FINAL REPORT

Coursera IBM Data Science Certification

Capstone Project - The Battle of Neighbourhoods

SATYAJEET DAS

Introduction:

Problem Background:

New York City, often called simply New York and abbreviated as NYC, 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 km2), New York is also the most densely populated major city in the United States.[6] Located at the southern tip of the U.S. state of New York, the city is the centre of the New York metropolitan area, the largest metropolitan area in the world by urban landmass.[10] With almost 20 million people in its metropolitan statistical area and approximately 23 million in its combined statistical area, it is one of the world's most populous megacities. New York City has been described as the cultural, financial, and media capital of the world, significantly influencing commerce,[11] entertainment, research, technology, education, politics, tourism, art, fashion, and sports.

New York City is America's classic immigrant city; it was the major historic gateway for the country's eastern and southern European arrivals a century ago and is a major receiving centre today. Its immigrant history, the composition and extraordinary diversity of its current immigrant streams, and its institutions have combined to make it an immigrant city like no other in the United States. New York's contemporary immigrant population stands out for its extraordinary diversity. What is remarkable is the large number from so many different countries. In 2000, the top three groups—Dominicans, Chinese, and Jamaicans—were just under 30 percent of all the foreign-born. No other country accounted for more than five percent, and there were substantial numbers of many West Indian, Latin American, Asian, and European nationalities. The incredible ethnic diversity of New York's immigrants is matched by the heterogeneity of their skills. The mixture of nationalities has ensured a mix of class and occupational origins. In 2000, nearly a quarter of foreign-born New Yorkers age 25 and older had a college degree; at the other end of the spectrum, 35 percent had not completed high school.

New York City, set aside it is the most diverse city of USA, it's also a major point of entry for immigrants; the term "melting pot" was coined to describe densely populated immigrant neighborhoods on the Lower East Side. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world.

With its diverse culture, comes social, economic as well as cuisine diversity.

Problem Description:

- With advent globalisation, people from varied background tend to settle across diverse destinations of the world. We are considering the case of New York City, the most diverse community of the world. Being the economic hub, it attracts people from a varied background. A city with such diversity needs to be explored in detail, this will be done in the first part.
- For the second part, we will analyse three types of restaurants (Italian, Chinese and Indian) across NYC. The ability to eat and the enjoyment of eating are important aspects of a good quality of life. In other words, "being able to eat what I want, when I want" makes us feel good. Hence restaurants are a major factor while selecting a neighbourhood. Apart from that, the major events of life, personal and professional, are celebrated in restaurants. Acquaintances become friends around a table in the safe and controlled environment of a restaurant. Individuals become lovers across a restaurant table, sometimes. The aforementioned facts are some of the major inspiration for analysing restaurants.
- For the third part, we will be analysing the ease of commute across NYC in a simple way. We are assuming the person will use public transport.
  We will analyse the number of **Bus Stops** and **Metro Stations** in the neighbourhood of NYC. The case of cabs is excluded as their service can be acquired in even remote location. But the fact that they are about three to four times costlier as compared to metros and buses, leads us to believe that a normal person will prefer a metro or bus compared to a taxi. It's an inherent tendency of human to save money. And for daily commute buses and metros far too economical compared to their counterpart taxis. Who likes a spendthrift, right!! Nobody will be willing to pay more if other economical and convenient methods are available.

This project will be divided into three parts:

1. We will the exploring New York City
2. Exploring Restaurants around New York City (Italian, Chinese and Indian)
3. Exploring the ease of commute across New York City (Metros and Buses)

Target audience for this project:

- People looking for neighborhood to rent an apartment
- People looking for neighborhood based on cuisine
- People looking to open a restaurant
- People looking to explore the various venues that NYC provides

To find the answers to the following questions:

- Which neighbourhoods can be considered similar in NYC, show them in NYC map?
- Various types of venues in NYC?
- Number of Italian restaurants in NYC?
- what is best location in New York City for Italian Cuisine?
- Number of Chinese restaurants in NYC?
- what is best location in New York City for Chinese Cuisine?
- Number of Indian restaurants in NYC?
- what is best location in New York City for Indian Cuisine?
- what is best location in New York City considering Buses as the mode of transport?
- what is best location in New York City considering Metros as the mode of transport?

Data Section:

For this project uses the following data:

- New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude.
  - Data source : https://cocl.us/new_york_dataset
  - Description: we will use this data set to explore various neighbourhoods of New York City

- Venues in each neighbourhood of New York City.
  - Data source : Foursquare API
  - Description: By using this API we will get all the venues in each neighbourhood. We can filter these venues to get Italian, Indian and Chinese restaurants as well as various transportation system the city offers (Metros and Buses).

