Battle of Neighbourhoods

Introduction

A restaurant is to be opened in a borough of New York. For this problem Brooklyn, NY is selected. The neighbourhood in which the restaurant should be opened is to be decided.The deciding factor in selection is the proximity to other popular restaurants. Since it is very competitive to establish a restaurant in a neighbourhood with plenty of popular restaurants, a neighbourhood with least dining options is a good choice. Other factors in location determination include accessibility and distance from the city center. In this project we will be focusing only on the proximity to other similar venues.

This type of analysis will be useful for stakeholders planning on investing in restaurants and fast food joints.

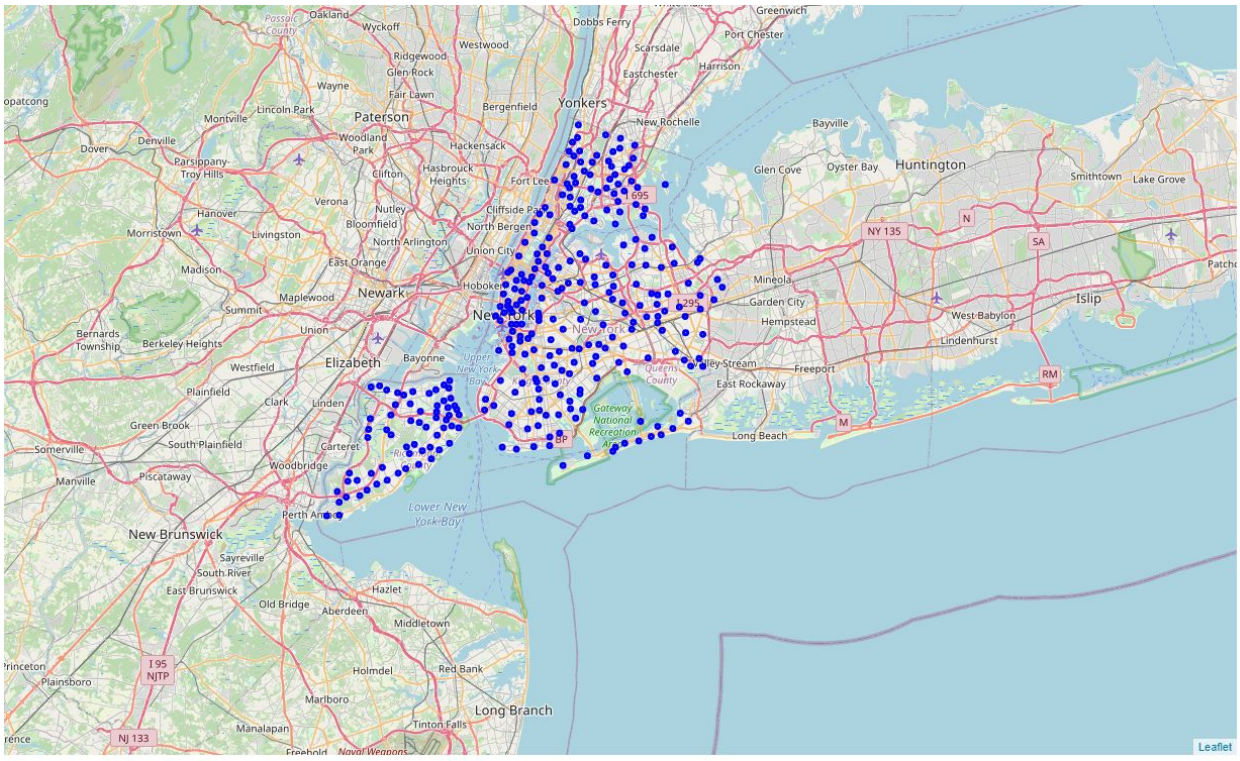
Datasets

New York has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighborhood.New York data can be downloaded from:
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
The above link redirects to a json file which can be downloaded.
From the above dataset features of each borough and neighbourhood can be extracted Into dataframe. Since we will be focusing in Brooklyn, data for only Brooklyn will be sliced from the dataframe to create a new dataframe.
Using folium a map of brooklyn can be visualized and neighbourhoods marked.

Foursquare api is then used in each of the neighbourhoods to determine the top venues in that neighbourhood. The venues are stored in a dataframe and frequency of each category of venue is determined. From the data we can see that the maximum venues are pizza plaza followed by coffee shops.

