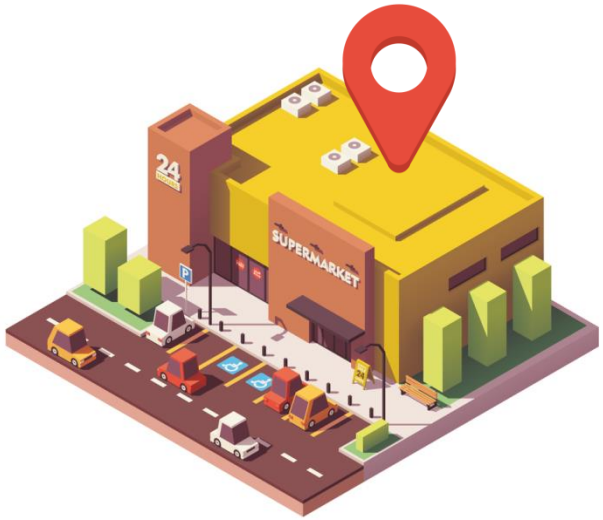
Applied Data Science Capstone Project

The Battle of Neighborhoods

Coursera - IBM Data Science Professional Certificate

Finding an optimal location for a Supermarket in Brooklyn, New York City



Submitted by:

Somraj Chowdhury

Business Problem

Problem:

This project will aim at finding a suitable and optimal location for a Supermarket in Brooklyn, in the City of New York, United States.

Problem Description:

A supermarket is a self-service shop offering a wide variety of food, beverages and household products, organized into sections and shelves. With the increase in the population, the demand and availability of essential resources like food and other households increase and supermarkets serve this very purpose of providing the resources under one roof.

Target Audience:

This project is aimed at those individuals and business groups who are interested to invest in opening a supermarket in Brooklyn, New York.

Data

New York City Data:

This dataset contains data of the boroughs and the neighborhoods that exist in each borough along with the geographical coordinates of each neighborhood.

This dataset holds data of a total of ***5*** boroughs and ***306*** neighborhoods.

Supermarket Data from Foursquare API:

Information about the existing supermarkets in Brooklyn, New York City like the name of supermarkets in a neighborhood and their location along with geographical coordinates etc. will be obtained from the Foursquare API.

Population Data for each Brooklyn neighborhood:

Population data for each Brooklyn neighborhood was created by finding the population of each neighborhood individually from the web.

Methodology

Load data and explore NYC neighborhoods:

- Download the dataset from the server which is a ***json*** file.
- Extract the required data from the ***features key*** of the downloaded json file.
- Transform the json data into a pandas dataframe with the following features:

	Borough	Neighborhood	Latitude	Longitude

GeoPy

This library is used to convert an address into its latitude and longitude values.

Folium

This library is used for visualizing data with the help of interactive maps.

Explore Brooklyn Neighborhoods:

- Since we want to find a suitable and optimal location for opening a supermarket in the neighborhoods of Brooklyn borough, we create a dataframe that only consists of Brooklyn neighborhoods and all the related data.
- Obtain the latitude and longitude values of Brooklyn using the geopy library by giving the address as '***Brooklyn, NY***' and then using the folium library create a map of Brooklyn with the 70 neighborhoods superimposed on top.

Neighborhood in Brooklyn

Explore supermarkets in Brooklyn using Foursquare API:

- In order to use Foursquare services, you need to create free developer account first.
- Once created, you are provided with unique ***CLIENT_ID*** and ***CLIENT_SECRET*** which will be required to request venue data in your application and very importantly that shouldn't be shared with anyone else.

