Finding Optimal Location to open new Drugstore in Brooklyn

Anchal Sojatiya

30 May 2019

Table of contents

Introduction

Brooklyn is the most populous borough of New York City, with an estimated 2,648,771 residents in 2017. Brooklyn has been coterminous with Kings County, the most populous county in the U.S. state of New York and the second-most densely populated county in the United States. Today, if each borough were ranked as a city, Brooklyn would rank as the third-most populous in the U.S., after Los Angeles and Chicago. In an area so densely populated, drugstore is a very important utility.

Business Problem

In this project we will try to find an optimal location for a Drug Store. Specifically, this report will be targeted to stakeholders interested in opening a Drug Store in New York, USA.

Since there are lots of Drug Stores in New York we will try to detect locations that are not already crowded with Drugstores. We are also particularly interested in areas with no Drug Stores in vicinity. We would also prefer locations as close to city center as possible.

We will use our data science powers to generate a few most promising neighborhoods based on these criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

Data

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

- number of existing Drug Stores in the neighborhood
- distance to Drug Store in the neighborhood, if any
- distance of neighborhood from city center We decided to use regularly spaced grid of locations, centered around city center, to define our neighborhoods.

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

- centers of candidate areas will be generated algorithmically and approximate addresses of centers of those areas will be obtained using Bing Maps API reverse geocoding
- number of Drug Store and their location in every neighborhood will be obtained using Foursquare API
- coordinate of New York center will be obtained using Bing Maps API geocoding of well-known New York location (Brooklyn)

Bing Maps API
Bing Maps Platform (previously Microsoft Virtual Earth) is a geospatial mapping platform produced by Microsoft. It allows developers to create applications that layer location-relevant data on top of licensed map imagery. The imagery includes samples taken by satellite sensors, aerial cameras (including 45 degree oblique "bird's eye" aerial imagery licensed from Pictometry International), Street side imagery, 3D city models and terrain.

Bing Maps Platform also provides a point-of-interest database including a search capability. Microsoft uses the Bing Maps Platform to power its Bing Maps product.

Key features of the Bing Maps Platform include:

- Photo-based images with features such as Street side and 45 degree oblique "bird's eye" views (nominally including 4 views at 90 degree viewpoint increments) that present data in context while simplifying orientation and navigation.
- The ability to overlay standard or custom data points and layers with different themes.
- Building-level geocoding for more than 70 million addresses in the United States.
- Developer support options available.
- Set of APIs available upon which developers can build applications. [Source Wiki]

Using Bing Maps API, coordinate of Brooklyn, New York were found and grid of area candidates, equally spaced, centered around city center and within ~8km from Brooklyn. Neighborhoods were defined as circular areas with a radius of 300 meters, so the neighborhood centers would be 600 meters apart [as in Figure 1].
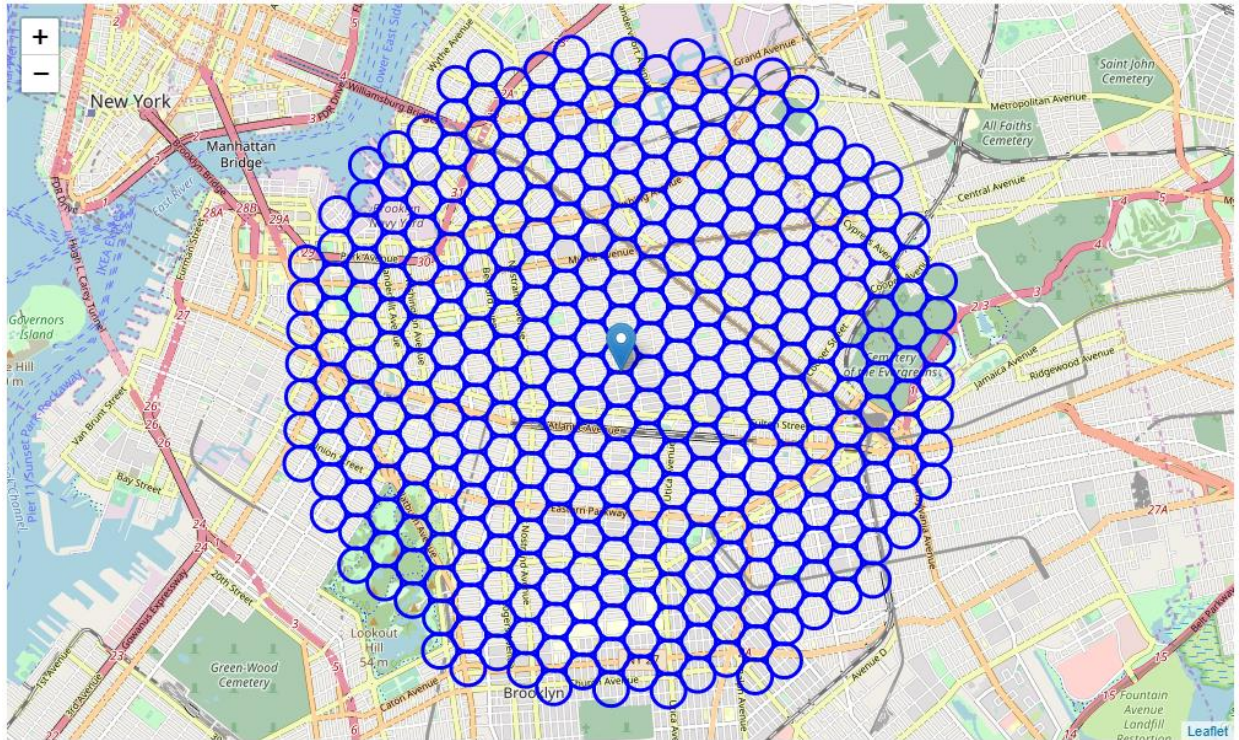