- GeoSpace data
  - Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm
  - Description: By using this geo space data we will plot the NYC map with its neighbourhood, NYC map with neighbourhood grouped into clusters.

Methodology:

Part-1

1. We import all the important libraries required.

2. We define a function ‘g_NYCdata’ which will use the URL: https://cocl.us/new_york_dataset , generate the JSON file and finally generate a data frame consisting list of Boroughs, Neighbourhoods along with their latitude and longitude of NYC.

3. We define a function ‘gvenues’ which will use Foursquare API and generate all the venues present in the radius of 750m of each neighbourhood of NYC.
4. Then we import **geocoder** and **folium** for construction of the map of NYC.
5. We plot a map of NYC with its neighbourhood marked.
6. We generate a data set with venue name, categories and their respective latitude and longitude.
7. WE create another data frame consisting of the number of venues in each neighbourhood.
8. We then display all unique venues number and names present in NYC.
9. Next, we categorise the venue using **the one-hot** technique.
10. We display the 15 most common venue of each neighbourhood.
11. We then cluster all neighbourhood of NYC using **K-mean** algorithm.
12. Plot the cluster of neighbourhoods in NYC map.

Part-2

1. In this part, we explore three types of restaurants that NYC provides. We will begin with Italian, followed by Chinese and end with Indian restaurants.
2. First, we will start with Italian restaurants.
3. We collect the number of Italian restaurants in each neighbourhood using the Foursquare API.
4. We create a data frame with Borough, Neighbourhood, ID and Name of Italian restaurants.
5. Create graphs to visualise the number of restaurants in top 5 borough and neighbourhood.
6. We select the borough and neighbourhood having the largest number of Italian restaurants.
7. Next, we will collect the number of Chinese restaurants in each neighbourhood using the Foursquare API.
8. We create a data frame with Borough, Neighbourhood, ID and Name of Chinese restaurants.
9. Create graphs to visualise the number of restaurants in top 5 borough and neighbourhood.
10. We select the borough and neighbourhood having the largest number of Chinese restaurants.
11. Finally, we will collect the number of Indian restaurants in each neighbourhood using the Foursquare API.
12. We create a data frame with Borough, Neighbourhood, ID and Name of Indian restaurants.
13. Create graphs to visualise the number of restaurants in top 5 borough and neighbourhood.
14. We select the borough and neighbourhood having the largest number of Indian restaurants.

Part-3

1. We will explore the ease of commute in NYC by exploring number of metro stations and bus stops in each neighbourhood.
2. Next methods are similar to presented in part-2, we will collect their number in each neighbourhood and subsequently plot graphs.

Result:

1.**Number of neighbourhoods in NYC**

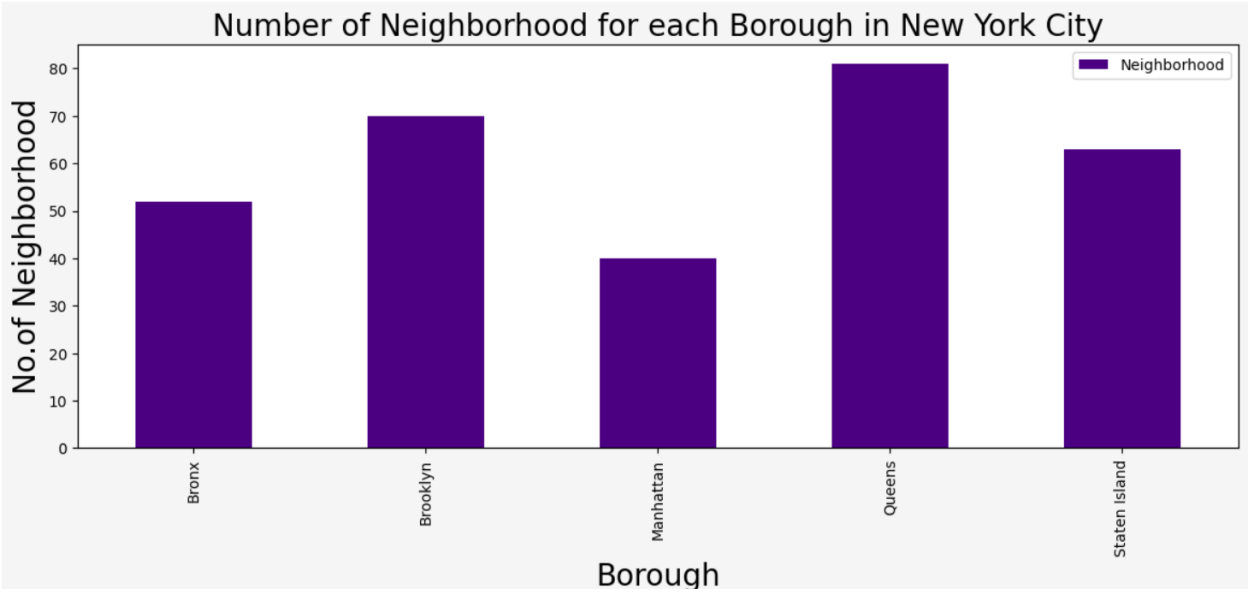
There are 306 neighbourhood in NYC.

```
[ ] NYC_data.head()
```

