The Battle of Neighborhoods

Veer Abhimanyu Singh

1. Introduction

1.1 Background

New York City (NYC), often called New York (NY), is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles (784 km^2), New York is also the most densely populated major city in the United States. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. It is home to the headquarters of the United Nations, New York is an important center for international diplomacy.

New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that is among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park. Broadway Theater is staged in neon-lit Times Square.

Many tech companies have their offices and headquarters situated in the Bay Area. If someone wants to start a restaurant business in this area, what should be the approach to start the same? There are already many restaurant chains and near the offices of almost every tech company. Therefore, how can someone address this problem?

1.2 Business Problem

New York city is a business center at world level. Several people have dream to start a business here and grow rapidly. As we know that New York comprises of 5 boroughs. In this capstone project, the main objective would be:

- Find the best spot for the next profitable restaurant business in Brooklyn.
- This Area has a lot of restaurants with different cuisines including Mexican, Italian, Indian etc
- Starting a restaurant without analysis won't yield good results

1.3 Interest

The project is beneficial to anyone who is looking to start a new restaurant business in Brooklyn area. The audience is quite large and may include stakeholders like: Businessman, Tourists.

2. Data acquisition and cleaning

2.1 Data sources

The data used for the capstone:

New York city data from the Json file available in the course. This file comprises of details of all boroughs with their corresponding Neighborhoods. Also, addition of their coordinates helps much throughout.

Four Square API is being used to get venues details for neighborhoods in New York. We will use explore query, search query and other ways to fully utilize the API functionalities.

2.2 Methodology Used:

- The data was extracted and cleaned, and the feature selection was made
- Map for all the offices in the bay area was plotted
- Foursquare API was used to figure out the nearest restaurants
- The restaurants were classified using KMeans ML algorithm
- The clustered restaurants were plotted along with the companies in the Bay Area

3. Exploratory Data Analysis

The data included longitude and latitude data for all 05 boroughs for New York.

A glimpse of the data downloaded in JSON format:

```
Out[7]: {'geometry': {'coordinates': [-73.84720052054902, 40.89470517661],
           'type': 'Point'},
          'geometry_name': 'geom',
          'id': 'nyu_2451_34572.1',
          'properties': {'annoangle': 0.0,
           'annoline1': 'Wakefield',
           'annoline2': None,
           'annoline3': None,
           'bbox': [-73.84720052054902,
            40.89470517661,
            -73.84720052054902,
            40.89470517661],
           'borough': 'Bronx',
           'name': 'Wakefield',
           'stacked': 1},
          'type': 'Feature'}
```

We cleaned the data and after featuring engineering, we got the dataframe in the following format:

Out[15]:

	Borough	Neighborhood	Latitude	Longitude
0	Brooklyn	Bay Ridge	40.625801	-74.030621
1	Brooklyn	Bensonhurst	40.611009	-73.995180
2	Brooklyn	Sunset Park	40.645103	-74.010316
3	Brooklyn	Greenpoint	40.730201	-73.954241
4	Brooklyn	Gravesend	40.595260	-73.973471

The dataframe included Borough, Neighborhood, Latitude and Longitude data. We filtered this data for only Brooklyn area for further analysis. Then, we explored neighborhood data for Brooklyn using FourSquare API.

In [28]:
```
brooklyn_venues = getNearbyVenues(names=brooklyn_data['Neighborhood'],
                                  latitudes=brooklyn_data['Latitude'],
                                  longitudes=brooklyn_data['Longitude'])
```

```
Bay Ridge
Bensonhurst
Sunset Park
Greenpoint
Gravesend
Brighton Beach
Sheepshead Bay
Manhattan Terrace
Flatbush
Crown Heights
East Flatbush
Kensington
Windsor Terrace
```

Then we looked at the number of venues for each neighborhood:

In [29]:
```
brooklyn_venues.groupby('Neighborhood').count()
```

Out[29]:

