Capstone Project: The Battle of Neighborhoods

Location Analysis and Recommendation

Introduction(Business Problem)

The ideology behind the problem taken into consideration is basically for people moving into a new city so that they can find their best suited area of preference for finding a stay. Based on the requirements of the client, the analysis is performed and a location is found with the maximum avaibility of these requirements. This is to help these new people to settle and get accustomed to the new city.

The problem taken at hand is explained below:

Problem Statement:

A person wants to find a place for his stay in New York City. The person loves Asian food and hence wants to find a location where it is available. Using Foresquare API, the areas of the city are analyzed and explored to find an area with maximum number of Asian Restaurants. Furthermore, the avaibility of Market Places, Health Care Stores, Shopping Malls ,Salons etc are also to taken into consideration while finding the best location. Finally a list of all venues in the selected neighborhood is given and plotted on the map.

A list of other requirements to be considered for the above problem are:

Essential Requirements:

1. Bank
2. Bakery
3. Pharmacy
4. Market/Store/Shop
5. Salon
6. Health and Beauty Service
7. Gym
8. Park
9. Spa

Non-Essential Requirements:

1. Tea Shop
2. Tea Room

Data:

New York City or NY is the most populous city of the United States of America and is the largest metropolitan area of the world. New York is divided into five Boroughs-Bronx, Brooklyn, Manhattan, Queens and Staten Islands. A Borough is a town or district of administrative unit.

Data Source: The data source from which this is obtained is: : https://geo.nyu.edu/catalog/nyu_2451_34572. This is a free to download dataset available via the link.

Data Description: The dataset consists of four columns namely the Boroughs, their respective neighborhoods and the latitudes and longitudes of each neighborhood.

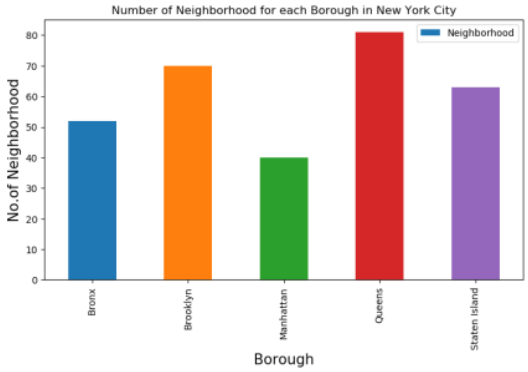
Now to find Asian Restaurants in the city, Foresquare API is used to find all the venues in the city consisting of Asian Restaurants in a radius of 1000 meters and having a limit of 100.The Borough with highest number of Asian Restaurants is found and then further analysis of the neighborhoods of that Borough is done.

Methodology:

1. **Loading and Understanding Data:**

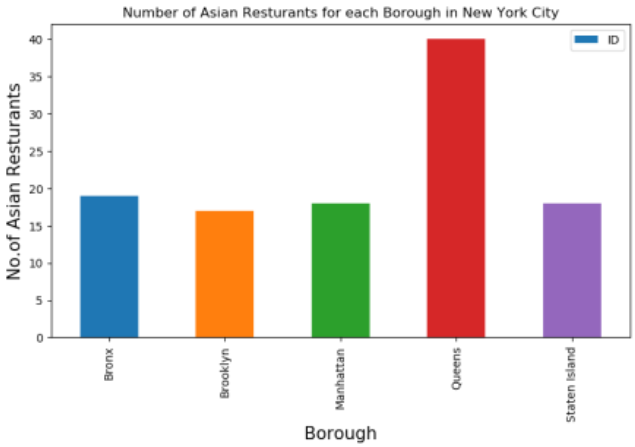
   Firstly we start by importing the libraries and installing if required.Then the data is loaded and analyzed for better understanding.

   - The shape of the dataframe is checked to get the number of rows and columns of the dataframe and the count of neighborhood in each Borough.
   - The latitudes and longitudes of New York City is found using geolocator and then the neighborhoods are printed as clusters on the ma of New York City.
   - The next thing is to visualize the neighborhoods graphically.



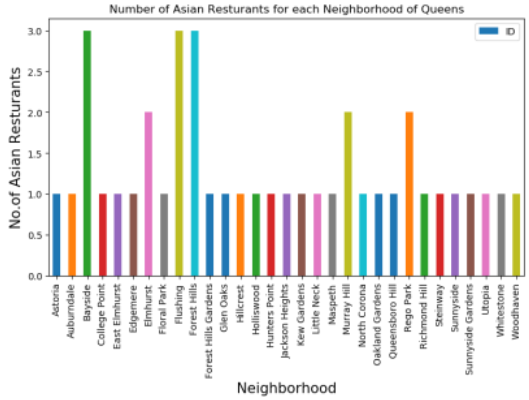
   From the graph,it is seen that Queens has the maximum number of neighborhoods.

2. **Foresquare API:**
   - Now the Client ID and Client Secrest as per the Foresquare API is taken into variables for use.
   - A function get_venues is defined which when called will give the venue details. It includes the latitude and longitude as its parameters.After which the function is called with a condition that the venues being fetched has its category as Asian Restaurants.This is done for the entire New York City.
   - Once a list of Asian Restaurants is available,it is stored in a dataframe and a plot is made between the Boroughs of the city and the number of Asian Restaurants in each Borough.