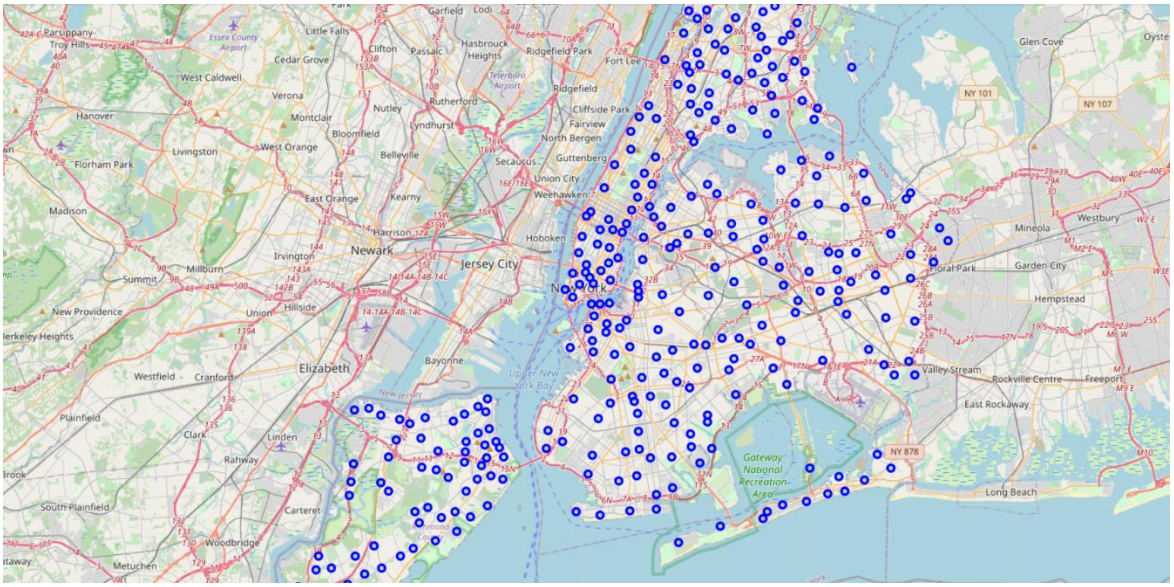
	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

```
[ ] NYC_data.shape
(306, 4)
```

The above code shows that are a total of 306 different Neighborhoods in New York.



2.Plot showing different neighbourhood in NYC



3.Number of venues in neighbourhood of NYC

```
[ ] NYC_venues.head()
```

	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
2	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
4	Wakefield	40.894705	-73.847201	Rite Aid	40.889062	-73.842993	Pharmacy

```
[ ] NYC_venues.shape
```

```
(15598, 7)
```