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bath Beach	50	50	50	50	50	50
Bay Ridge	91	91	91	91	91	91
Bedford Stuyvesant	26	26	26	26	26	26
Bensonhurst	29	29	29	29	29	29
Bergen Beach	6	6	6	6	6	6
Boerum Hill	84	84	84	84	84	84
Borough Park	19	19	19	19	19	19
Brighton Beach	44	44	44	44	44	44

We found about 286 unique categories. We performed one-hot encoding for all these categories. Then we looked at each neighborhood along with the top 5 most common venues.

```
----Bay Ridge----
                venue  freq
0                 Spa  0.07
1   Italian Restaurant  0.07
2         Pizza Place  0.05
3           Bagel Shop  0.03
4  American Restaurant  0.03


----Bedford Stuyvesant----
           venue  freq
0    Coffee Shop  0.12
1            Bar  0.08
2    Pizza Place  0.08
3  Deli / Bodega  0.08
4           Café  0.08
```

Based on this data, created new dataframe and looked at the top 10 venues for each neighborhood:

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
0	Bath Beach	Donut Shop	Pharmacy	Pizza Place	Sushi Restaurant	Kids Store	Italian Restaurant	Fast Food Restaurant	Women's Store	Chinese Restaurant	Mobile Phone Shop
1	Bay Ridge	Spa	Italian Restaurant	Pizza Place	Bagel Shop	American Restaurant	Bar	Greek Restaurant	Grocery Store	Ice Cream Shop	Sandwich Place
2	Bedford Stuyvesant	Coffee Shop	Pizza Place	Deli / Bodega	Café	Bar	BBQ Joint	Discount Store	Park	Other Repair Shop	Cocktail Bar
3	Bensonhurst	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Chinese Restaurant	Dessert Shop	Bakery	Bagel Shop	Noodle House	Liquor Store	Grocery Store
4	Bergen Beach	Harbor / Marina	Baseball Field	Athletics & Sports	Playground	Donut Shop	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flower Shop	Food
5	Boerum Hill	Coffee Shop	Bar	Dance Studio	Cosmetics Shop	Yoga Studio	Clothing Store	Middle Eastern Restaurant	Martial Arts Dojo	Spa	Kids Store

Then used this dataframe for building a clustering algorithm.

4. Machine Learning Algorithm

Ran k-means to cluster the neighborhood into 5 clusters as we were getting the optimal number of clusters as 05 based on Elbow method.

Cluster 01: The Italian Restaurant were the most common venue, followed by Pizza Place and Chinese Restaurant.

Cluster 1

In [41]:
```
brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 0, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]
```

Out[41]:

	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
4	Gravesend	Italian Restaurant	Metro Station	Pizza Place	Chinese Restaurant	Lounge	Spa	Bakery	Bus Station	Music Store	Men's Store

Cluster 02: There were multiple neighborhoods in the cluster 02, with Italian, Chinese, Mexican and Caribbean Restaurants were the most common venue, followed by Pizza Place and Fast Food Restaurant.

Cluster 2

```
In [42]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 1, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1])
```

Out[42]:

	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	Bagel Shop	American Restaurant	Bar	Greek Restaurant	Grocery Store	Ice Cream Shop	Sandwich Place
6	Sheepshead Bay	Dessert Shop	Turkish Restaurant	Restaurant	Sandwich Place	Yoga Studio	Pizza Place	Department Store	Creperie	Russian Restaurant	Café
7	Manhattan Terrace	Pizza Place	Chinese Restaurant	Ice Cream Shop	Donut Shop	Grocery Store	Japanese Restaurant	Cosmetics Shop	Nightclub	Steakhouse	Bank
8	Flatbush	Mexican Restaurant	Coffee Shop	Caribbean Restaurant	Pharmacy	Chinese Restaurant	Laundromat	Bank	Bagel Shop	Lounge	Middle Eastern Restaurant
10	East Flatbush	Department Store	Pharmacy	Supermarket	Caribbean Restaurant	Fast Food Restaurant	Hardware Store	Park	Moving Target	Wine Shop	Deli / Bodega
12	Windsor Terrace	Deli / Bodega	Italian Restaurant	Plaza	Diner	Park	Sushi Restaurant	Coffee Shop	Grocery Store	Bookstore	Bar

Cluster 03: The Caribbean Restaurant were the most common venue, followed by Fast Food and Pizza Place.

