Capstone Project

Introduction Section – Business Problem

New York city is much bigger trade center and dream to lot of growing business and people. We eventually know that New York comprises of 5 boroughs and in this project, we will try to explore BROOKLYN and what it has to offer to the growing city.

Problem Description and Analysis –

Every new person who enter the land of New York has lot of choices to hang in to enjoy be it cafe's, parks or hotels or some come for businesses. Below are some points –

1) Even with new age smartphone people still find it difficult to get choices of best attraction available nearby.
2) People who come to start their new businesses often find it difficult to locate the best place they can establish their startup and we will try to find appropriate location.
3) If every place is bifurcated by its common venues it become easy to know where to find best things.

Stakeholders – Businessman, Tourists.

Data Section-

The data used are –

1) 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.
2) 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.

Methodology Section –

In this project New York data is taken as a base which contains info of all borough and their corresponding neighborhoods. Some exploratory data analysis applied –

1) From the dataframe 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.

Machine learning technique –

We will use Clustering technique of Unsupervised Learning to segment the neighborhoods and cluster them using K-MEANS clustering to get the clusters 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.