We found 15598 venues in the neighbourhood of NYC

There are 15,598 venues in NYC

4.Number of categories of venues in NYC neighbourhood

458 different venues in NYC neighbourhood

```
[ ] print('There are {} uniques categories.'.format(len(NYC_venues['Venue Category'].unique())))

There are 458 uniques categories.

[ ] NYC_venues['Venue Category'].unique()

array(['Dessert Shop', 'Ice Cream Shop', 'Caribbean Restaurant',
       'Pharmacy', 'Donut Shop', 'Gas Station', 'Sandwich Place',
       'Supermarket', 'Fried Chicken Joint', 'Deli / Bodega',
       'Pizza Place', 'Fast Food Restaurant', 'Bakery', 'Food Truck',
       'Gift Shop', 'Bagel Shop', 'Dumpling Restaurant',
       'Spanish Restaurant', 'Discount Store', 'Seafood Restaurant',
       'Mexican Restaurant', 'Post Office', 'Shopping Mall',
       'Mobile Phone Shop', 'Department Store', 'Grocery Store',
       'Convenience Store', 'Mattress Store',
       'Paper / Office Supplies Store', 'Restaurant', 'Bus Station',
       'Chinese Restaurant', 'Hotel', 'Basketball Court', "Men's Store",
       'Baseball Field', 'Park', 'Harbor / Marina', 'Accessories Store',
       'Electronics Store', 'Laundromat', 'Music Venue',
       'Other Great Outdoors', 'Print Shop', 'Diner',
       'Furniture / Home Store', 'Historic Site', 'Burger Joint',
       'Bowling Alley', 'Breakfast Spot', 'Cocktail Bar',
       'Asian Restaurant', 'Nightclub', "Women's Store",
       'Food & Drink Shop', 'Sporting Goods Shop', 'Plaza', 'Coffee Shop',
       'River', 'Art Gallery', 'Playground', 'Athletics & Sports',
       'Japanese Restaurant', 'Tapas Restaurant', 'Greek Restaurant',
       'Bar', 'Café', 'Bank', 'Health & Beauty Service',
       'Thai Restaurant', 'Italian Restaurant', 'Wine Shop',
       'Frozen Yogurt Shop', 'Medical Supply Store', 'Gourmet Shop',
       'Latin American Restaurant', 'Pub', 'Beer Bar', 'Warehouse Store',
       'Trail', 'Wings Joint', 'Yoga Studio', 'Thrift / Vintage Store',
       'Candy Store', 'Gym', 'Rental Car Location', 'Supplement Shop',
       'Miscellaneous Shop', 'American Restaurant', 'Pet Store',
       'Steakhouse', 'Shopping Plaza', 'Metro Station', 'Jewelry Store',
       'Outdoor Sculpture', 'Bus Stop', 'Scenic Lookout',
       'Tennis Stadium', 'Video Game Store', 'Big Box Store',
       'Kids Store', 'College Stadium', 'Tennis Court',
```

5.Results after categorisation

Only 5 neighbourhood are shown for further results refer notebook.

	Neighbourhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Animal Shelter	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Ent
0	Allerton	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
1	Annadale	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.071429	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
2	Arden Heights	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
3	Arlington	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.090909	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	
4	Arrochar	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	

6.Top 15 common venues of each neighbourhood

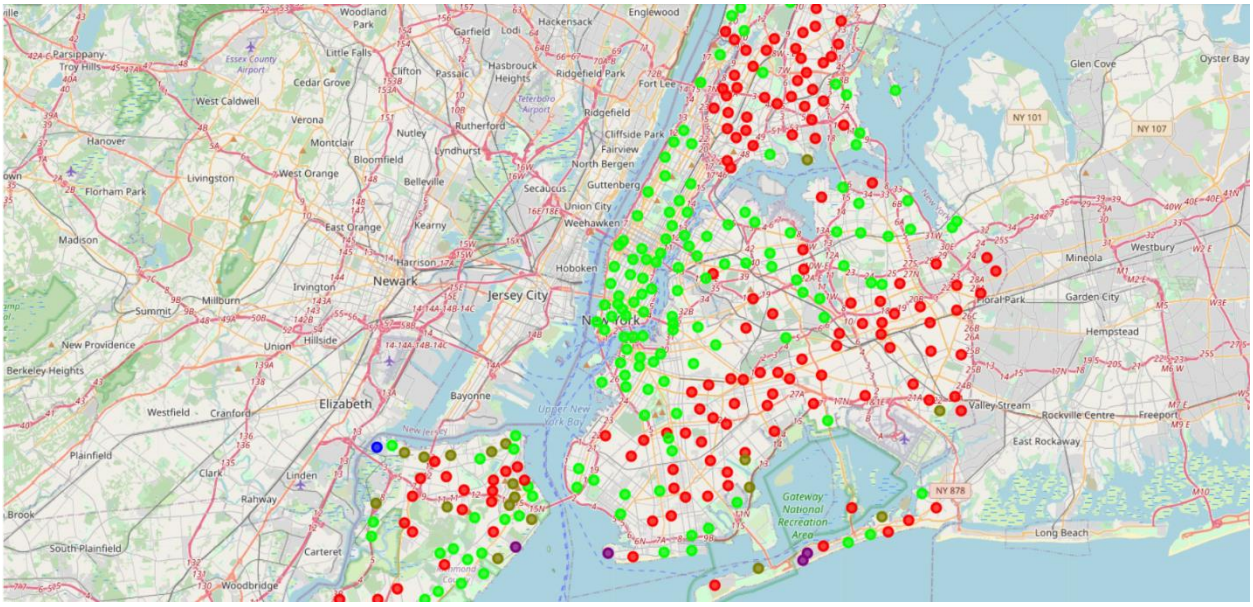
Only 5 neighbourhood are shown for further results refer notebook.