Cluster 3

```
In [43]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 2, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1])
```

Out[43]:

	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
47	Prospect Park South	Caribbean Restaurant	Fast Food Restaurant	Mobile Phone Shop	Grocery Store	Pizza Place	Pharmacy	Clothing Store	Donut Shop	Chinese Restaurant	Latin American Restaurant

Cluster 04: The Italian Restaurants were the most common venue, followed by Pizza Place, Sushi and Chinese Restaurants.

Cluster 4

```
In [44]: brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 3, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1])
```

Out[44]:

	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
1	Bensonhurst	Italian Restaurant	Sushi Restaurant	Ice Cream Shop	Chinese Restaurant	Dessert Shop	Bakery	Bagel Shop	Noodle House	Liquor Store	Grocery Store
2	Sunset Park	Bakery	Latin American Restaurant	Mexican Restaurant	Bank	Pizza Place	Gym	Mobile Phone Shop	Pharmacy	Sandwich Place	Supplement Shop
3	Greenpoint	Bar	Pizza Place	Boutique	Cocktail Bar	Mexican Restaurant	Café	Record Shop	Coffee Shop	Bakery	Yoga Studio
5	Brighton Beach	Restaurant	Beach	Eastern European Restaurant	Russian Restaurant	Gourmet Shop	Sushi Restaurant	Bank	Fast Food Restaurant	Mobile Phone Shop	Pharmacy
9	Crown Heights	Pizza Place	Grocery Store	Museum	Café	Playground	Supermarket	Farmers Market	Coffee Shop	Sushi Restaurant	Bakery
11	Kensington	Grocery Store	Sandwich Place	Ice Cream Shop	Pizza Place	Thai Restaurant	Restaurant	Japanese Restaurant	Outdoors & Recreation	Mobile Phone Shop	Mexican Restaurant
13	Prospect Heights	Bar	Cocktail Bar	Mexican Restaurant	Bakery	Gourmet Shop	Pizza Place	Thai Restaurant	Wine Shop	Southern / Soul Food Restaurant	Coffee Shop

Cluster 05: The Italian Restaurant were the most common venue, followed by Pizza Place.

Cluster 5

In [45]:
```
brooklyn_merged.loc[brooklyn_merged['Cluster Labels'] == 4, brooklyn_merged.columns[[1] + list(range(5, brooklyn_merged.shape[1]))]
```

Out[45]:

	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
23	Fort Greene	Italian Restaurant	Flower Shop	Pizza Place	Coffee Shop	Yoga Studio	Spa	Mexican Restaurant	French Restaurant	Cocktail Bar	New American Restaurant

5. Conclusions

Restaurants were among the most common venue among most of the neighborhoods. However, there were a number of neighborhood where restaurants were not among the top 03 common venue, like: Kensington and East Flashbush.

For East Flashbush: Departmental Stores, Pharmacy and Supermarket were among the top 03 most common venue.

For Kensington: Grocery Store, Sandwich Place and Ice Cream Shop were among the top 03 most common venue.

Both the places look as a good option to start a new restaurant, however East Flashbush looks a better option as there is no food related option among the top 03. Therefore, we can conclude as East Flashbush as the best neighborhood to start a new restaurant.

6. Future directions

We will need to look into more details for the neighborhood, like population density, residential or office area, number of offices, number of apartments, etc. These additional points will help us to make more efficient decision.