Figure 1

Now, using reverse geocoding, address of center of these circular areas were found and copied into a dataframe [as in Figure 2].

	Latitude	Longitude	X	Y	Distance from center	Address	Zip Code
0	40.672450	-73.893741	-5.828720e+06	9.854911e+06	5992.495307	2131 Pitkin Ave, New York, NY 11207	11207
1	40.675981	-73.893604	-5.828120e+06	9.854911e+06	5840.376700	2722 Atlantic Ave, New York, NY 11207	11207
2	40.679513	-73.893467	-5.827520e+06	9.854911e+06	5747.173218	38 Arlington Ave, New York, NY 11207	11207
3	40.683045	-73.893331	-5.826920e+06	9.854911e+06	5715.767665	34 Bulwer Pl, New York, NY 11207	11207
4	40.686577	-73.893194	-5.826320e+06	9.854911e+06	5747.173218	Jackie Robinson Pkwy, New York, NY 11385	11385
5	40.690110	-73.893057	-5.825720e+06	9.854911e+06	5840.376700	8115 Cypress Ave, New York, NY 11385	11385
6	40.693642	-73.892920	-5.825120e+06	9.854911e+06	5992.495307	8097 59th St, New York, NY 11385	11385
7	40.667243	-73.897962	-5.829620e+06	9.855430e+06	5855.766389	570 Blake Ave, New York, NY 11207	11207
8	40.670774	-73.897825	-5.829020e+06	9.855430e+06	5604.462508	213-259 Georgia Ave, New York, NY 11207	11207
9	40.674306	-73.897689	-5.828420e+06	9.855430e+06	5408.326913	317 Liberty Ave, New York, NY 11207	11207

Figure 2

Foursquare API

Launched in November 2009, Foursquare is a database of more than 105 million places worldwide and an API that enables location data for the world's top companies and more than 150,000 developers rely on Foursquare to power geo-tagging, venue search and more in their apps. [Source Foursquare].

Foursquare API was used to get info on Drug Store in each neighborhood. A total of 209 Drugstores were found in our area of interest. [Figure 3]

```
List of Drugstore
--------------------------
('5502cc22498e1d83355b252b', 'Health Mart Pitkin Pharmacy', 40.67272907587055, -73.89074423000926, '2232 Pitkin Ave, Brooklyn, NY 11207, Unite
d States', 254, True, -5828661.61209925, 9854524.899534753)
('4c3f413de26920a1a4285be7', 'Umamah Pharmacy, Inc.', 40.67350260361088, -73.89100261438382, '2232 Pitkin Ave, Brooklyn, NY 11207, United Stat
es', 259, True, -5828531.2856114805, 9854562.17665745)
('4b606be0f964a520efe429e3', 'Fulton Family Pharmacy', 40.677496, -73.891267, '668 Fulton Street (S Elliot Place), New York, NY, United State
s', 291, True, -5827854.411283976, 9854616.373297282)
('4cacd36ba6e08cfa59b4ae94', 'CVS pharmacy', 40.692585, -73.896353, '63-57 (fresh pond road), Ridgewood, NY, United States', 312, True, -58253
12.874544177, 9855348.935941638)
('4e3c5a85d164b529119e34a9', 'New Lots Care Pharmacy Inc.', 40.66716980448785, -73.89710737118067, '404 New Lots Ave (New Jersey Ave), Brookly
n, NY 11207, United States', 72, True, -5829629.654663829, 9855319.621782755)
('564b443b498ef87929bb09ea', 'Brooklyn Pharmacy', 40.669193, -73.8973, '589 Sutter Ave (Georgia Avenue), Brooklyn, NY 11207, United States', 1
81, True, -5829286.89414519, 9855354.61778675)
('5c9aba58f0b490002c0ddaf5', 'Walgreens', 40.70045740630135, -73.89729081602957, '6036 Myrtle Avenue (Summerfield Street), Ridgewood, NY 1138
5, United States', 290, True, -5823980.630266709, 9855509.297608403)
('589c90afa36ecd0141c936d1', "Nature's Way Pharmacy", 40.700723430946304, -73.89748692512512, '60-27 Myrtle Ave, New York, NY 11385, United St
ates', 270, True, -5823936.237939329, 9855535.946146956)
('4dac8e266e81162ae804b798', 'ROCKAWAY RX, INC', 40.66138744354248, -73.90364527702332, '710 Rockaway Ave (Livonia ave), Brooklyn, NY 11212, U
nited States', 348, True, -5830636.21791031, 9856136.15524829)
('4f1ed127e4b0d72e6b907da6', 'dumont pharmacy', 40.66500074247297, -73.90522952052349, 'Brooklyn, NY 11212, United States', 277, True, -583002
8.676887867, 9856358.907898411)
...
Total: 209
```

