The Battle of the Neighbourhoods - Report

Introduction & Business Problem:

Problem Background:
The City of New York is the most attractive city in the US. It is diverse with good business opportunities. All walks of industry swarm to obtain their spot in this major city, meaning it also very competitive.

Problem Description:
A friend of mine decided to open his second restaurant in New York. However, he seeks some consultant insight on an ideal location and seek my help to run some analysis to support on it.

We will be looking into various angle of data to study and understand better about New York city in term of the population, competitors, the surrounding environment and also the ingredient convenience supply.

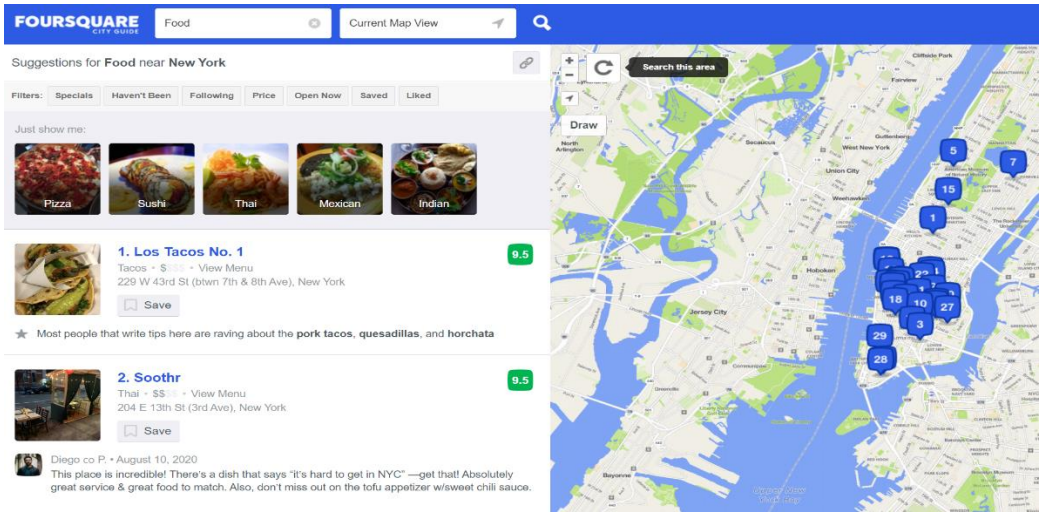
Data Source:

We will be analysing some of the data for this Project.

Foursquare API Data:

https://foursquare.com/city-guide

Foursquare provide various information of different venues within the neighbourhood. The data include location latitude and longitude, venue, explore details, comments from visitor and photos.



NYU Spatial Data Repository:

https://geo.nyu.edu/catalog/nyu_2451_34572

This New York City Neighbourhood name point file was created as a guide to New York City's neighbourhood that appear on web resources.

Methodology:

To obtain an optimum location in New York City for the new restaurant

Analysis Approach:

We will load the data json file and apply with python language. The data frame consists of the coordinate from both NYU SDR and Foursquare. We will apply geopy, folium, pandas, numpy and random. We will finally visualize the neighbourhood and recommendation.

```
Scarborough_grouped_clustering = Scarborough_grouped.drop('Neighborhood', 1) # delet column
kmeans = KMeans(n_clusters=3, random_state=0).fit(Scarborough_grouped_clustering) # divide into 3 group
kmeans.labels_
```

```
array([2, 2, 0, 0, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 0, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 0, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 0, 2, 2, 2, 2, 2, 0, 0, 2, 2, 2, 2, 2, 2, 2, 0, 2, 2, 2,
       0, 2, 2, 2, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,
       2, 2, 2, 2, 2, 2, 2, 2, 2, 0], dtype=int32)
```

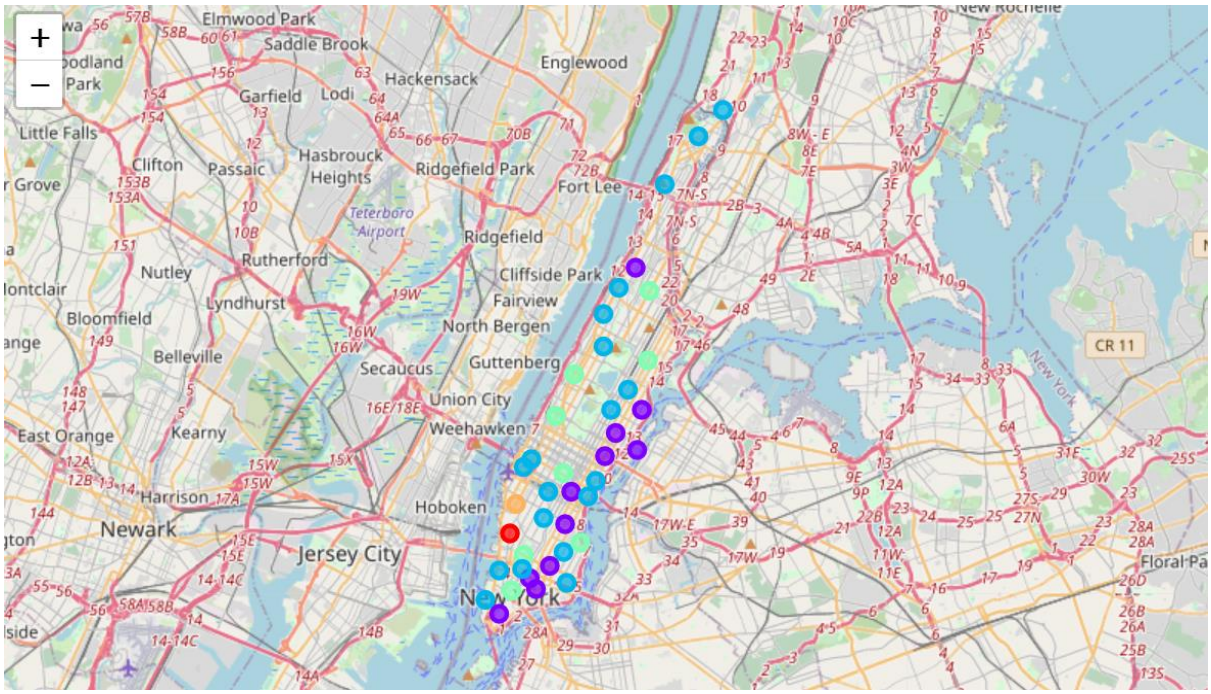
```
columns = ['Neighborhood']
for ind in np.arange(num_top_venues):  #
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

neighborhoods_venues_sorted = pd.DataFrame(columns=columns) # new a data table
neighborhoods_venues_sorted['Neighborhood'] = Scarborough_grouped['Neighborhood']

for ind in np.arange(Scarborough_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(Scarborough_grouped.iloc[ind, :], num_top_venues)
    # reconstruct the data structure from the largeset number to the least

neighborhoods_venues_sorted.head()
```

Examining the cluster.



Results:

As per our finding from most of the neighbourhood having common venues of grocery stores and restaurants. Manhattan is a preferable location to open the restaurant from the dense population and better development neighbourhood. There are still other factors to consider such as shop cost and actual surrounding.

Discussion:

My friend will be able to make his decision and narrow down the option.

Conclusion:

Every neighbourhood has a unique attraction. Bigger crowd will offer better opportunities for the restaurant to prosper