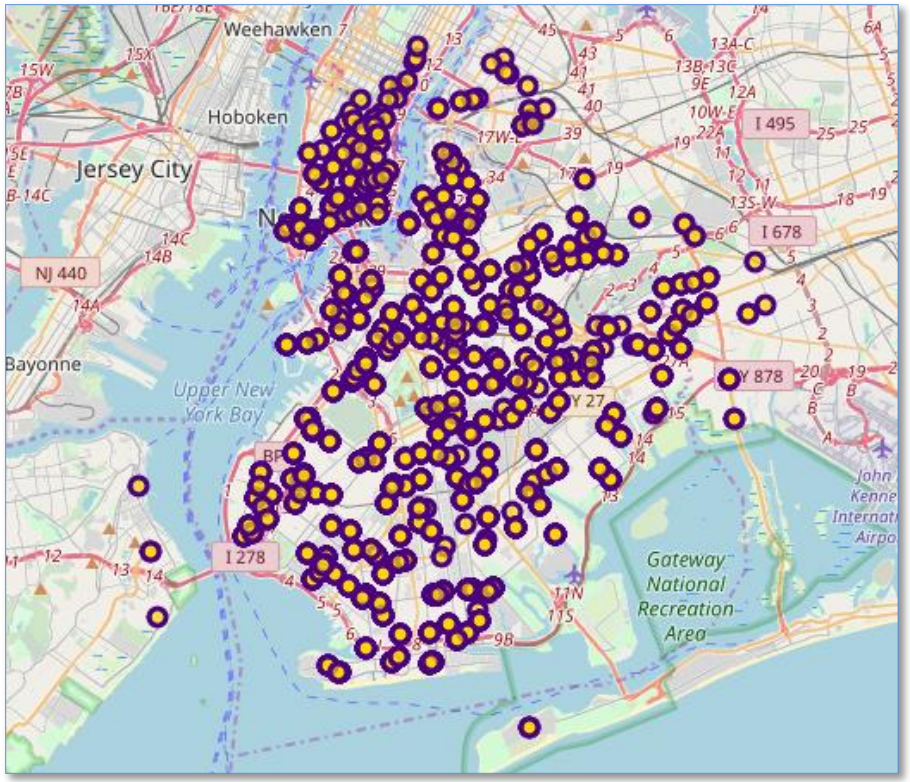
FOURSQUARE
FIND YOUR PLACE

- In order to retrieve venue specific data for example data about restaurants, schools etc. you need to specify a specific ***category ID*** associated with it. In this problem, we want data related to supermarkets in Brooklyn; therefore we select the supermarket category

- The data returned by the Foursquare API will be in ***json*** format and all the important information that we require about the venue (supermarket) is stored in the ***items*** key of the retrieved json data.
- This supermarket data is added to the Brooklyn neighborhood data so as to know which supermarket is in which neighborhood.
- In this capstone project, we have retrieved information about all supermarkets within ***4km*** radius of the Brooklyn borough.

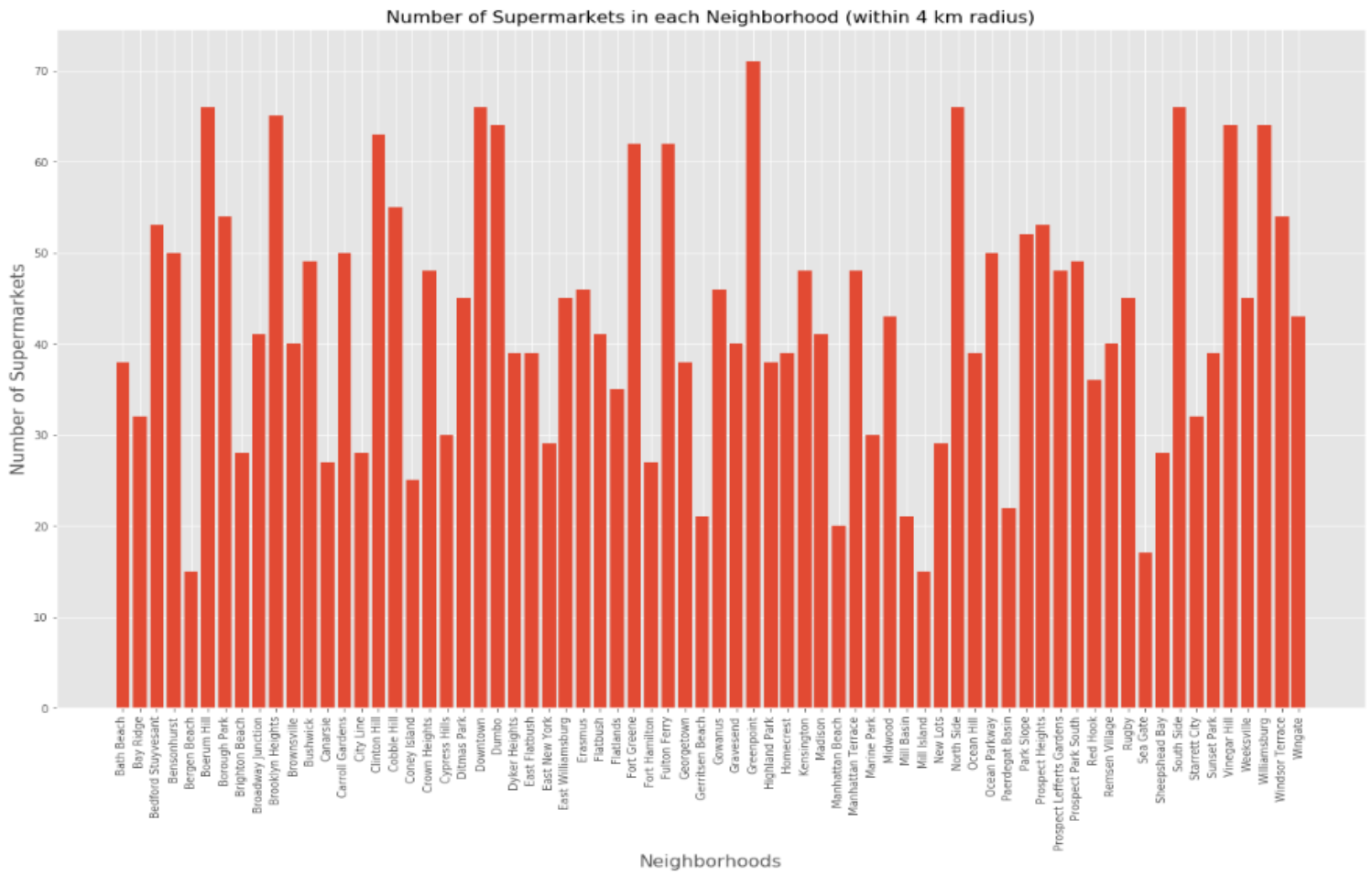
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Supermarket Name	Supermarket Latitude	Supermarket Longitude
0	Bay Ridge	40.625801	-74.030621	Met Fresh Supermarket	40.616528	-74.034003
1	Bay Ridge	40.625801	-74.030621	Metropolitan CityMarket	40.617375	-74.030735
2	Bay Ridge	40.625801	-74.030621	CTown Supermarkets	40.629234	-74.022803
3	Bay Ridge	40.625801	-74.030621	Jmart 新世界超市	40.610080	-74.001221
4	Bay Ridge	40.625801	-74.030621	Foodtown	40.619927	-74.032301
5	Bay Ridge	40.625801	-74.030621	Scaturro Supermarkets	40.629409	-74.005051
6	Bay Ridge	40.625801	-74.030621	Food Dynasty	40.611275	-74.008544
7	Bay Ridge	40.625801	-74.030621	Bingo Wholesale	40.629211	-73.999403
8	Bay Ridge	40.625801	-74.030621	Key Food Fresh & Natural	40.628058	-74.029126
9	Bay Ridge	40.625801	-74.030621	Fei Long Market	40.633397	-74.011286