Figure 3

All the collected Drugstore in our area of interest were plotted on map [Figure 4].
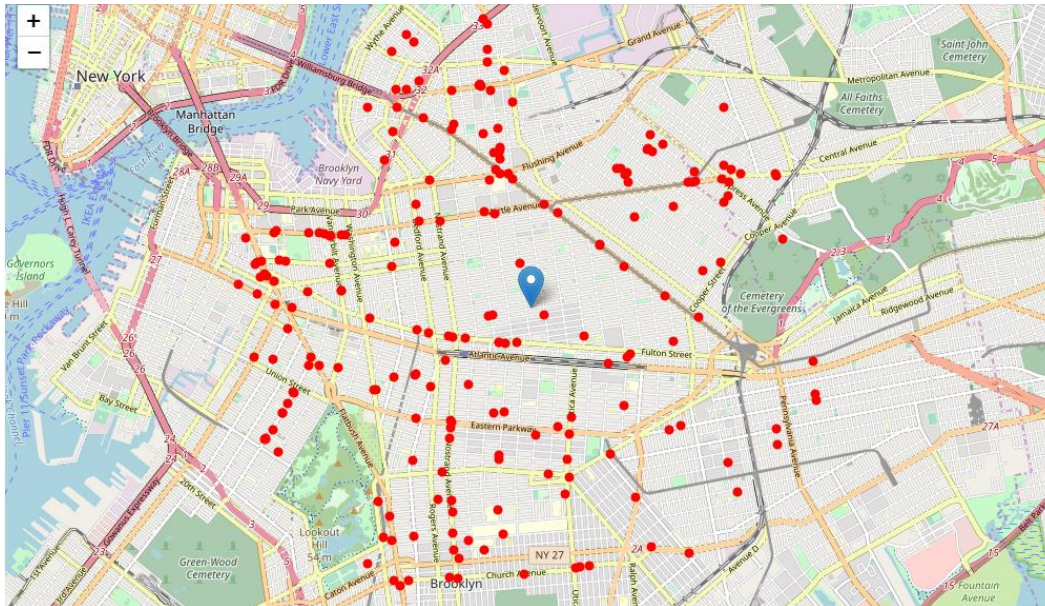

Figure 4

So now we have the entire Drugstores in area within few kilometers from Brooklyn. We also know which drugstores exactly are in vicinity of every neighborhood candidate center.

This concludes the data gathering phase - we're now ready to use this data for analysis to produce the report on optimal locations for a new Drugstore!

Methodology

As our project we will direct our efforts on detecting areas of Brooklyn that have low drugstore density. We will limit our analysis to area ~6km around city center.

In first step we have collected the required data: location of every drugstore within 6km from Brooklyn center (according to Foursquare categorization).

Second step in our analysis will be calculation and exploration of 'drugstore density' across different areas of Brooklyn - we will use heat maps to identify a few promising areas close to center with low number of drugstores and focus our attention on those areas.

In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration locations with no more than two drugstores in radius of 400 meters. We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Analysis

We performed some basic explanatory data analysis and derive some additional info from our raw data. And count the number of drugstores in every area candidate and calculate the distance to nearest Drugstore from every area candidate center (not only those within 300m - we want distance to closest one, regardless of how distant it is).