From the graph ,it is evident that Queens has the maimum number of Asian Restaurants.So further analysis is done for queens.

3. **Exploring Queens:**
    - The dataframe consisting of all Asian Restaurant details is sliced only for the Borough Queens .
    - The original dataframe is also sliced for Queens details only.
    - After the slicing,a graph is plotted between the neighborhoods of Queens and the number of Asian Restaurants to find out the neighborhoods in which maximum number of Asian Restaurants is available in a radius of 1000 meters.



From the graph,it is evident that Bayside,Flushing and Forest Hills have the highest and equal number of neighborhoods.Hence we have to further explore each of these.

4. **Exploring the Neighborhoods:**
    - Three generalised functions are made to be used for analyzing each neighborhood.One is to fetch the “URL” based on the latitudes and longitudes of thelocation.The other is to use the URL to generate the request results.The final one is to get the category type of each venue when called.
5. **Exploring the first Neighborhood-Bayside:**
    - In the sliced dataframe of the original one for queens details,the neighborhood location is found and its latitudes and longitudes as well.The latitudes and

longitudes are passed to get the url from the function and the url is then passed to get the results.

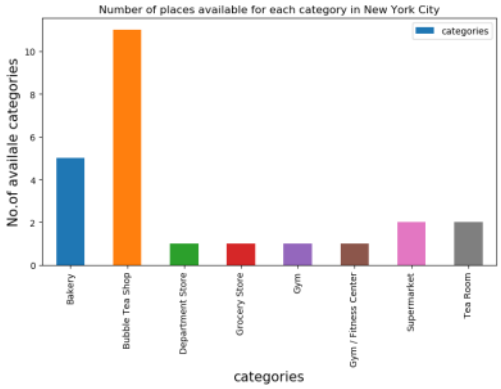
- Once the results are found,a call is madeto get the category types and arrange them into a new dataframe meant for Bayside neighborhood.
- As per the requirements of the person stated in the problem statement,all the requirements are arranged into a new dataframe and a plot is made between the required categories and their counts.

From the graph,the following observations are made:

```
Pharmacy                    4
Bakery                      3
Supplement Shop             2
Bank                        2
Park                        1
Bakery                      3
Supplement Shop             2
Bank                        2
Park                        1
Bubble Tea Shop             1
Health & Beauty Service     1
Gym                         1
Gym / Fitness Center        1
Spa                         1
Tea Room                    1
Salon / Barbershop          1
Bubble Tea Shop             1
Health & Beauty Service     1
Gym                         1
Gym / Fitness Center        1
Spa                         1
Tea Room                    1
Salon / Barbershop          1
```

6. **Exploring Flushing:**

- All the methods as stated for Bayside are followed and the same graph is plotted for it.



From the graph,the following observations are made:

Bubble Tea Shop	11
Bakery	5
Supermarket	2
Tea Room	2
Gym / Fitness Center	1
Grocery Store	1
Department Store	1
Gym	1

7. **Exploring Forest Hills:**

- A graph is plotted like the previous ones:

From the graph,the following observations are made:

Bakery	6
Park	3
Gym / Fitness Center	2
Grocery Store	2
Supermarket	1
Pharmacy	1
Health & Beauty Service	1
Bank	1
Salon / Barbershop	1

8. **Comparison:**
    - From the graphical representations and the observations,a comparison of the neighborhoods are made in correspondence to the requirements mentioned in the problem statement.
    - Although the avaibility of the requirements may be less in count for the chosen neighborhood,the criteria is to find most of the requirements in the given neighborhoods.

Results:

Best Suited Neighborhood Recommendation:

Comparing the avaibility of the requirements in the three neighborhoods,the best suited neighborhood is Forest Hills.It is chosen because despite the less count of a few categories available,it fulfills almost all the requirement list as per the problem for a radius of 1000 meters.

The second recommendation is Bayside.

Further Analysis of Forest Hills:

A list of all the venues in the neighborhood is given.Thereafter,the latitudes and longitudes of Forest Hills are found and all the venues in the neighborhood are clustered on the map.

Discussions:

- It is to be noted that the criterion for comparison may change based on the requirement of the person and so may the city of loctaion.
- For a new city,a new dataset of the city details shall be used while for new requirements,minor changes in the categories can be made.
- The three neighborhoods :Bayside,Flushing and Forest Hills have certain categories which are named differently from the requirements mentione above:
  1)For the requirement of Market/Store/Shop, Forest Hills has a Grocery Store and a Supermarket.
  2)For the requirement of Market/Store/Shop, Flushing has a Grocery Store,a Department Store and a Supermarket. For the requirement of Tea Shop,Flushing has Bubble Tea Shop.
  3)For the requirement of Market/Store/Shop,Bayside has a Supplement Store. For the requirement of Tea Shop,Bayside has Bubble Tea Shop.
- Furthermore,beautification can be added.

Conclusion:

To sum up,the following were done throughout the project:

- Finding out the Asian Restaurants in New York City.
- Finding the Borough with maximum Asian Restaurants.
- Exploring the neighborhoods for highest number of Asian Restaurants for the above result of the Borough.
- Further exploration of the above selected neighborhoods for the other requirements as per the list in the above problem statement.
- Finding the neighborhood which fulfills most of the essential requirements and recommending it.
- Analyzing the recommended neighborhood further to give a list of all venues in the area and make a cluster of them on the map of the neighborhood.