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'}
```

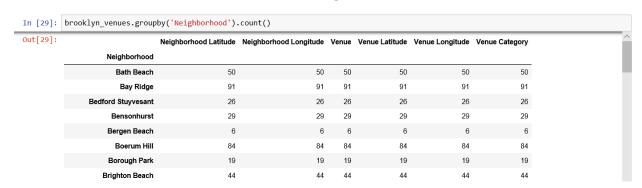
We cleaned the data and after featuring engineering, we got the dataframe in the following format:

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:



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
          Café 0.08
```

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

10th Most Common Venue	9th Most Common Venue	8th Most Common Venue	7th Most Common Venue	6th Most Common Venue	5th Most Common Venue	4th Most Common Venue	3rd Most Common Venue	2nd Most Common Venue	1st Most Common Venue	Neighborhood	:
Mobile Phone Shop	Chinese Restaurant	Women's Store	Fast Food Restaurant	Italian Restaurant	Kids Store	Sushi Restaurant	Pizza Place	Pharmacy	Donut Shop	Bath Beach	0
Sandwich Place	Ice Cream Shop	Grocery Store	Greek Restaurant	Bar	American Restaurant	Bagel Shop	Pizza Place	Italian Restaurant	Spa	Bay Ridge	1
Cocktail Bar	Other Repair Shop	Park	Discount Store	BBQ Joint	Bar	Café	Deli / Bodega	Pizza Place	Coffee Shop	Bedford Stuyvesant	2
Grocery Store	Liquor Store	Noodle House	Bagel Shop	Bakery	Dessert Shop	Chinese Restaurant	Ice Cream Shop	Sushi Restaurant	Italian Restaurant	Bensonhurst	3
Food	Flower Shop	Fish Market	Fish & Chips Shop	Filipino Restaurant	Donut Shop	Playground	Athletics & Sports	Baseball Field	Harbor / Marina	Bergen Beach	4
Kids Store	Spa	Martial Arts Dojo	Middle Eastern Restaurant	Clothing Store	Yoga Studio	Cosmetics Shop	Dance Studio	Bar	Coffee Shop	Boerum Hill	5

Then used this dataframe for building a clustering algorithm.

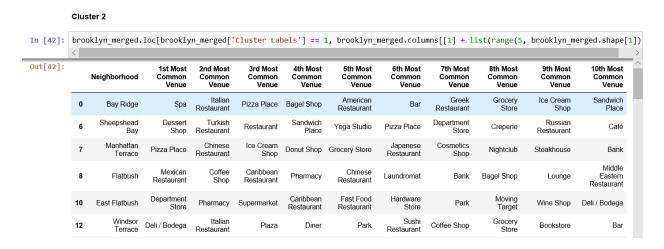
4. Machine Learning Algorithm

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.

	Clus	ter 1											
In [41]:	broo	klyn_merged.	loc[brookly	/n_merged['C	luster Label	s'] == 0, br	ooklyn_merg	ed.columns[[1] + list(range(5, br	ooklyn_merg	ed.shape[1])	
	<											>	
Out[41]:	N	leighborhood	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 03: The Caribbean Restaurant were the most common venue, followed by Fast Food and Pizza Place.



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



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



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.