Capstone Project - Brooklyn Clustering Analysis

1.Introduction

1.1 Problem

In this project we will find a location in Brooklyn. This report will be targeted to foreign students interested in neighborhood in Brooklyn, New York.

We will try to detect the neighborhoods where have good environment for theirs studying.

2. Data acquisition and cleaning

2.1 Data source

2.2.1 New York Boroughs & Neighborhood dataset from https://cocl.us/new_york_dataset

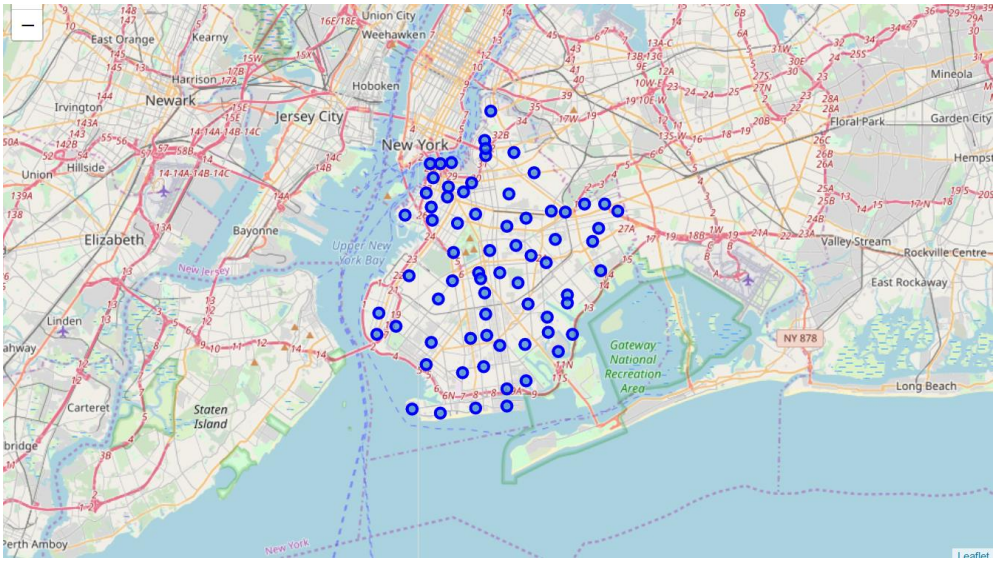
2.2.2 Venue categories in each neighborhood will be obtained using Foursquare API

2.2 Data cleaning

New York Boroughs & Neighborhood dataset are downloaded. Then borough, neighborhood, longitude and latitude columns are extracted and converted to data frame. Then we filter only rows which have borough is Brooklyn for analysis. Then we used Foursquare API to get some venue categories of neighborhoods in Brooklyn and merge with Brooklyn dataset.

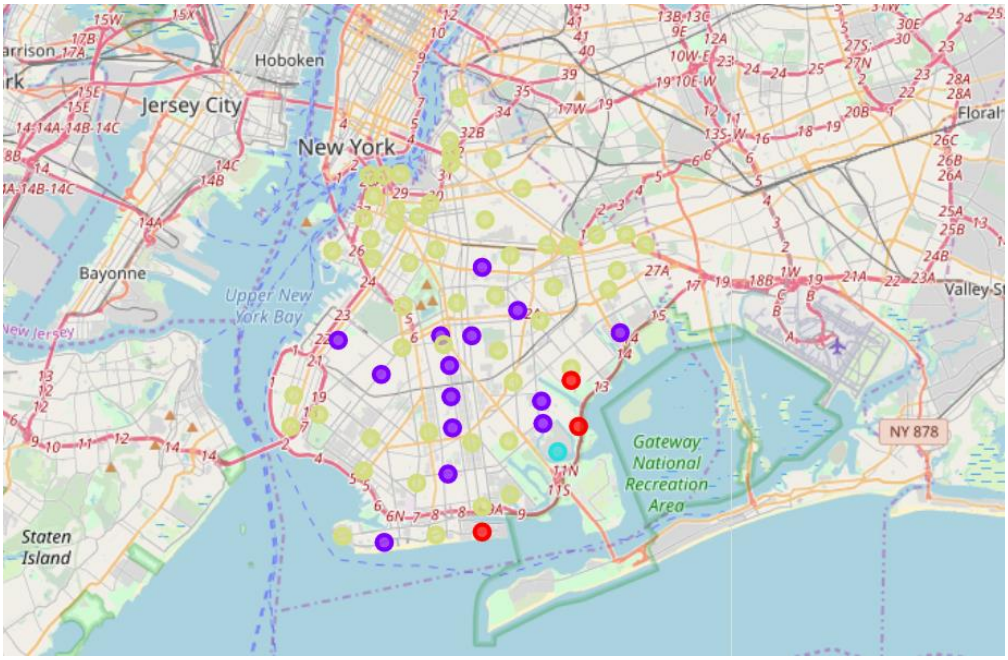
3. Exploratory Data Analysis

We visualize Brooklyn and neighborhoods in a map



4. Predictive Modeling

In this project we will use Foursquare API for gathering venue categories in each neighborhood. Then we will create clusters (using k-means clustering) of those locations. Then we can examine each cluster and determine the location for foreign students.