Using the folium library, we then create a map of Brooklyn with the all supermarkets in the neighborhoods superimposed on top.

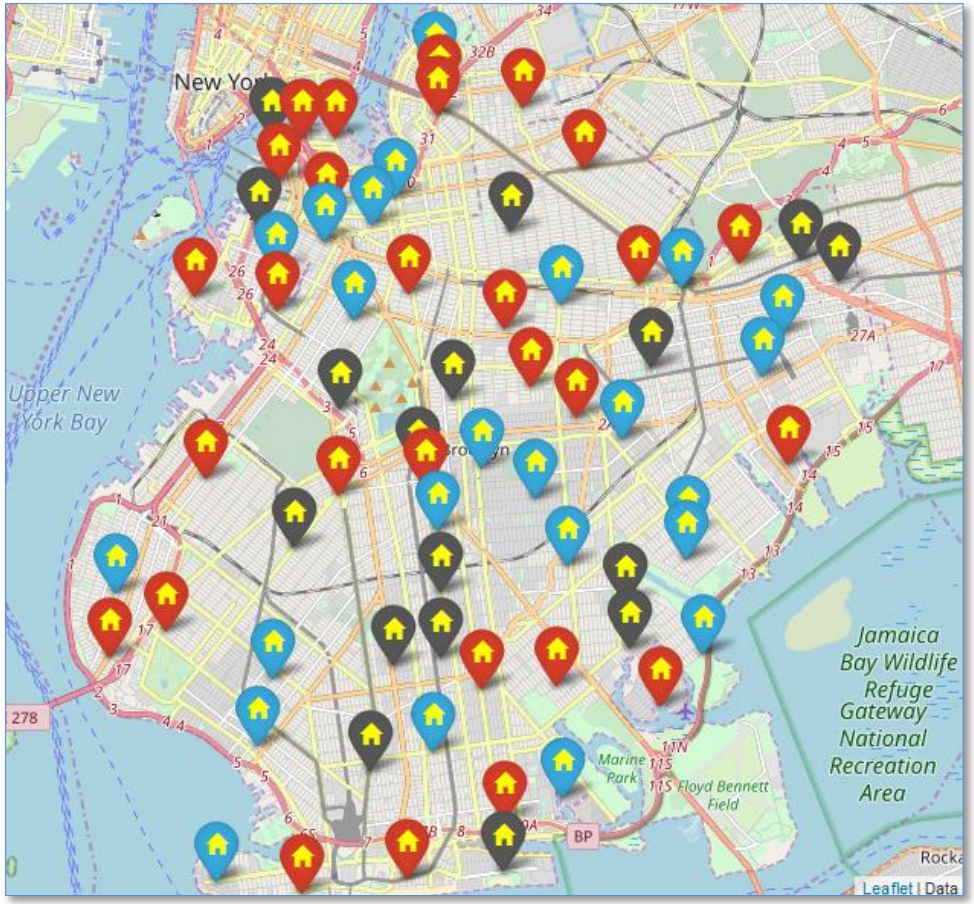


Supermarket in Brooklyn neighborhoods

We find the ***total number of supermarkets*** in each neighborhood and then visualize it using bar chart.



