Capstone Project – The Battle of Neighbourhoods

Introduction **to the Problem**

New York City is, demographically speaking one of the largest cities in the world and the largest in the United States of America. Home to over 18.4 million people.

Over the last decade the city has experienced exponential economic growth to an estimated $1.751 trillion GDP. Over 8% of the country's total GDP, despite the country only accounting for 0.008% of the country's square miles. Manhattan, the City's economic hub accounting for a tiny 0.001%.

This extreme competition for space has saw the cities expansion and "gentrification" of other boroughs within the City. For the sake of this study we will be looking at Brooklyn. Which is the cities 2nd largest and most affected by the aforementioned expansions due to its close proximity to Manhattan and its convenient transport links, in particular with JFK airport.

We will use this project to try and assist in common issues that people have when moving to / visiting lesser known or up and coming Boroughs of popular cities;

1. Finding the best attractions available in the area.
2. Identifying great places to start a business.
3. Identifying attraction saturation in individual areas.
4. Identifying potential gaps in the market for potential businesses within each area.

The target stakeholders for this study would be entrepreneurial businesspersons as well as visitors of the city.

Introduction to the *Data* used

In this section I have outlined below the data that was used for this part of the course and the rationale behind those choices;

1. Firstly, we used the New York City data that was available from this course via JSON file. The file consisted of coordinates for all boroughs and their corresponding Neighbourhoods which made it perfect for the purpose of this course.

2. Secondly, we used the Four Square API to extract venue details within the neighbourhoods of New York as the purpose of this study, such details are essential in identifying and extrapolating suggestions and recommendations. In this part of the course we used explorer queries and search queries to fully utilize the powerful API functionality available through Four Square.

***Methodology* behind the project**

Due to size and nature of the data required to extrapolate some informative recommendations in this project a methodology was required to be applied. Within this some exploratory data analysis was applied;

1) From the data frame we will subset the dataset to get only the records for BROOKLYN.
2) Data cleaning with removal of all NaN values plus any blank fields.
3) Visualize all neighborhood of BROOKLYN using FOLIUM library.
4) Fetching Latitude and Longitude of neighborhoods using Geopy Library
5) Normalizing the data fetched from Foursqare API.

In addition to the above, it's worth also mentioning some of the Machine learning techniques that have been utilized in this project. Generally speaking, we will be using clustering techniques of unsupervised learning in order to segment the neighborhoods.

We will be utilizing the K-MEANS algorithm to cluster neighborhoods with similarities in their venues.

Results Section;
When we run this clustering on Brooklyn dataset we get the 5 clusters having neighborhood which are similar based on venues they have nearby. Below are top venues in each neighborhood.

```
----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


----Bensonhurst----
                venue  freq
0  Italian Restaurant  0.10
1             Bakery  0.07
2    Sushi Restaurant  0.07
3        Dessert Shop  0.07
4      Ice Cream Shop  0.07
```

Sample of cluster formed;

Cluster 2

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

2]:

	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 Co
0	Bay Ridge	Spa	Italian Restaurant	Pizza Place	Bagel Shop	American Restaurant	Bar	Greek Restaurant	Grocer Store
6	Sheepshead Bay	Dessert Shop	Turkish Restaurant	Restaurant	Sandwich Place	Yoga Studio	Pizza Place	Department Store	Creper

Discussion Section;

According to results we see that most common venues come out to be restaurants and Coffee shops, which means any new businessman can start restaurants provided that they need to compete with existing established chains but if he provides top class facilities to get to top.

Conclusion;

In this Project we have used Foursquare API with Machine learning techniques to provide the best results in segmenting the neighborhood according to their venues.