Final Report of the Capstone Project – The Battle of Neighborhoods Finding a Better Place in Brooklyn, NY

Introduction

This project will help people to better explore facilities around their neighborhood. It will help people making smart and efficient decision on selecting great neighborhood out of numbers of other neighborhoods in Brooklyn, New York. The project will compare various location and venues around each of the neighborhoods like cafes, schools, super markets, medical stores, shopping centers, theatres, hospitals, etc. This project aims to create an analysis of features for people wishing to move to Brooklyn to search for the best neighborhood as a comparative analysis between neighborhoods. An additional feature is included as comparison which is average rent. It will help people to get awareness of the area and neighborhood before moving to a new city, state, country or place for their work or to start a new fresh life.

Data

For this project we wil be using data available in the New York week 3 assignment.
Link: https://geo.nyu.edu/catalog/nyu_2451_34572

Also we will be using data of average rent from this link :https://www.rentcafe.com/average-rent-market-trends/us/ny/brooklyn/

Foursquare API data
We will use data of different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information. Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

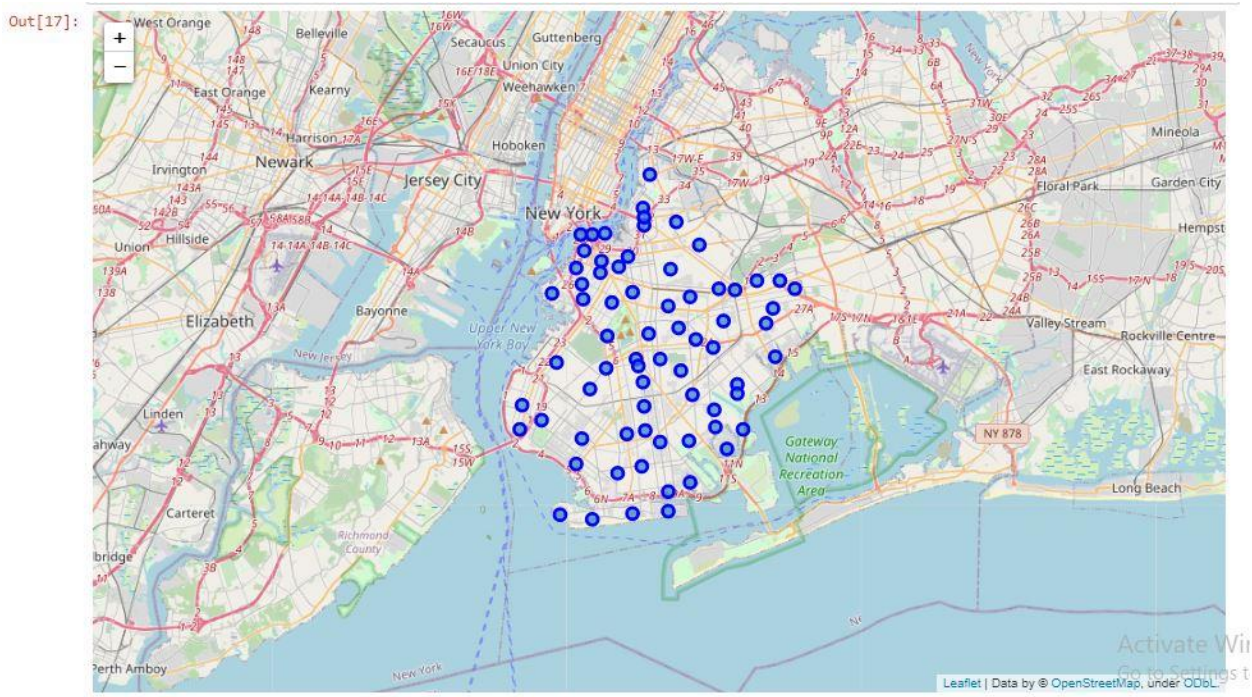
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meters.

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Name of the venue e.g. the name of a store or restaurant
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Rent data
This will contain the average rent of neighborhoods in Brooklyn by postcode

Map of Brooklyn



Methodology Section

Clustering Approach:

To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York and

Toronto. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

Using K-Means Clustering Approach

```
In [39]: neighborhoods_venues_sorted.insert(0, 'Clustering Label', kmeans.labels_)

         brooklyn_merged = brooklyn_data

         brooklyn_merged = brooklyn_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

         brooklyn_merged.head()
```

Out[39]:

	Borough	Neighborhood	Latitude	Longitude	Clustering Label	Clustering Labels1	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7 C
0	Brooklyn	Bay Ridge	40.625801	-74.030621	0.0	0.0	Yoga Studio	Thai Restaurant	Café	Spa	Lounge	Caucasian Restaurant	Su Re
1	Brooklyn	Bensonhurst	40.611009	-73.995180	1.0	1.0	Park	Yoga Studio	Creperie	Deli / Bodega	Department Store	Dessert Shop	Di
2	Brooklyn	Sunset Park	40.645103	-74.010316	0.0	0.0	Pizza Place	Bagel Shop	Pharmacy	Mobile Phone Shop	Fried Chicken Joint	Latin American Restaurant	Ba
3	Brooklyn	Greenpoint	40.730201	-73.954241	0.0	0.0	Butcher	Grocery Store	Vegetarian / Vegan Restaurant	Polish Restaurant	Arts & Crafts Store	Italian Restaurant	Fr Ve St
4	Brooklyn	Gravesend	40.595260	-73.973471	0.0	0.0	Lounge	Yoga Studio	Food Truck	Deli / Bodega	Department Store	Dessert Shop	Di

```
In [40]: del brooklyn_merged["Clustering Labels1"]
```


Most Common venues near Neighborhood

In [27]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

columns = ['Neighborhood']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Neighborhood'] = brooklyn_grouped['Neighborhood']

for ind in np.arange(brooklyn_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(brooklyn_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted.head()
```

Out[27]:

	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	Yoga Studio	Thai Restaurant	Café	Spa	Lounge	Caucasian Restaurant	Sushi Restaurant	Chinese Restaurant	Pool Hall	Pizza Place
1	Bedford Stuyvesant	Gym / Fitness Center	Grocery Store	Bus Station	Yoga Studio	Dry Cleaner	Fast Food Restaurant	Farmers Market	Falafel Restaurant	Dumpling Restaurant	Discount Store
2	Bensonhurst	Park	Yoga Studio	Creperie	Deli / Bodega	Department Store	Dessert Shop	Diner	Discount Store	Dog Run	Dry Cleaner
3	Boerum Hill	Bakery	Yoga Studio	Food Truck	Deli / Bodega	Department Store	Dessert Shop	Diner	Discount Store	Dog Run	Dry Cleaner
4	Brighton Beach	Coffee Shop	Diner	Liquor Store	Restaurant	Sushi Restaurant	Food & Drink Shop	Optical Shop	Bookstore	Dessert Shop	Varenyky restaurant

?context=cpdaas

Work Flow:

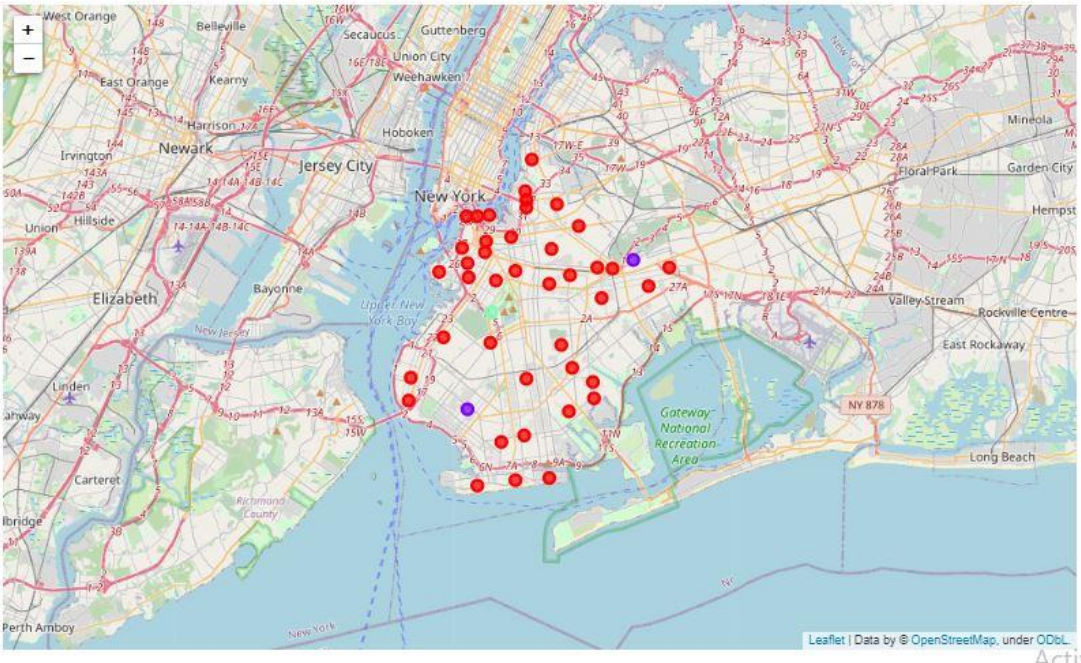
Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Results Section

Map of Clusters in Brooklyn

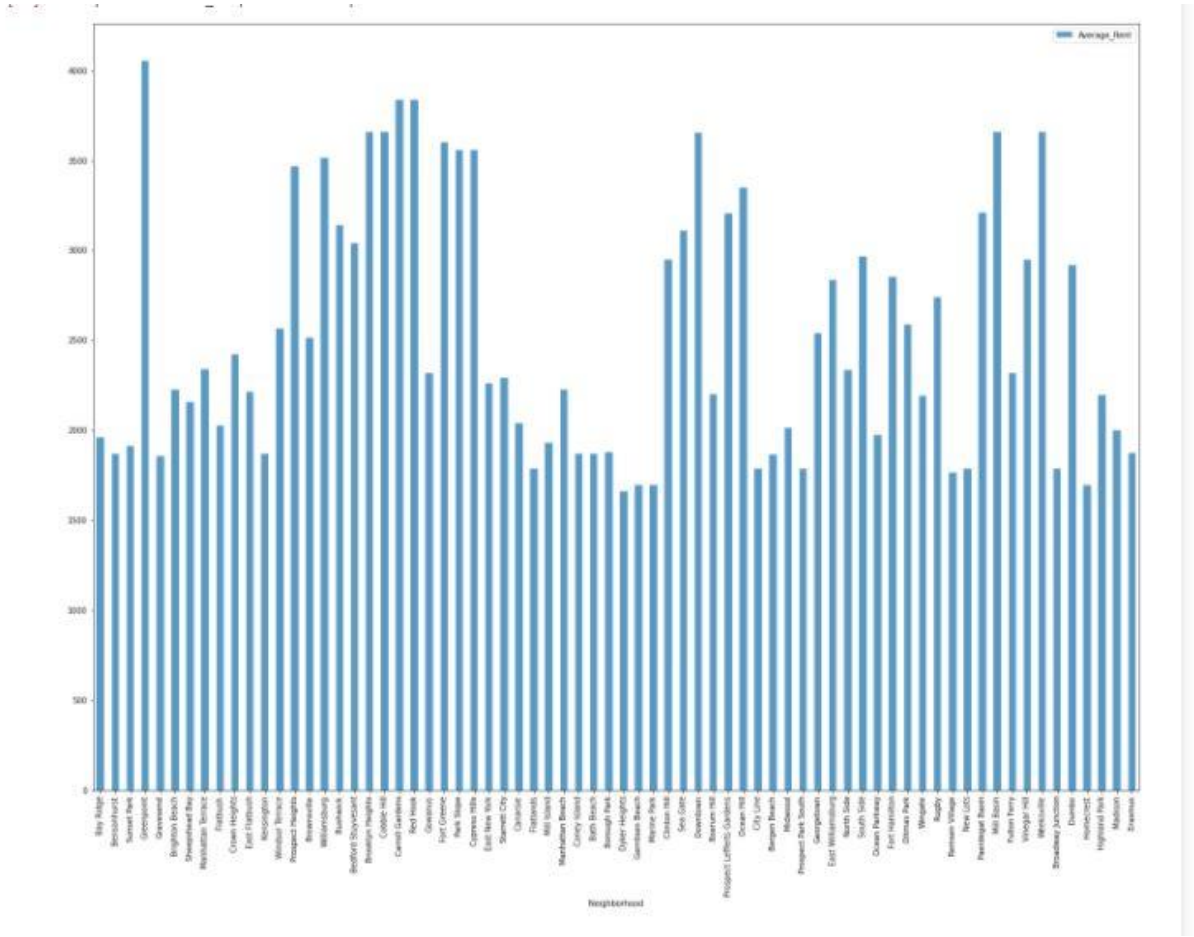
```
map_clusters
```

Out[49]:




```
In [50]: brooklyn_merged.loc[brooklyn_merged['Clustering Label'] == 0, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape
[1]))]]
```


Average Rent by Clusters in Brooklyn



Foursquare API:

This project have used Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

Discussion Section

Problem Which We Tried to Solve

The major purpose of this project, is to suggest a better neighborhood in a new city for the person who will be moving there. Social presence in society in terms of likeminded people. Connectivity to the airport, bus stand, city center, markets and other daily needs things nearby. Sorted list of house in terms of average rent prices.

Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 3 different clusters and for 128 different latitude and longitude from dataset, which have very-similar neighborhoods around them. Using the charts above results presented to a particular neighborhood based on average house prices and school rating have been made.

I feel rewarded with the efforts and believe this course with all the topics covered is well worthy of appreciation. This project has shown me a practical application to resolve a real situation that has impacting personal and financial impact using Data Science tools. The mapping with Folium is a very powerful technique to consolidate information and make the analysis and decision better with confidence.

Future Work*:*

This project can be continued for making it more precise in terms to find best house in Scarborough. Best means on the basis of all required things (daily needs or things we need to live a better life) around and also in terms of cost effective.

Libraries which were used to develop the Project:
Pandas: For creating and manipulating data frames.
Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
Scikit Learn: For importing k-means clustering.
JSON: Library to handle JSON files.
XML: To separate data from presentation and XML stores data in plain text format.
Geocoder: To retrieve Location Data.
Beautiful Soup and Requests: To scrap and library to handle http requests.
Matplotlib: Python Plotting Module.