It was found that Average distance to closest drugstore from each area center was: 497.67274709 53074. So on average Drugstore was found within ~500m from every area center candidate. A m ap showing heatmap / density of Drugstores was created to extract some meaningful info from th em. [Figure 5]
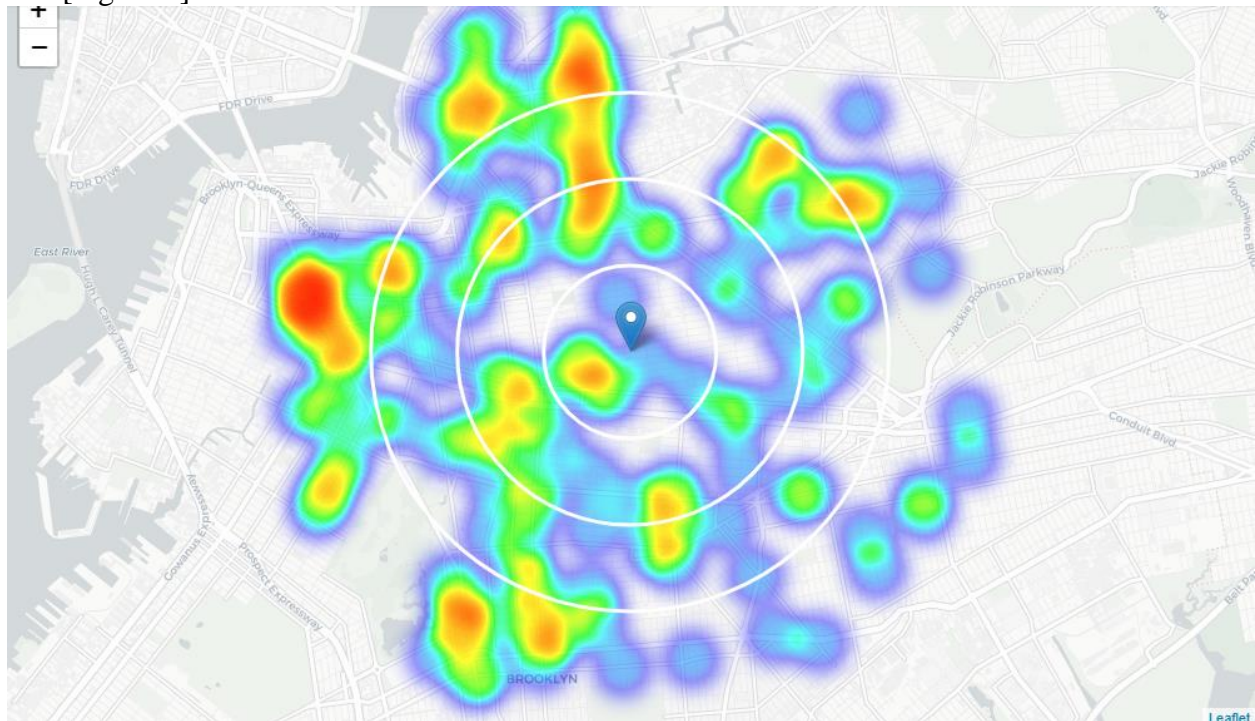

Figure 5

This map indicates higher density of existing Drugstore in west, north and north-east from Brooklyn, with closest pockets of low Drugstore density positioned center, east and south-east from city center.

Based on this our further analysis was based on areas center, east and south-east from Brooklyn center. This placed our location candidates mostly in Bedford-Stuyvesant, Crowns Height and Brownsville.
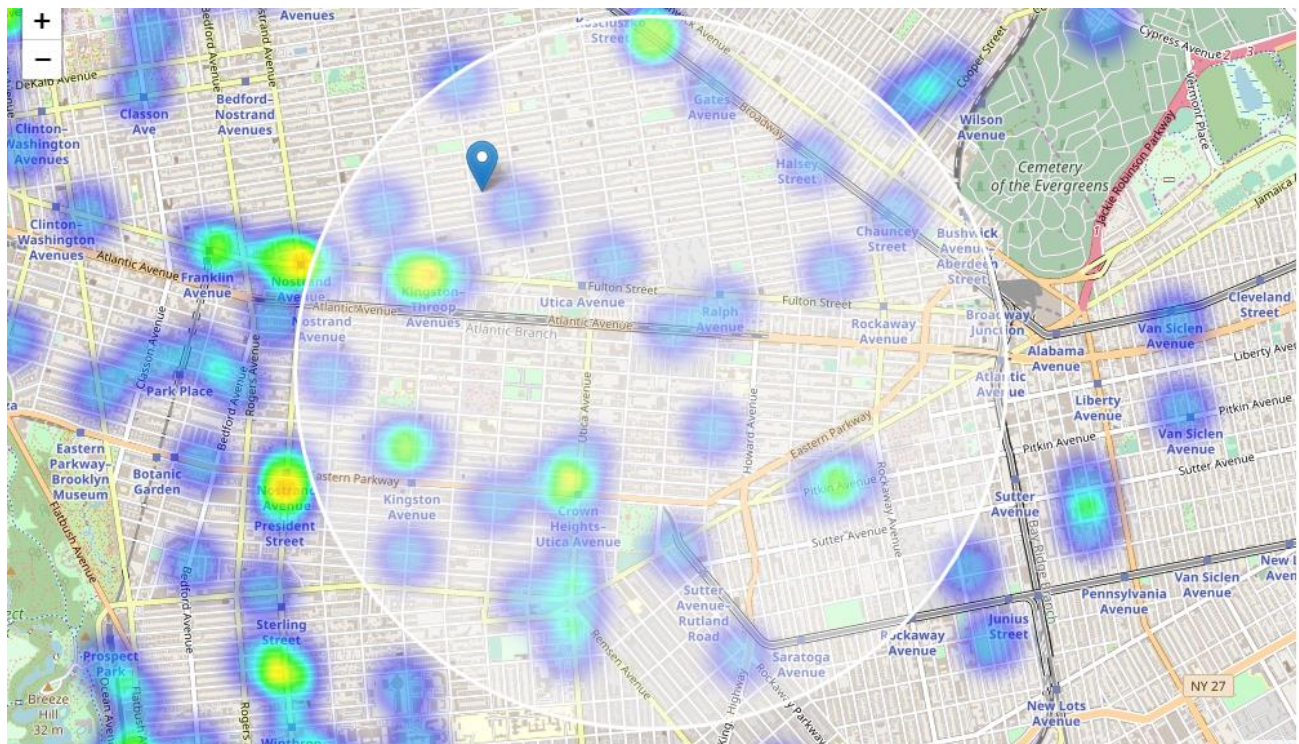
So circle was shifted to our desired location [Figure 6]

