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

Md Tanzil Hoque Chowdhury

Middle Tennessee State University

June 15, 2021

IBM Data Science Professional Certificate Capstone Project

Introduction/ Business Problem

- The main objective is to find the best location in Brooklyn to open a new coffee shop.
- Clustered neighborhoods and near other restaurants.
- 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.
- 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.
- Research is done using data science methodology and machine learning techniques such as clustering to find an optimal location for the coffee shop.

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.

Methodology

- First we read the json data and found that all the relevant data is in the features key.
- We extracted this important information from the features key.
- 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 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.

Methodology

- '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 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.
- We put the information into a pandas data frame having 15 top venues in each neighborhood.

Clustering

- 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.

Results

- We used Folium to map the neighborhood data. We showed the neighborhoods with the cluster labels.
- 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.

Results

- 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.

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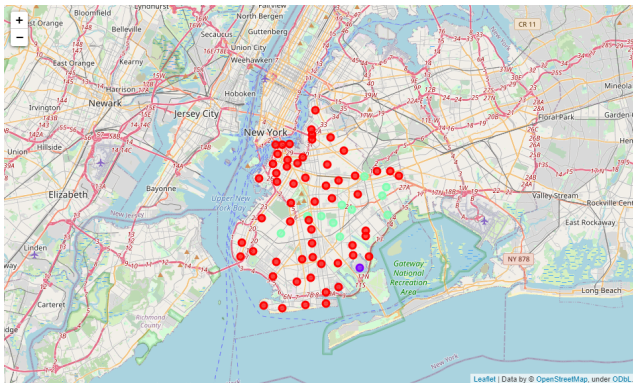
Figure: Different Clusters in Brooklyn, New York city.

Results

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

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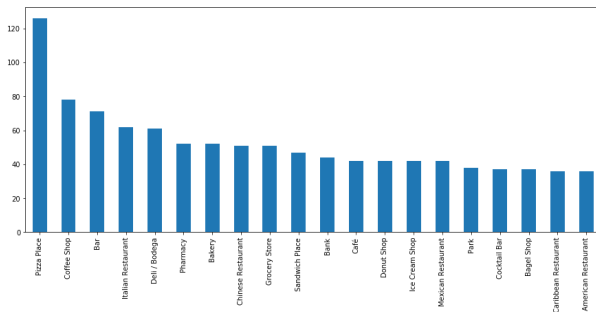
Figure: Neighborhoods in the Brooklyn having coffee shop or not.

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.

- 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.

Cluster Analysis

- 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.
- 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.

Conclusion

- 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.

The End