	Neighbourhood	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	11th Most Common Venue	12th Most Common Venue	13th Most Common Venue	14th Most Common Venue	15th Most Common Venue
0	Allerton	Pizza Place	Donut Shop	Sandwich Place	Supermarket	Bar	Pharmacy	Fast Food Restaurant	Discount Store	Fried Chicken Joint	Mobile Phone Shop	Bike Trail	Clothing Store	Mexican Restaurant	Bus Station	Caribbean Restaurant
1	Annadale	Restaurant	Pizza Place	Sushi Restaurant	Diner	Liquor Store	Home Service	Park	Food	Train Station	American Restaurant	Cosmetics Shop	Deli / Bodega	Farmers Market	Financial or Legal Service	Fast Food Restaurant
2	Arden Heights	Spa	Food	Coffee Shop	Pizza Place	Liquor Store	Mexican Restaurant	Sandwich Place	Pharmacy	Bus Stop	Optical Shop	Flower Shop	Event Space	Exhibit	Factory	Falafel Restaurant
3	Arlington	Bus Stop	Deli / Bodega	Juice Bar	Coffee Shop	Construction & Landscaping	American Restaurant	Snack Place	Farmers Market	Farm	Fish Market	Fast Food Restaurant	Falafel Restaurant	Field	Filipino Restaurant	Film Studio
4	Arrochar	Baseball Field	Italian Restaurant	Liquor Store	Beach	Deli / Bodega	Bus Stop	Pizza Place	Sandwich Place	Sculpture Garden	Park	Outdoors & Recreation	Food Truck	Supermarket	Playground	Middle Eastern Restaurant

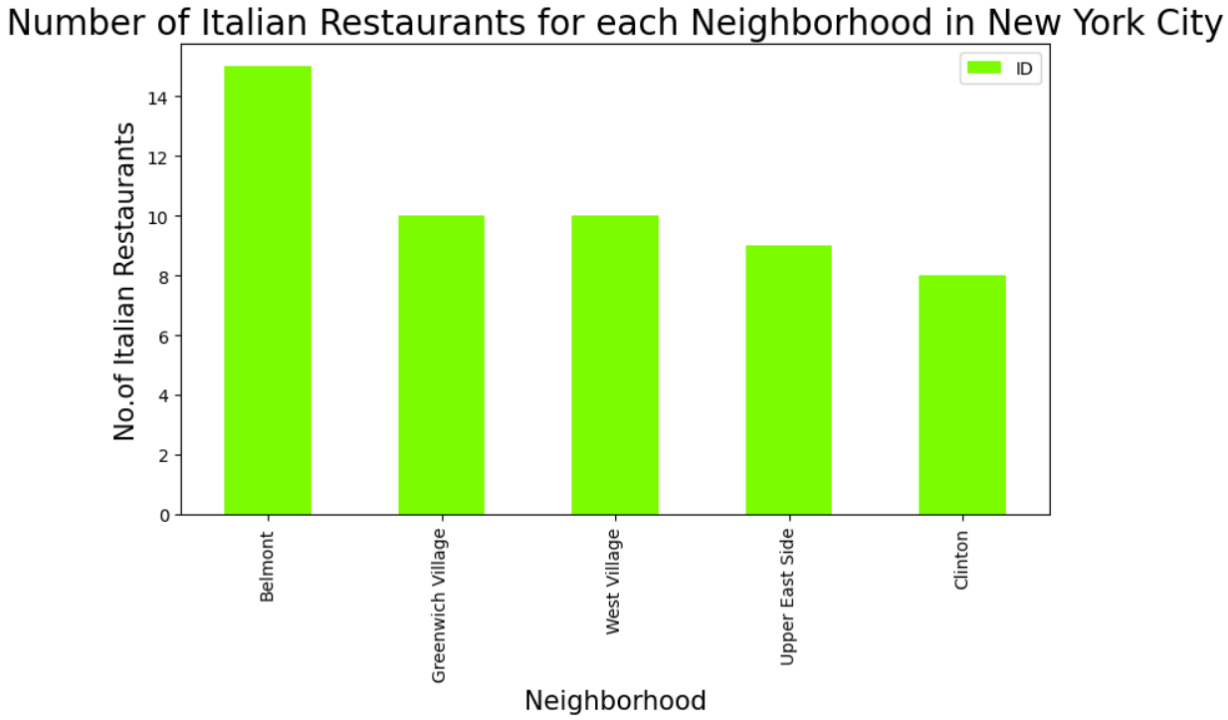
The above table shows the most common venues in any Neighnourhood of NYC.