Figure 6

This nicely covers all the pockets of low drugstore density in Bedford-Stuyvesant, Crown Heights and Brownsville closest to Brooklyn center.

A denser grid of location candidates restricted to our new region of interest was created with location candidates 300m apart. I was found that 252 candidate neighborhood centers were generated
.
Further number of drugstore in vicinity (radius of 500 meters was used) and distance to closest Drugstore was calculated [Figure 7].

Latitude	Longitude	X	Y	Drugstores nearby	Distance to Drugstore
40.680892	-73.942882	-5.827470e+06	9.861303e+06	6	177.344747
40.672989	-73.945184	-5.828820e+06	9.861563e+06	2	387.449114
40.674754	-73.945119	-5.828520e+06	9.861563e+06	1	327.993537
40.676520	-73.945054	-5.828220e+06	9.861563e+06	1	457.009863
40.678286	-73.944989	-5.827920e+06	9.861563e+06	2	402.052419

Figure 7

These locations were filtered with no more than two drugstores in radius of 400 meters. It was fo und that. Locations with no more than two drugstores nearby was 230, Locations with no Drugst ore within 400m was 138and Locations with both conditions met were 135.
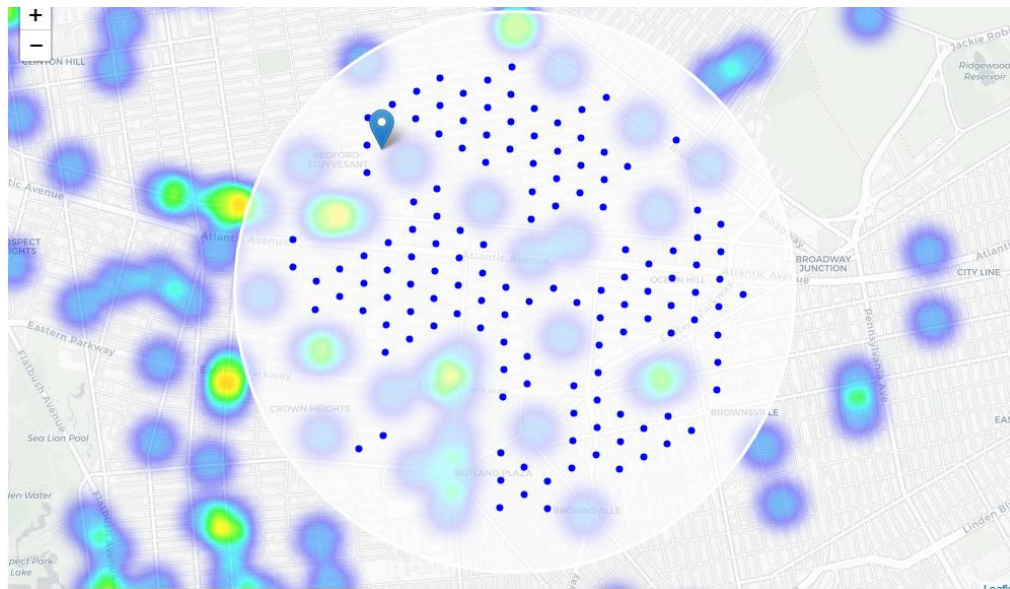

