Finding an optimal location to open a coffee Shop in Brooklyn, New York City.

Md Tanzil Hoque Chowdhury

June 15, 2021

1 Introduction / Business Problem

The objective of this project is to analyze and select the best location in Brooklyn, New York City to open a new coffee shop. Searching for the optimal location is challenging as we need to consider many things before we open a coffee shop in a place. If the location is very saturated with other coffee shops, then it will not be a good idea to open a new coffee shop in that particular place as it will be a very tough competition with other coffee shops, which are already existed and popular in that area. Rather, we should choose an area where we can identify the demand of a coffee shop based on any available coffee shops and other popular venues. To address this business problem, research will be done using data science methodology and machine learning techniques such as clustering to find an optimal location for the coffee shop.

2 Data Acquisition

To solve the business problem of finding the most suitable location in Brooklyn, New York city, we will need the following data:

- List of boroughs and neighborhoods in the city of New York.
- List of zip codes for each neighborhood.
- Latitude and longitude coordinates of those neighborhoods to plot maps and tie in venue data.
- Information about different categories of venues such as restaurant cluster locations, whether a Coffee shop already exists or not.

3 Methodology

The New York data is in a Json format. So, first we read the json data and found that all the relevant data is in the features key, which is basically a list of the neighborhoods. So, we extracted this important information from the features key. Then we converted the Jason data to pandas data frame. We found that the data frame has 5 boroughs and 306 neighborhoods. Then we used the geopy library to get the latitude and longitude values of New York City and visually represent it using the python folium library. We calculated the unique borough of

New York city, and we chose the Brooklyn as this is the place of our interest. We checked the neighborhoods of the Brooklyn. We visualized Brooklyn and its neighborhoods and calculated the latitude and longitude of the Brooklyn. Then we used Foursquare to find all the neighborhoods and different venues in Brooklyn. 'Bay Ridge' neighborhood was selected and, we got the top venues that are in 'Bay Ridge' within a radius of 5000 meters using Foursquare. We sent the GET request and examine the results. We noticed that all the information is in the items key. We extracted information from the items key and convert into pandas data frame. We got the information of different venues and extracted the categories of different venues. We explored neighborhoods in Brooklyn and created a function to repeat the same process to all the neighborhoods in Brooklyn and checked how many venues were returned for each neighborhood. We calculated the unique venue categories and found that 288 unique venue categories. Each neighborhood was analyzed. We grouped rows by neighborhood and by taking the mean of the frequency of occurrence of each category and printed each neighborhood along with the top 5 most common venues. Then this information was put into a pandas data frame having 15 top venues in each neighborhood. We ran k-means clustering technique to cluster the neighborhood into 3 different clusters. In each cluster we checked is there any coffee shop available in each neighborhood of the cluster or not. If there is no coffee shop in the neighborhoods of same cluster then we checked other categories related with coffee shop and based on this information we finally decided where to open the coffee shop.

4 Results

We used Folium to map the neighborhood data. We showed the neighborhoods with the cluster labels. Notice that there are total 3 clusters. Red, green and purple. The red cluster has 58 neighborhoods, the green cluster has 11 neighborhoods, and the purple cluster has only 1 neighborhood which is considered as an outlier. We checked the presence of coffee shop in each neighborhood of each cluster and found that out of 58 neighborhoods in cluster 1 (red cluster), 32 neighborhoods have coffee shops and 26 neighborhoods do not have coffee shop. In cluster 3 (green cluster), out of 11 neighborhoods, only 1 neighborhood has coffee shop and other 10 neighborhoods do not have coffee shop. We also checked the top 15 venues in each neighborhoods based on the popularity and choice of the people living in that area. We found that the most common venues in neighborhoods of green cluster are like restaurant, pizza place, donut shop, juice shop and so on. Based on this information we can assume that as there is only 1 coffee shop in this cluster of 11 neighborhoods, it will be a good place to open a coffee shop in this cluster.