7.Map with clustered neighbourhood

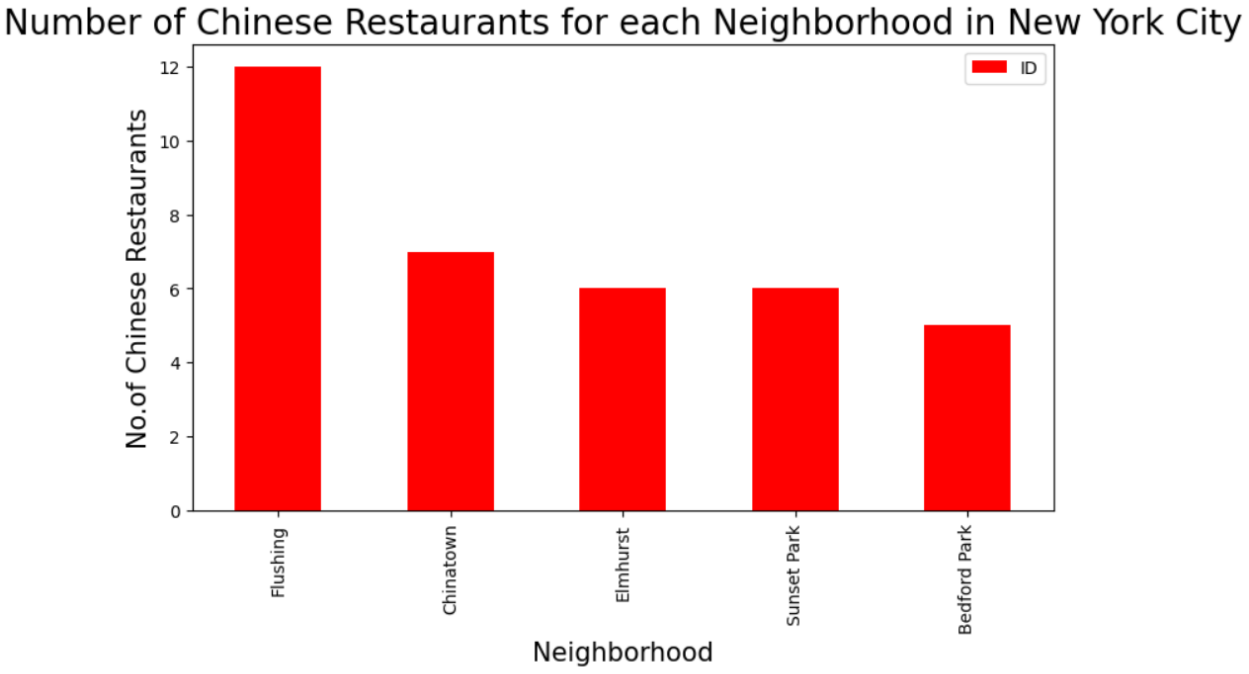
	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue	12th Most Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	3	Supermarket	Pharmacy	Donut Shop	Gas Station	Ice Cream Shop	Fast Food Restaurant	Bakery	Bagel Shop	Caribbean Restaurant	Pizza Place	Gift Shop	Sandwich Place
1	Bronx	Co-op City	40.874294	-73.829939	3	Mattress Store	Pizza Place	Bus Station	Fast Food Restaurant	Accessories Store	Paper / Office Supplies Store	Grocery Store	Men's Store	Mexican Restaurant	Mobile Phone Shop	Fried Chicken Joint	Music Venue
2	Bronx	Eastchester	40.887556	-73.827806	3	Caribbean Restaurant	Fast Food Restaurant	Diner	Pizza Place	Historic Site	Seafood Restaurant	Bakery	Shopping Mall	Grocery Store	Nightclub	Asian Restaurant	Bowling Alley
3	Bronx	Fieldston	40.895437	-73.905643	0	Bus Station	Park	Plaza	Athletics & Sports	River	Art Gallery	Coffee Shop	Playground	Pizza Place	Field	Filipino Restaurant	Zoo Exhibit
4	Bronx	Riverdale	40.890834	-73.912585	0	Park	Diner	Japanese Restaurant	Bank	Bar	Plaza	Pizza Place	Mexican Restaurant	Bagel Shop	Burger Joint	Baseball Field	Medical Supply Store