Figure 8

All these location were plot on map [Figure 8]. Any of those locations would be a potential candidate for a new Drugstore, at least based on nearby competition.
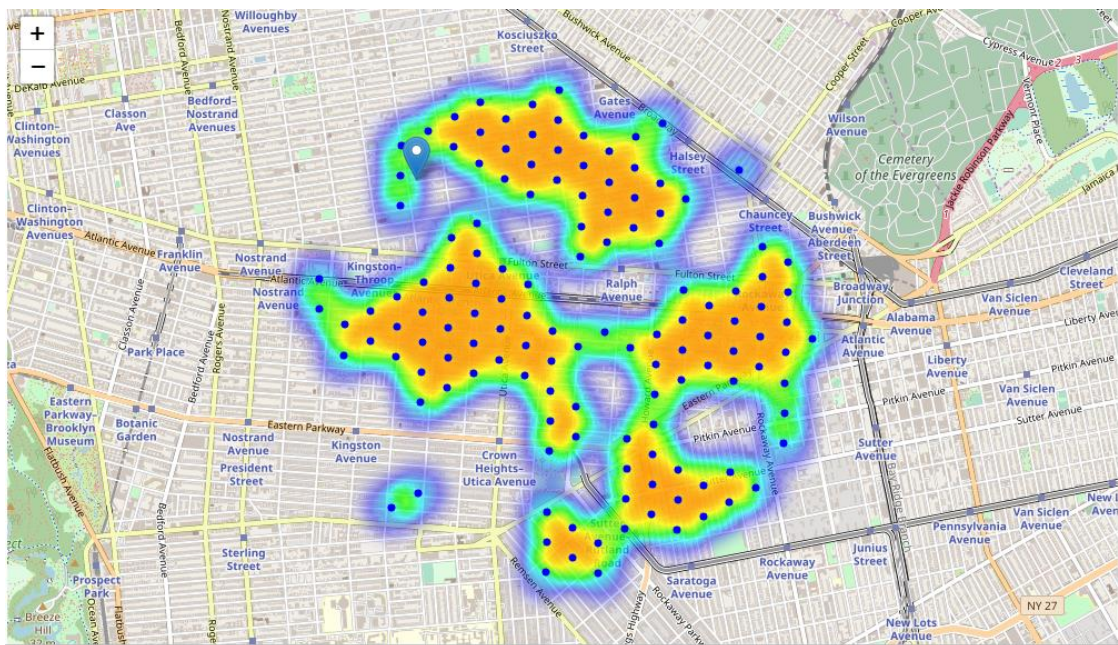
Then these locations were plotted on heat map [Figure 9]

Figure 9

What we have now is a clear indication of zones with no Drugstores nearby.

Let us now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis. We will use k means clustering for this.

K-means clustering

K-means clustering is a method of vector quantization that is popular for cluster analysis. K-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