4.1 Does Brooklyn have any space for another coffee shop?

To answer this question, we need to count the venue categories collected for all zip codes in the study, then sort them in descending order, and graph them with Python's default matplotlib. Results are shown below in the bar chart. From the bar chart we can see that coffee shop is the 2nd most common place in Brooklyn. As it is the 2nd most popular place in Brooklyn and as there is only 1 coffee shop in the 3rd cluster, it will be good idea to open a coffee shop in the neighborhoods of the 3rd cluster.

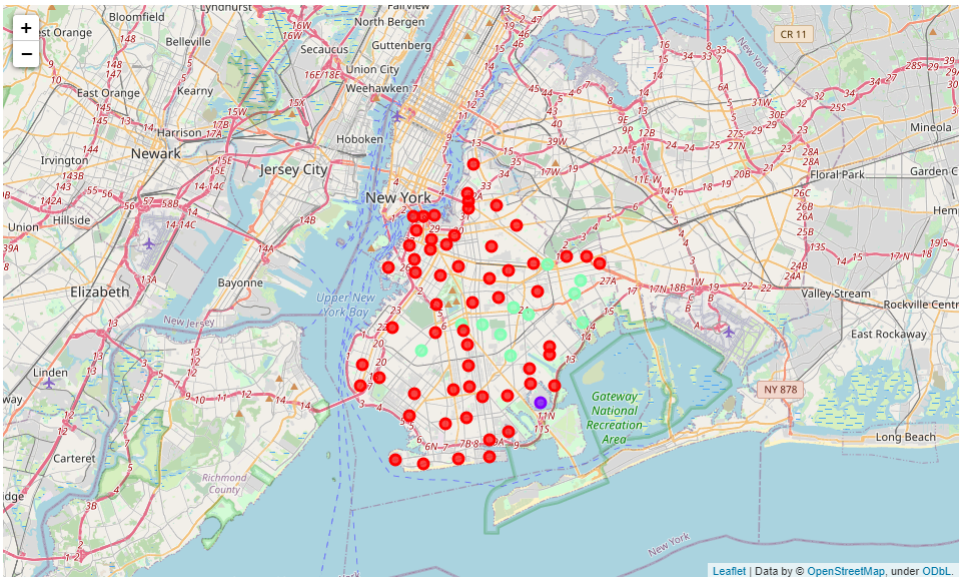

Figure 1: Different Clusters in Brooklyn, New York city.

	Borough	Neighborhoods	Latitude	Longitude	Cluster Labels	Coffee Shop Exists	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank
10	Brooklyn	East Flatbush	40.641718	-73.936103	2	False	Chinese Restaurant	Park	Pharmacy	Moving Target	Wine Shop	Supermarket	Departmer Stor
26	Brooklyn	East New York	40.669926	-73.880699	2	False	Fast Food Restaurant	Asian Restaurant	Metro Station	Caribbean Restaurant	Pharmacy	Salon / Barbershop	Pizza Plac
27	Brooklyn	Starrett City	40.647589	-73.879370	2	False	Donut Shop	Pharmacy	Bus Stop	Supermarket	Caribbean Restaurant	Shopping Mall	Chines Restaurar
29	Brooklyn	Flatlands	40.630446	-73.929113	2	False	Pharmacy	Fast Food Restaurant	Fried Chicken Joint	Caribbean Restaurant	Discount Store	Video Store	Paper Offic Supplie Stor
34	Brooklyn	Borough Park	40.633131	-73.990498	2	True	Bank	Pharmacy	Fast Food Restaurant	Pizza Place	Bistro	Coffee Shop	Caf
47	Brooklyn	Prospect Park South	40.647009	-73.962613	2	False	Department Store	Grocery Store	Fast Food Restaurant	Clothing Store	Caribbean Restaurant	Mobile Phone Shop	Donu Sho
56	Brooklyn	Rugby	40.655572	-73.926882	2	False	Sandwich Place	Caribbean Restaurant	Bank	Fried Chicken Joint	Gas Station	Grocery Store	Beac
roll output; double click to hide			40.652117	-73.916653	2	False	Caribbean Restaurant	Fast Food Restaurant	Salad Place	Spa	Supermarket	Gas Station	Caf
58	Brooklyn	New Lots	40.662744	-73.885118	2	False	Pharmacy	Fried Chicken Joint	Grocery Store	Pizza Place	Discount Store	Breakfast Spot	Foo
64	Brooklyn	Broadway Junction	40.677861	-73.903317	2	False	Gas Station	Donut Shop	Nightclub	Tunnel	Sandwich Place	Speakeasy	Breakfas Spo
69	Brooklyn	Erasmus	40.646926	-73.948177	2	False	Caribbean Restaurant	Pharmacy	Ice Cream Shop	Supermarket	Convenience Store	Mobile Phone Shop	Ban

