking good. What we have now is a clear indication of zones with low number of pharmacies in vicinity. Let us now cluster those locations to create centers of zones containing good locations.

Neighborhood	Cluster Labels	Borough	CommunityBoard	Area	Census	Latitude	Longitude	Pharmacy_Name	Pharma_Latitude	Pharma_Longitude
Bensonhurst	8	10	10	10	10	10	10	10	10	10
Brighton Beach	2	20	20	20	20	20	20	20	20	20
Brooklyn Heights	2	21	21	21	21	21	21	21	21	21
Canarsie	4	12	12	12	12	12	12	12	12	12
Clinton Hill	0	6	6	6	6	6	6	6	6	6
Cobble Hill	8	9	9	9	9	9	9	9	9	9
Coney Island	7	2	2	2	2	2	2	2	2	2
Cypress Hills	7	1	1	1	1	1	1	1	1	1
Dumbo	0	5	5	5	5	5	5	5	5	5
East Flatbush	5	4	4	4	4	4	4	4	4	4
Erasmus	0	7	7	7	7	7	7	7	7	7
Flatlands	4	13	13	13	13	13	13	13	13	13
Fort Greene	2	20	20	20	20	20	20	20	20	20
Fort Hamilton	0	7	7	7	7	7	7	7	7	7
Fulton Ferry	7	1	1	1	1	1	1	1	1	1
Fulton Mall	6	46	46	46	46	46	46	46	46	46
Georgetown	9	17	17	17	17	17	17	17	17	17
Gerritsen Beach	7	1	1	1	1	1	1	1	1	1

Cluster the dataset using the k means

Looking good. What we have now is a clear indication of zones with low number of pharmacies in vicinity. Let us now cluster those locations to create centers of zones containing good locations.

Neighborhood	Cluster Labels	Borough	CommunityBoard	Area	Census	Latitude	Longitude	Pharmacy_Name	Pharma_Latitude	Pharma_Longitude
Coney Island	7	2	2	2	2	2	2	2	2	2
Cypress Hills	7	1	1	1	1	1	1	1	1	1
Fulton Ferry	7	1	1	1	1	1	1	1	1	1
Gerritsen Beach	7	1	1	1	1	1	1	1	1	1
Gowanus	7	1	1	1	1	1	1	1	1	1
Mapleton	7	1	1	1	1	1	1	1	1	1
Remsen Village	7	2	2	2	2	2	2	2	2	2
Rugby	7	1	1	1	1	1	1	1	1	1
Bushwick	7	1	1	1	1	1	1	1	1	1
Carroll Gardens	7	2	2	2	2	2	2	2	2	2
Greenwood Heights	7	2	2	2	2	2	2	2	2	2

After reviewing the clusters filter the data set with least density

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Coney Island	40.543439	-74.197644
1	Brooklyn	Cypress Hills	40.543439	-74.197644
2	Brooklyn	Fulton Ferry	40.543439	-74.197644
3	Brooklyn	Gerritsen Beach	40.543439	-74.197644
4	Brooklyn	Gowanus	40.543439	-74.197644
5	Brooklyn	Mapleton	40.543439	-74.197644
6	Brooklyn	Remsen Village	40.543439	-74.197644
7	Brooklyn	Rugby	40.543439	-74.197644
8	Brooklyn	Bushwick	40.543439	-74.197644
9	Brooklyn	Carroll Gardens	40.543439	-74.197644
10	Brooklyn	Greenwood Heights	40.543439	-74.197644

Results and Discussion

After reviewing the pharmacies across the Brooklyn neighborhoods, the above locations came out as the least number of pharmacies.

This, of course, does not imply that those zones are actually optimal locations for a new Pharmacy!

Purpose of this analysis was to only provide info on areas in Brooklyn where a new pharmacy can be established This will give an initial list of optimal locations to stakeholders where a further deep evaluation will be required to nail down the location considering additional factors.

Conclusion

Purpose of this project was to identify areas in Brooklyn with low number of pharmacies in order to aid stakeholders in narrowing down the search for optimal location. By calculating density distribution from Foursquare data, we have first identified general boroughs that justify further analysis, and then generated extensive collection of locations which satisfy some basic requirements regarding existing nearby pharmacies. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) to be used as starting points for final exploration by stakeholders.

Final decision on optimal location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, zoning, city approvals, proximity to hospitals, real estate availability, prices etc.