Methodology

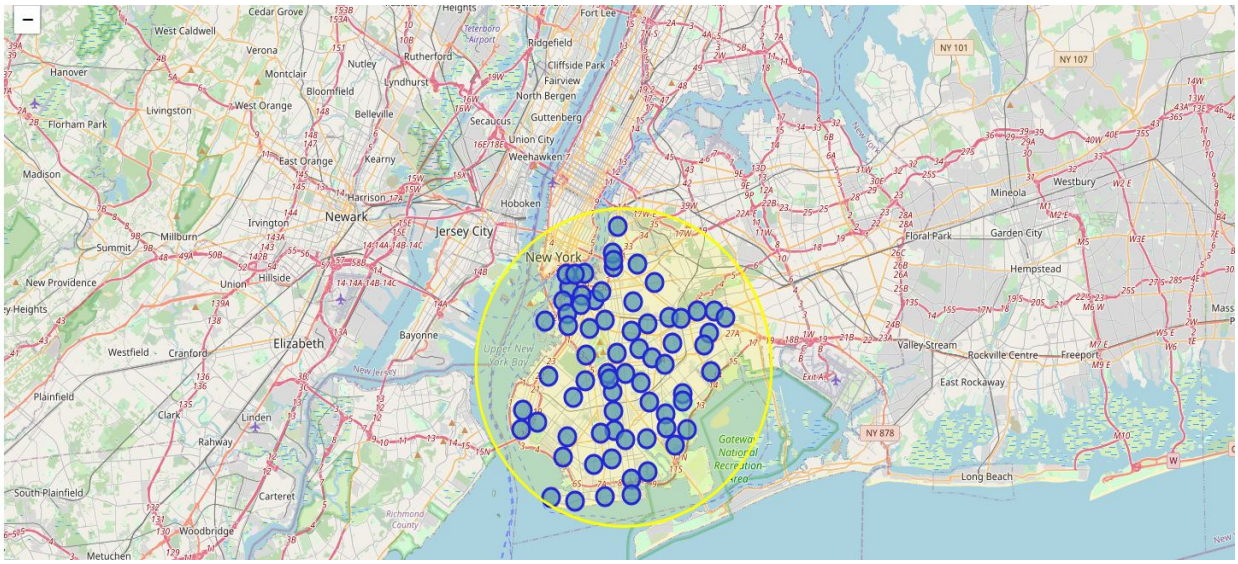
The datasets for all the neighbourhoods of a borough are organized. Foursquare location data is used to identify the top 100 venues in each neighbourhood. Then a dataframe is created containing the location of each venue along with its borough,neighbourhood and category.Using value_counts() the number of unique venues are determined. Further analysis is performed on the data to determine the frequency of each category in each neighbourhood. With this we can determine the concentration of restaurants at any neighbourhood and avoid that location.
After cleaning and processing the datasets, machine learning algorithms are used to determine clusters in the borough. A cluster size of 8 was chosen and K Means Clustering was performed. Based on these clusters the concentration of restaurants in each cluster was observed. Also only top 5 venues in each neighbourhood were taken into consideration. All clusters are then superimposed on the map of brooklyn using folium for visualization.



