Final Report - Capstone Project - The Battle of Neighborhoods :

The reasons for running business in Brooklyn, NY

1. Introduction

The purpose of this Project is to help people in exploring better 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, NY.

Lots of people are migrating to various states of USA and needed lots of research for exploring jobs snd doing business for their living. This project is for those people who are looking for better neighborhoods. For ease of accessing to Cafe, School, Super market, medical shops, grocery shops, mall, theatre, hospital, like minded people, etc.

This Project aim to create an analysis of features for a people migrating to Brooklyn to search a best neighborhood as a comparative analysis between neighborhoods. 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.

2. Data Section

Data Link : https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Foursquare API Data:

We will need data about 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.

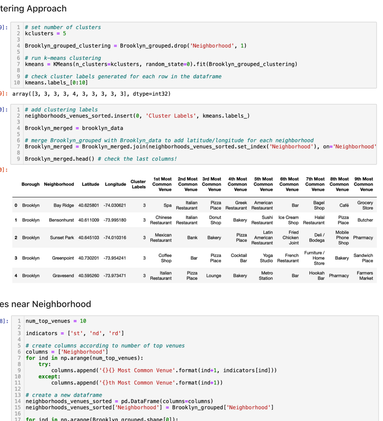
After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be 100 meter.

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:

- Neighborhood
- Neighborhood Latitude
- Neighborhood Longitude
- Venue
- Name of the venue e.g. the name of a store or restaurant
- Venue Latitude
- Venue Longitude
- Venue Category

Map of Brooklyn

```
In [14]:
map_brooklyn
```



3. Methodology Section

Clustering Approach:

We decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods in a big city like New York. 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 [29]:
# set number of clusters
kclusters = 5

Brooklyn_grouped_clustering = Brooklyn_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(Brooklyn_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
Out[29]: array([3, 3, 3, 3, 4, 3, 3, 3, 3, 3], dtype=int32)
```

```
In [30]:
# add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Labels', kmeans.labels_)

Brooklyn_merged = brooklyn_data

# merge Brooklyn_grouped with Brooklyn_data to add latitude/longitude for each neighborhood
Brooklyn_merged = Brooklyn_merged.join(neighborhoods_venues_sorted.set_index('Neighborhood'), on='Neighborhood')

Brooklyn_merged.head() # check the last columns!
```

Out[30]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Brooklyn	Bay Ridge	40.625801	-74.030621	3	Spa	Italian Restaurant	Pizza Place	Greek Restaurant	American Restaurant	Bar	Bagel Shop	Cafe	Grocery Store	Hookah Bar
1	Brooklyn	Bensonhurst	40.611009	-73.995180	3	Chinese Restaurant	Italian Restaurant	Donut Shop	Bakery	Sushi Restaurant	Ice Cream Shop	Halal Restaurant	Pizza Place	Butcher	Pet Store
2	Brooklyn	Sunset Park	40.645103	-74.010316	3	Mexican Restaurant	Bank	Bakery	Pizza Place	Fried Chicken Joint	Latin American Restaurant	Deli / Bodega	Middle Eastern Restaurant	Pharmacy	Gym
3	Brooklyn	Greenpoint	40.730201	-73.954241	3	Coffee Shop	Bar	Pizza Place	Cocktail Bar	French Restaurant	Furniture / Home Store	Yoga Studio	Bakery	Sandwich Place	Deli / Bodega
4	Brooklyn	Gravesend	40.596260	-73.973471	3	Italian Restaurant	Pizza Place	Lounge	Bakery	Metro Station	Bar	Hookah Bar	Pharmacy	Farmers Market	Mexican Restaurant

Most Common venues near Neighborhood

```
In [28]:
num_top_venues = 10

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

# create columns according to number of top venues
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))

# create a new dataframe
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_to

neighborhoods_venues_sorted.head()
```

Out[28]:

	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	Bath Beach	Chinese Restaurant	Pizza Place	Bubble Tea Shop	Sushi Restaurant	Gas Station	Pharmacy	Italian Restaurant	Donut Shop	Fast Food Restaurant	Asian Restaurant
1	Bay Ridge	Spa	Italian Restaurant	Pizza Place	Greek Restaurant	American Restaurant	Bar	Bagel Shop	Cafe	Grocery Store	Hookah Bar
2	Bedford Stuyvesant	Bar	Coffee Shop	Cafe	Pizza Place	Bus Stop	Bus Station	Fruit & Vegetable Store	Fried Chicken Joint	Gift Shop	Gourmet Shop
3	Bensonhurst	Chinese Restaurant	Italian Restaurant	Donut Shop	Bakery	Sushi Restaurant	Ice Cream Shop	Halal Restaurant	Pizza Place	Butcher	Pet Store
4	Bergen Beach	Harbor / Marina	Baseball Field	Playground	Athletics & Sports	Park	Other Nightlife	Pakistani Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Gym

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.

4. Results Section

Map of Clusters in Brooklyn