Cluster 1

```
brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 0, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

58]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
31	Manhattan Beach	Sandwich Place	Harbor / Marina	Ice Cream Shop	Playground	Pizza Place	Food	Bus Stop	Beach	Café	Factory
45	Bergen Beach	Harbor / Marina	Baseball Field	Playground	Athletics & Sports	Donut Shop	Food & Drink Shop	Food	Flower Shop	Fish Market	Fish & Chips Shop
59	Paerdegat Basin	Child Care Service	Asian Restaurant	Harbor / Marina	Bus Line	Food	Factory	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant

Cluster 2

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

[69]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
2	Sunset Park	Bank	Bakery	Latin American Restaurant	Mexican Restaurant	Pizza Place	Pharmacy	Gym	Fried Chicken Joint	Mobile Phone Shop	Deli / Bodega
7	Manhattan Terrace	Pizza Place	Donut Shop	Bagel Shop	Convenience Store	Ice Cream Shop	Mobile Phone Shop	Chinese Restaurant	Steakhouse	Coffee Shop	Grocery Store
8	Flatbush	Coffee Shop	Deli / Bodega	Caribbean Restaurant	Bank	Pharmacy	Mexican Restaurant	Plaza	Lounge	Chinese Restaurant	Juice Bar
9	Crown Heights	Pizza Place	Café	Museum	Playground	Burger Joint	Electronics Store	Candy Store	Bakery	Bagel Shop	Coffee Shop
27	Starrett City	Supermarket	Pharmacy	Caribbean Restaurant	Pizza Place	American Restaurant	Intersection	Bus Stop	Donut Shop	Fish & Chips Shop	Farmers Market
32	Coney Island	Baseball Stadium	Pizza Place	Caribbean Restaurant	Monument / Landmark	Theme Park Ride / Attraction	Brewery	Skating Rink	Music Venue	Beach	Dessert Shop
34	Borough Park	Bank	Café	Pizza Place	Pharmacy	Deli / Bodega	Men's Store	Fast Food Restaurant	Farmers Market	Restaurant	Coffee Shop

Cluster 3

```
brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 2, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

70]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
30	Mill Island	Pool	Filipino Restaurant	Factory	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Field	Fish & Chips Shop	Event Service

Cluster 4

```
brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 3, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]]
```

[71]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bay Ridge	Italian Restaurant	Spa	Greek Restaurant	Pizza Place	Grocery Store	American Restaurant	Bar	Pharmacy	Sandwich Place	Sushi Restaurant
1	Bensonhurst	Dessert Shop	Italian Restaurant	Ice Cream Shop	Sushi Restaurant	Pizza Place	Donut Shop	Chinese Restaurant	Liquor Store	Supermarket	Cha Chaan Teng
3	Greenpoint	Bar	Pizza Place	Cocktail Bar	Coffee Shop	Yoga Studio	Mexican Restaurant	Café	Record Shop	Sushi Restaurant	French Restaurant
4	Gravesend	Bus Station	Deli / Bodega	Pizza Place	Chinese Restaurant	Italian Restaurant	Bakery	Lounge	Donut Shop	Gym	Bar
5	Brighton Beach	Russian Restaurant	Restaurant	Beach	Eastern European Restaurant	Pharmacy	Gourmet Shop	Sushi Restaurant	Bank	Mobile Phone Shop	Convenience Store
6	Sheepshead Bay	Dessert Shop	Turkish Restaurant	Sandwich Place	Yoga Studio	Buffet	Italian Restaurant	Harbor / Marina	Karaoke Bar	Grocery Store	Miscellaneous Shop
10	East Flatbush	Department	Pharmacy	Moving Target	Supermarket	Fast Food	Caribbean	Chinese	Park	Hardware Store	Food & Drink

5. Results and Discussion

Our analysis shows that we should recommend foreign students to live in the neighborhoods in cluster 2 because there are venue categories for them.

6. Conclusion

Purpose of this project was to identify the neighborhoods in Brooklyn by using k-mean clustering for segmenting locations based on venue categories the neighborhoods