New York City

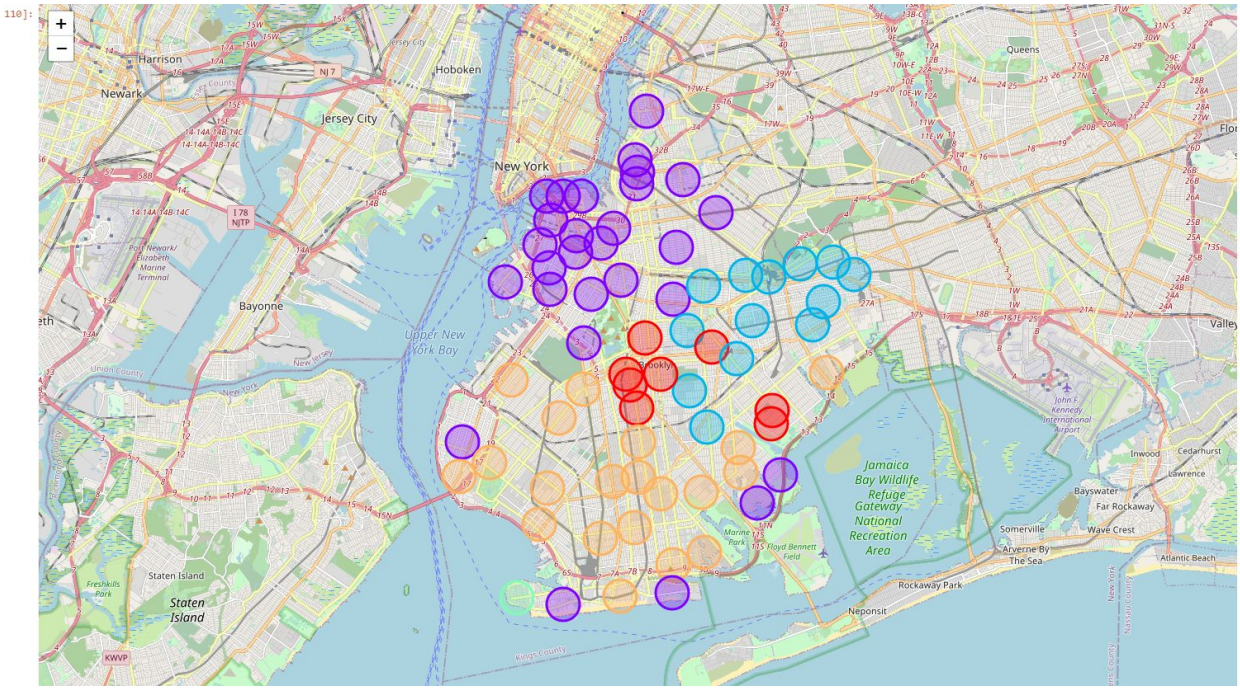
Results

Below is the folium generated map of Brooklyn with its neighborhoods indicated by the blue circles.



Brooklyn with all neighborhoods

After using K Means clustering, all 70 neighborhoods were grouped into 5 different clusters. Based on these clusters, location determination was made.



Clusters

For example below is the screenshot of top 5 venues of cluster 5.

Cluster 5

```
[118]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 4, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

[118]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
1	Bensonhurst	Bank	Pizza Place	Bakery	Japanese Restaurant	Chinese Restaurant
2	Sunset Park	Bakery	Mexican Restaurant	Pizza Place	Chinese Restaurant	Asian Restaurant
4	Gravesend	Pizza Place	Donut Shop	Bakery	Gym	Grocery Store
5	Brighton Beach	Beach	Sushi Restaurant	Bakery	Restaurant	Pharmacy
6	Sheepshead Bay	Russian Restaurant	Sushi Restaurant	Italian Restaurant	Turkish Restaurant	Pizza Place
7	Manhattan Terrace	Pizza Place	Bank	Donut Shop	Sushi Restaurant	Grocery Store
11	Kensington	Pizza Place	Grocery Store	Gas Station	Thai Restaurant	Café
27	Starrett City	Pizza Place	Mobile Phone Shop	Department Store	Pharmacy	Park
33	Bath Beach	Pizza Place	Bank	Japanese Restaurant	Chinese Restaurant	Bakery
34	Borough Park	Pizza Place	Bank	Grocery Store	Sandwich Place	Pharmacy
35	Dyker Heights	Gym / Fitness Center	Bagel Shop	Pizza Place	Cosmetics Shop	Chinese Restaurant
36	Gerritsen Beach	Donut Shop	Ice Cream Shop	Playground	Harbor / Marina	Deli / Bodega
37	Marine Park	Clothing Store	Sandwich Place	Cosmetics Shop	Furniture / Home Store	American Restaurant
46	Midwood	Pizza Place	Bakery	Grocery Store	Ice Cream Shop	Café
48	Georgetown	Italian Restaurant	Bank	Supermarket	Chinese Restaurant	Donut Shop
52	Ocean Parkway	Grocery Store	Donut Shop	Bank	Fast Food Restaurant	Sushi Restaurant
53	Fort Hamilton	Pizza Place	Japanese Restaurant	Bagel Shop	Gym / Fitness Center	Bar
60	Mill Basin	Clothing Store	Pizza Place	Italian Restaurant	Pharmacy	Bank
66	Homecrest	Bank	Grocery Store	Russian Restaurant	Pharmacy	Donut Shop
68	Madison	Pizza Place	Sushi Restaurant	Mobile Phone Shop	Bagel Shop	Pharmacy

Cluster 5

DIscussions

Pizza plaza is the most popular venue in all of Brooklyn followed by Coffee shops while bath houses, print shops are least popular. Also the accuracy of the above code depends on the location data provided by **Foursquare api**, accuracy of **geopy**. Also the above model assumed each neighborhood as a circle with a radius of 1 km.

Conclusion

So from examining the above clusters we can predict the neighbourhood where food places should be opened. For example opening a restaurant at **cluster 4 in the neighbourhood of Sea Gate** would be appropriate since there are no popular restaurants nearby.

The above procedure can be applied to all boroughs of New York to determine location for any type of commercial establishment.