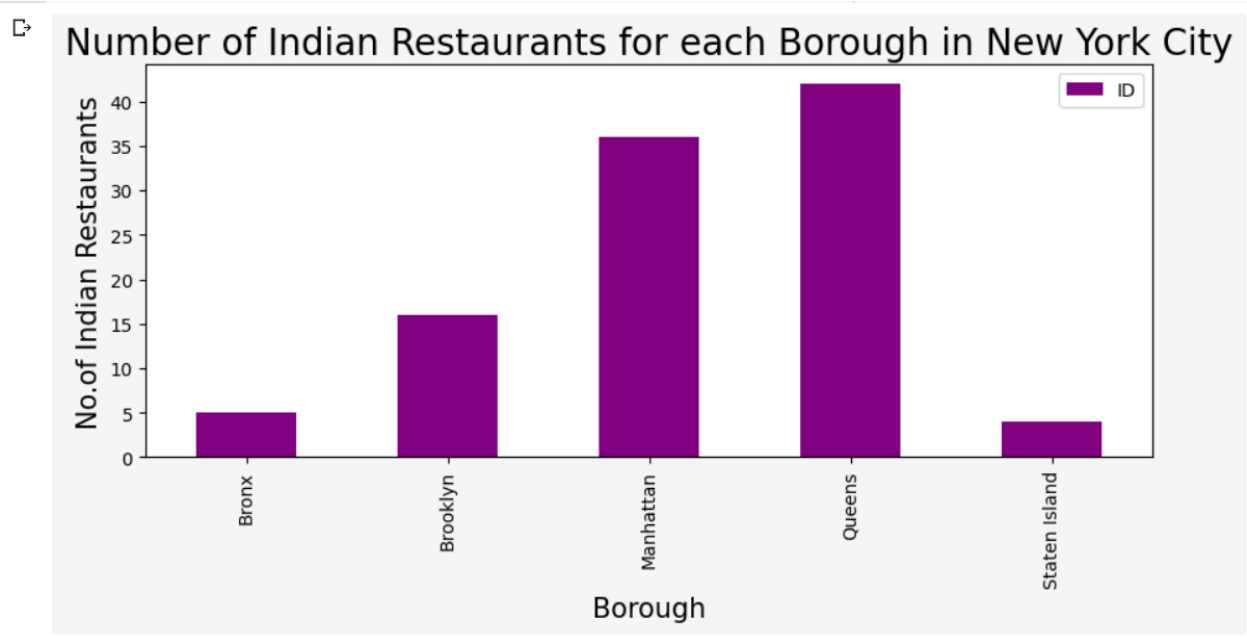
8. Bar Plot presenting showing number of Italian restaurants in Borough and Neighborhood of NYC.