Figure 2: Neighborhoods in the Brooklyn having coffee shop or not.

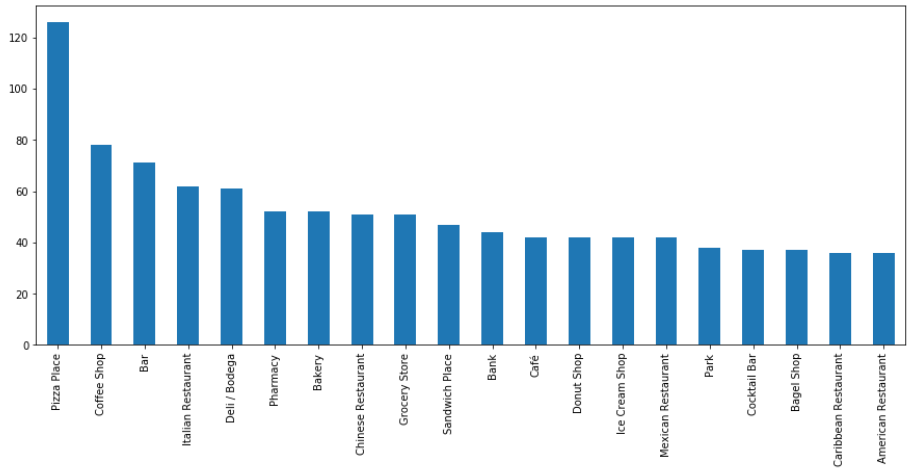

Figure 3: Different venues and their number in the Brooklyn.

5 Discussion

5.1 Cluster Analysis

Examination of the clusters completely indicates Cluster 3 is best having 10 neighborhoods absent of any coffee shop competition out of 11 neighborhoods. These neighborhoods have different other restaurants like coffee shop that clearly indicates that opening a coffee shop will be good decision in these areas. Also, as there is only 1 coffee shop in these neighborhoods there will be not such tough competition with other restaurants and there will be good profit for the coffee shop owner. Cluster 1 has lots of coffee shop and thus it will not be a wise decision to open a coffee shop in this cluster as the coffee shop owner has to face a tough competition there. Cluster 2 is outlier with only 1 neighborhood. So we can ignore this cluster.

5.2 Recommendations

Based on the above analysis we can suggest that Rugby or East Flatbush neighborhoods should be a good choice to open a coffee shop in one of these areas as most common venues in these are are Sandwich Place, Caribbean Restaurant, Bank, Fried Chicken Joint, Gas Station, Grocery Store, Beach, Mobile Phone Shop, Supermarket, Pharmacy, Pizza Place, Seafood Restaurant, Food Court and so on. So, as these places have high demands in these areas, opening a coffee shop in any one of these areas should be a good decision as it will not only be a good source of profit for the coffee shop owner but also people in these areas do not need to go to another neighborhoods of the 1st cluster to fulfill their cravings for coffee. They can easily get it in their own or nearest neighborhoods.

6 Conclusions

The main aim of this project was to find an optimal place to open a coffee shop in Brooklyn, New York city. Neighborhood data from the Brooklyn was used along with zip code of each neighborhoods. We used location data to find the venue data details from Foursquare and calculations using Python code. This base data set was loaded, cleaned, and prepared for cluster analysis and other operation. The Folium library was used for map related visualizations, and

the standard python matplotlib was used for standard graphs. It was a very interesting project and I really enjoyed a lot working on this project.