To determine optimum value of k we use k-mean elbow method

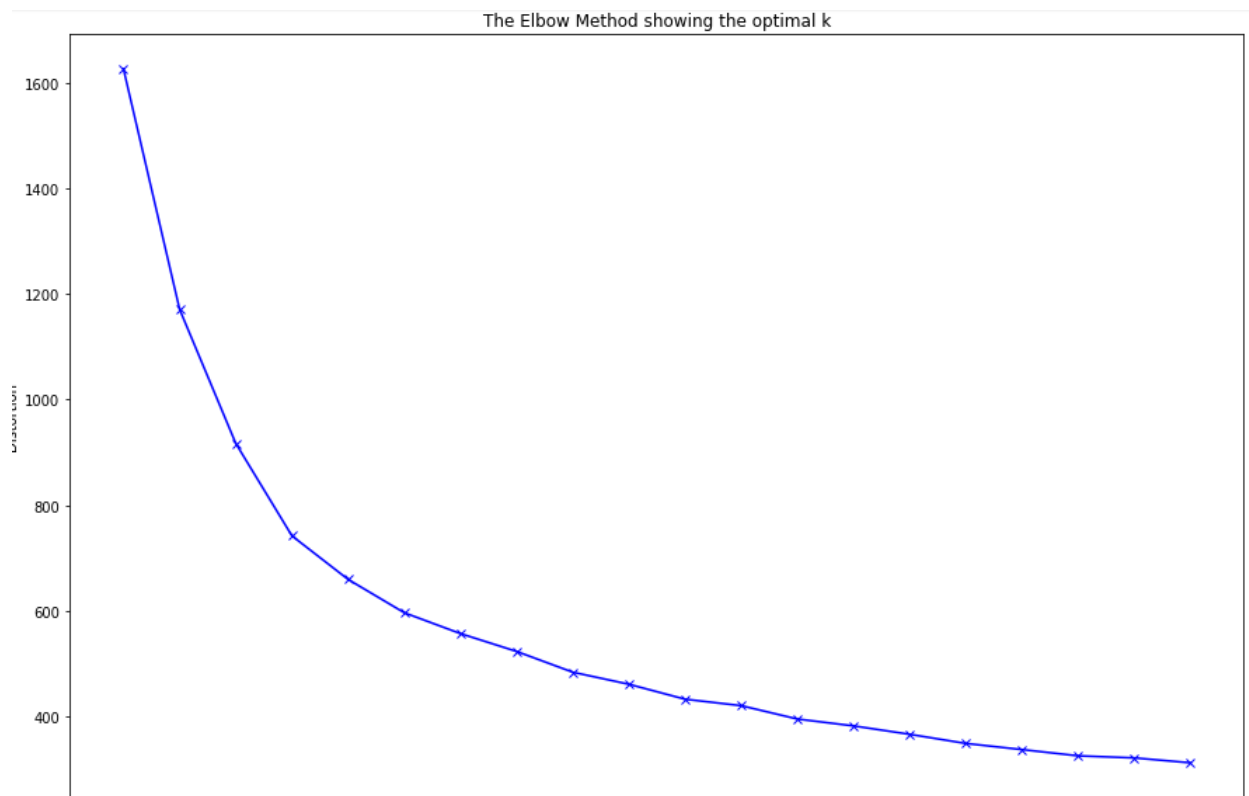From figure 10 it is clear that best value of k is 9
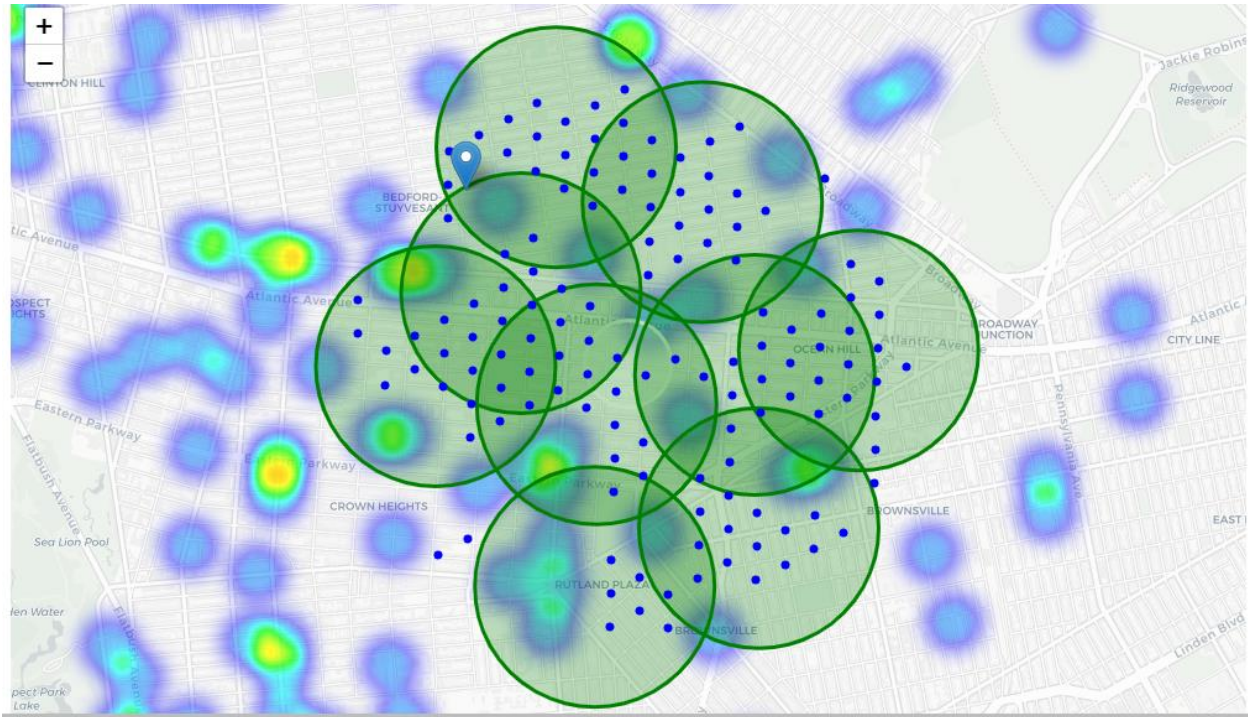

Figure 10

Figure 11

The clusters represent groupings of most of the candidate locations and cluster centers are placed nicely in the middle of the zones 'rich' with location candidates.[Figure 11]

As, addresses of these cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics, we reverse geocode those candidate area centers to get the addresses which can be presented to stakeholders[Figure 12] and these locations were plotted on map [Figure 13]