We create a visualization using folium where we group the neighborhoods based on the number of supermarkets present in that neighborhood and distinguish each group of neighborhoods using colored folium markers.



Neighborhoods with **15 - 30** supermarkets.

Neighborhoods with **31 - 45** supermarkets.

Neighborhoods with **>= 46** supermarkets.

Population and portion data of each Brooklyn neighborhood:

After determining number of supermarkets in each neighborhood, we then add the population data for each of the Brooklyn neighborhood along with a feature specifying in which portion of Brooklyn borough the neighborhood resides in.

	Neighborhood	Portion of Borough	Population	Number of Supermarkets
0	Bath Beach	Southwestern Brooklyn	29931	38
1	Bay Ridge	Southwestern Brooklyn	79371	32
2	Bedford Stuyvesant	Northern Brooklyn	158000	53
3	Bensonhurst	Southwestern Brooklyn	152000	50
4	Bergen Beach	Southern Brooklyn	45231	15

We can divide the neighborhoods in Brooklyn borough into ***6*** unique portions:

1. **Central Brooklyn**
2. **Eastern Brooklyn**
3. **Northern Brooklyn**
4. **Northwestern Brooklyn**
5. **Southern Brooklyn**
6. **Southwestern Brooklyn**

Result

After partitioning Brooklyn borough into 6 portions, we have made the following observations in terms of the **total number of supermarkets** and the **total population** in each borough portion and a ***score*** that will help in finding an optimal location to open a supermarket.

Portion 1:
Central Brooklyn

```
Number of neighborhoods in Central Brooklyn: 15
Total number of supermarkets in Central Brooklyn: 684
Total population in Central Brooklyn: 1197391
Score: 1751
```

Portion 2:
Eastern Brooklyn

```
Number of neighborhoods in Eastern Brooklyn: 8
Total number of supermarkets in Eastern Brooklyn: 253
Total population in Eastern Brooklyn: 450720
Score: 1782
```

Portion 3:
Northern Brooklyn

```
Number of neighborhoods in Northern Brooklyn: 7
Total number of supermarkets in Northern Brooklyn: 387
Total population in Northern Brooklyn: 706258
Score: 1825
```

Portion 4:
Northwestern Brooklyn

```
Number of neighborhoods in Northwestern Brooklyn: 15
Total number of supermarkets in Northwestern Brooklyn: 845
Total population in Northwestern Brooklyn: 271829
Score: 322
```

Portion 5:
Southern Brooklyn

```
Number of neighborhoods in Southern Brooklyn: 17
Total number of supermarkets in Southern Brooklyn: 524
Total population in Southern Brooklyn: 959338
Score: 1831
```

Portion 6:
Southwestern Brooklyn

```
Number of neighborhoods in Southwestern Brooklyn: 8
Total number of supermarkets in Southwestern Brooklyn: 304
Total population in Southwestern Brooklyn: 589085
Score: 1938
```

Discussion

In the results section above, we calculated scores for all 6 portions of the borough. This score is calculated by dividing the ***total population in a portion*** by the ***total number of supermarkets in that portion***.

What is the significance of this score?

- A *lower score* indicates that there is sufficient number of supermarkets for the present population.
- A *higher score* signifies that there is lesser number of supermarkets compared to the population of the portion.

Based on the analysis of each portion, we note that:

- **Portion 1 - Central Brooklyn** has the highest population of 1.2 million (approx.) which means a higher chance of customer expectancy.
- **Portion 4 - Northwestern Brooklyn** has the lowest score which means there is more business and competition among the supermarkets.
- **Portion 6 - Southwestern Brooklyn** has the highest score which means there is a deficiency in the number of supermarkets.

Conclusion

The results of this capstone project are obtained by performing analysis on limited data. The purpose of this project was to find an optimal and suitable location for opening a supermarket in the Brooklyn borough in order to aid stakeholders with best locations that could benefit the business and the growth of the brand. If a neighborhood or location has a high density of supermarkets, then it means that there is a high demand in those locations. Also locations with larger population will need more supermarkets in order to cater their needs. The final decision on selection of an optimal location will be taken by the stakeholders based on characteristics such as popularity of the location and proximity to major roads etc.

Therefore, the individual or business group interested in investing in opening a supermarket in the Brooklyn borough can invest in the Northwestern Brooklyn where there is much more competition in the supermarket business or can invest in Central Brooklyn where there is the highest population of all the 6 portions.