```
In [32]:
map_clusters
```


Examine Clusters

Cluster 1

```
In [33]:
Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 0, Brooklyn_merged.columns[[1] + list(range(5, Brookly
```

Out[33]:

	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
28	Canarsie	Asian Restaurant	Caribbean Restaurant	Gym	Thai Restaurant	Food	Yoga Studio	Outdoors & Recreation	Outdoor Gym	Other Repair Shop	Other Nightlife
59	Flatbush	Asian Restaurant	Auto Garage	Food	Yoga Studio	Pakistani Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Gym	Other Repair Shop	Other Nightlife

Cluster 2

```
In [34]:
Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 1, Brooklyn_merged.columns[[1] + list(range(5, Brookly
```

Out[34]:

	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	Locksmith	Yoga Studio	Other Great Outdoors	Pakistani Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Gym	Other Repair Shop	Other Nightlife

Cluster 3

```
In [36]:
Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 2, Brooklyn_merged.columns[[1] + list(range(5, Brookly
```

Out[36]:

	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
12	Windsor Terrace	Deli / Bodega	Pizza	Grocery Store	Diner	Park	Ice Cream Shop	Bookstore	Thrift / Vintage Store	Beer Store	Butcher
26	East New York	Deli / Bodega	Fast Food Restaurant	Pizza Place	Spanish Restaurant	Bus Stop	Caribbean Restaurant	Salon / Barbershop	Event Service	Convenience Store	Fried Chicken Joint
35	Dyker Heights	Burger Joint	Bagel Shop	Pizza	Playground	Golf Course	Yoga Studio	Other Great Outdoors	Outdoors & Recreation	Outdoor Gym	Other Repair Shop
43	Ocean Hill	Deli / Bodega	Supermarket	Bakery	Southern / Soul Food Restaurant	Bus Stop	Playground	Food	Convenience Store	Mexican Restaurant	Chinese Restaurant

Cluster 4

```
In [37]:
Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 3, Brooklyn_merged.columns[[1] + list(range(5, Brookly
```

Out[37]:

	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	Spa	Italian Restaurant	Pizza Place	Greek Restaurant	American Restaurant	Bar	Bagel Shop	Cafe	Grocery Store	Hookah Bar
1	Bensonhurst	Chinese Restaurant	Italian Restaurant	Donut Shop	Bakery	Sushi Restaurant	Ice Cream Shop	Halal Restaurant	Pizza Place	Butcher	Pet Store
2	Sunset Park	Mexican Restaurant	Bank	Bakery	Pizza Place	American Restaurant	Fried Chicken Joint	Deli / Bodega	Middle Eastern Restaurant	Pharmacy	Gym
3	Greenpoint	Coffee Shop	Bar	Pizza Place	Cocktail Bar	Yoga Studio	French Restaurant	Furniture / Home Store	Bakery	Sandwich Place	Deli / Bodega
4	Gravesend	Italian Restaurant	Pizza Place	Lounge	Bakery	Metro Station	Bar	Hookah Bar	Pharmacy	Farmers Market	Men's Store
9	Brighton Beach	Russian Restaurant	Donut Shop	Eastern European Restaurant	Beach	Restaurant	Sushi Restaurant	Pharmacy	Gourmet Shop	Bank	Mobile Phone Shop

cluster 5

```
In [38]:
Brooklyn_merged.loc[Brooklyn_merged['Cluster Labels'] == 4, Brooklyn_merged.columns[[1] + list(range(5, Brookly
```

Out[38]:

	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
45	Bergen Beach	Harbor / Marina	Baseball Field	Playground	Athletics & Sports	Park	Other Nightlife	Pakistani Restaurant	Outlet Store	Outdoors & Recreation	Outdoor Gym

The Location:

Brooklyn is a popular destination for new immigrants in USA to reside. As a result, it is one of the most diverse and multicultural areas in the Greater Brooklyn Area, being home to various religious groups and places of worship. Although immigration has become a hot topic over the past few years with more governments seeking more restrictions on immigrants and refugees, the general trend of immigration into USA has been one of on the rise.

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.

5. Discussion Section

Brooklyn is considered one of the most famous cities in New York City. Each year, there are millions of tourists. Therefore, in different areas of the city, it is important to analyze the potential to open a business. But because there is still a lot of information about the environment. both in terms of cost of living and competition Therefore, it is not yet clear what the outcome of the business should be and where.

6. Conclusion Section

In this project, using k-means cluster algorithm I separated the neighborhood into 10(Ten) different clusters and for 103 different lattitude and logitude 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.