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

1923 Bergen St, New York, NY 11233--11233       => 3.1km from Brooklyn
1 Jewel McCoy Ln, New York, NY 11213--11213      => 1.0km from Brooklyn
125 Sutter Ave, New York, NY 11212--11212        => 4.0km from Brooklyn
757 Putnam Ave, New York, NY 11221--11221        => 0.9km from Brooklyn
1474 Prospect Pl, New York, NY 11213--11213      => 2.2km from Brooklyn
632 Macdonough St, New York, NY 11233--11233     => 2.1km from Brooklyn
959 St Marks Ave, New York, NY 11213--11213      => 1.6km from Brooklyn
49 E 92nd St, New York, NY 11212--11212          => 3.7km from Brooklyn
2277 Pacific St, New York, NY 11233--11233       => 3.8km from Brooklyn
```
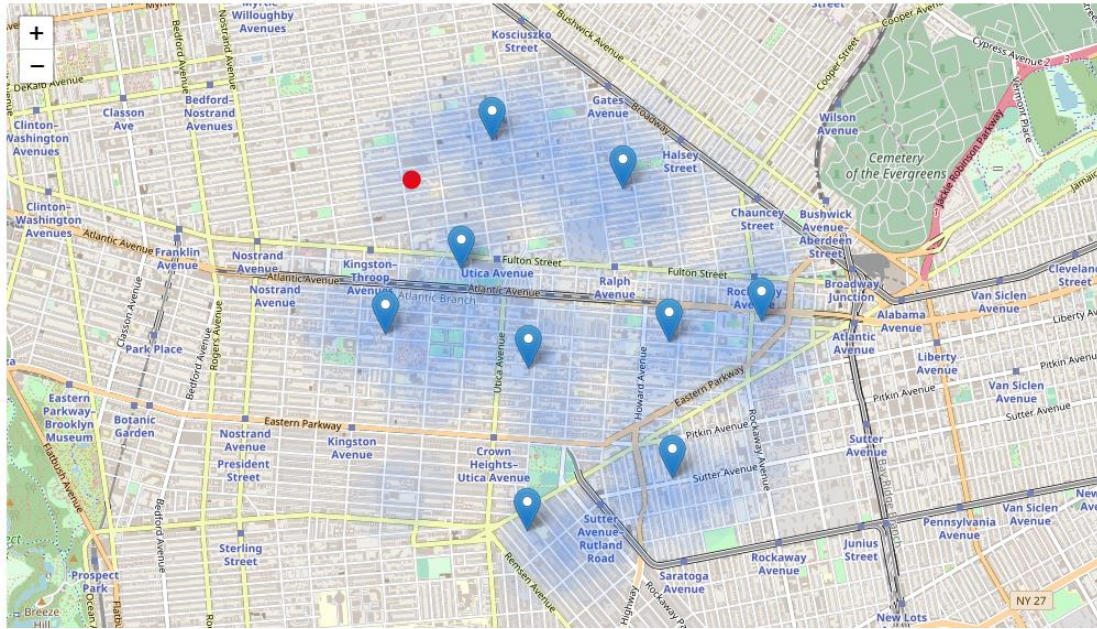
Figure 12

Figure 13

Results and Discussion

Our analysis shows that although there is are 209 drugstores in Brooklyn, there are pockets of low drugstore density fairly close to city center. Highest concentration of restaurants was detected west, north and north-east of Brooklyn, so we focused our attention to areas center, east and south-east, corresponding to Bedford-Stuyvesant, Crowns Height and Brownsville.

After directing our attention to this more narrow area of interest, we first created a dense grid of location candidates (spaced 300m apart); those locations were then filtered so that those with no more than two drugstores in radius of 500 meters were removed.

Those location candidates were then clustered to create zones of interest which contain greatest number of location candidates. Addresses of centers of those zones were also generated using reverse geocoding to be used as markers/starting points for more detailed local analysis based on other factors.

Result of all this is 9 zones containing largest number of potential new drugstore locations based on number of and distance to existing drugstores. This, of course, does not imply that those zones are actually optimal locations for a new drugstore! Purpose of this analysis was to only provide info on areas close to Brooklyn Center but not crowded with existing drugstore - it is entirely possible that there is a very good reason for small number of drugstore in any of those areas, reasons which would make them unsuitable for a new drugstore regardless of lack of competition in the area. Recommended zones should therefore be considered only as a starting point for more detailed analysis which could eventually result in location which has not only no

nearby competition but also other factors taken into account and all other relevant conditions met.

Conclusion

Purpose of this project was to identify Brooklyn areas close to center with low number of drugstores in order to aid stakeholders in narrowing down the search for optimal location for a new Drugstore. By calculating drugstore density distribution from foursquare data we have collection of locations which satisfy some basic requirements regarding existing nearby Drugstore. Clustering of those locations was then performed in order to create major zones of interest (containing greatest number of potential locations) and addresses of those zone centers were created to be used as starting points for final exploration by stakeholders.

Final decision on optimal drugstore location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended zone, taking into consideration additional factors like population, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.