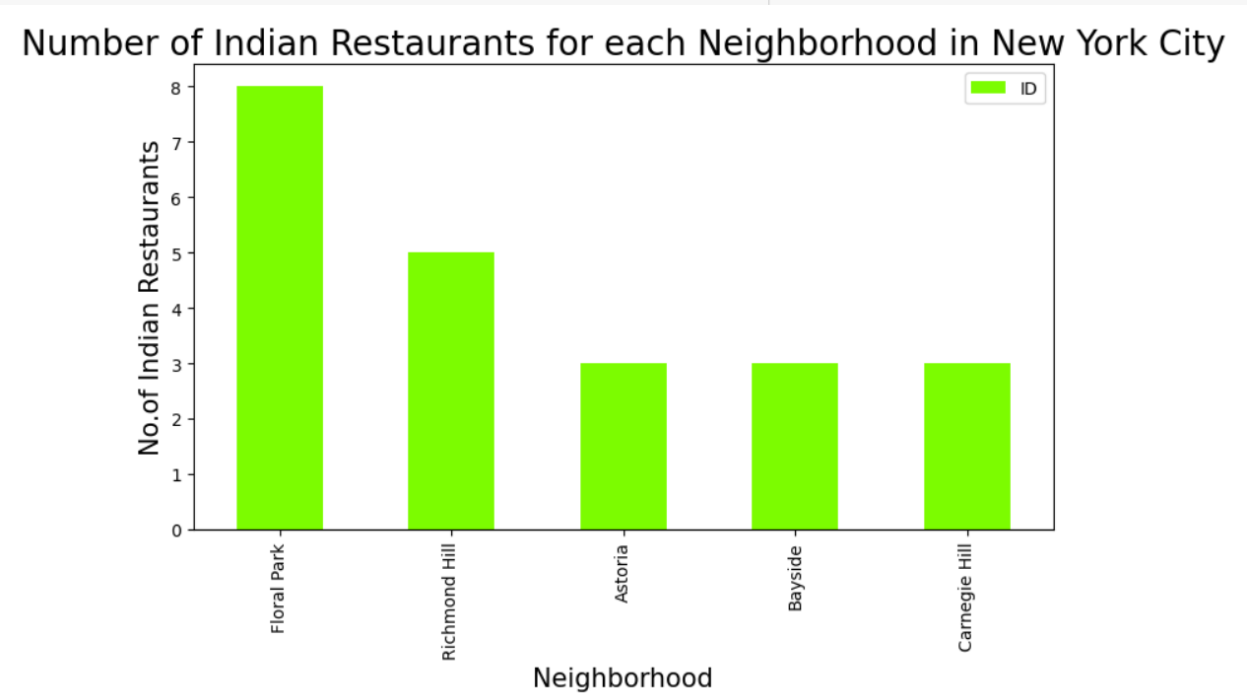


9. Bar Plot presenting showing number of Chinese restaurants in Borough and Neighborhood of NYC.

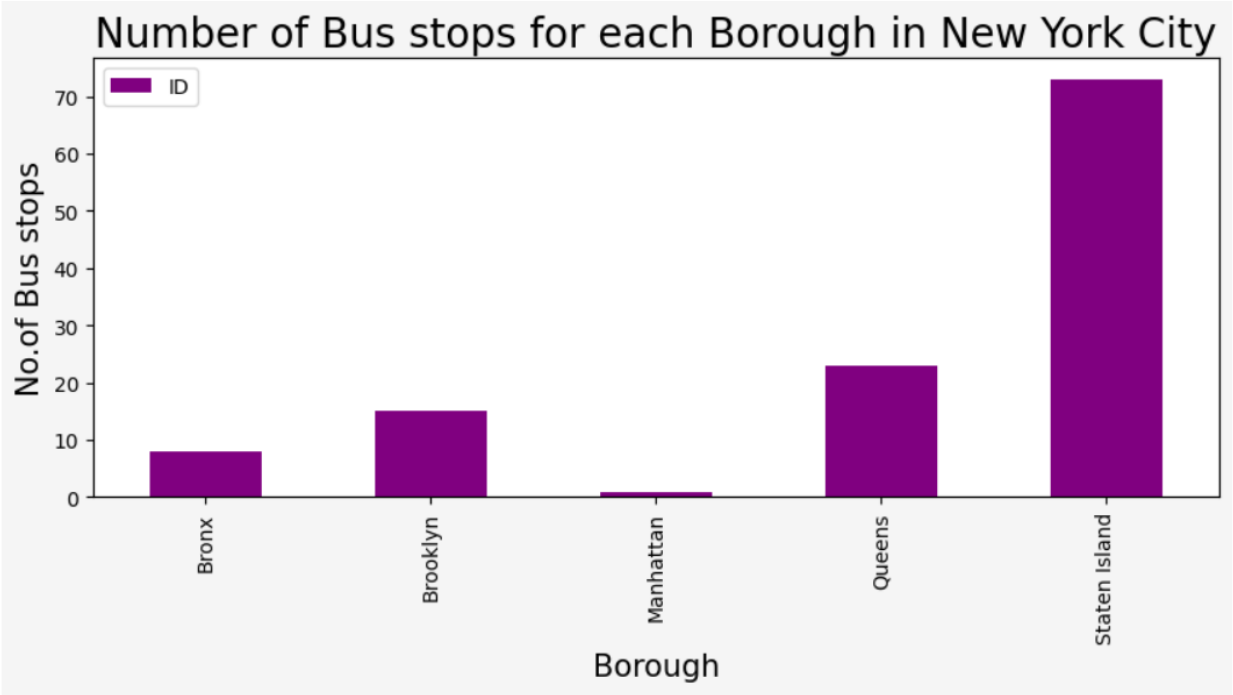


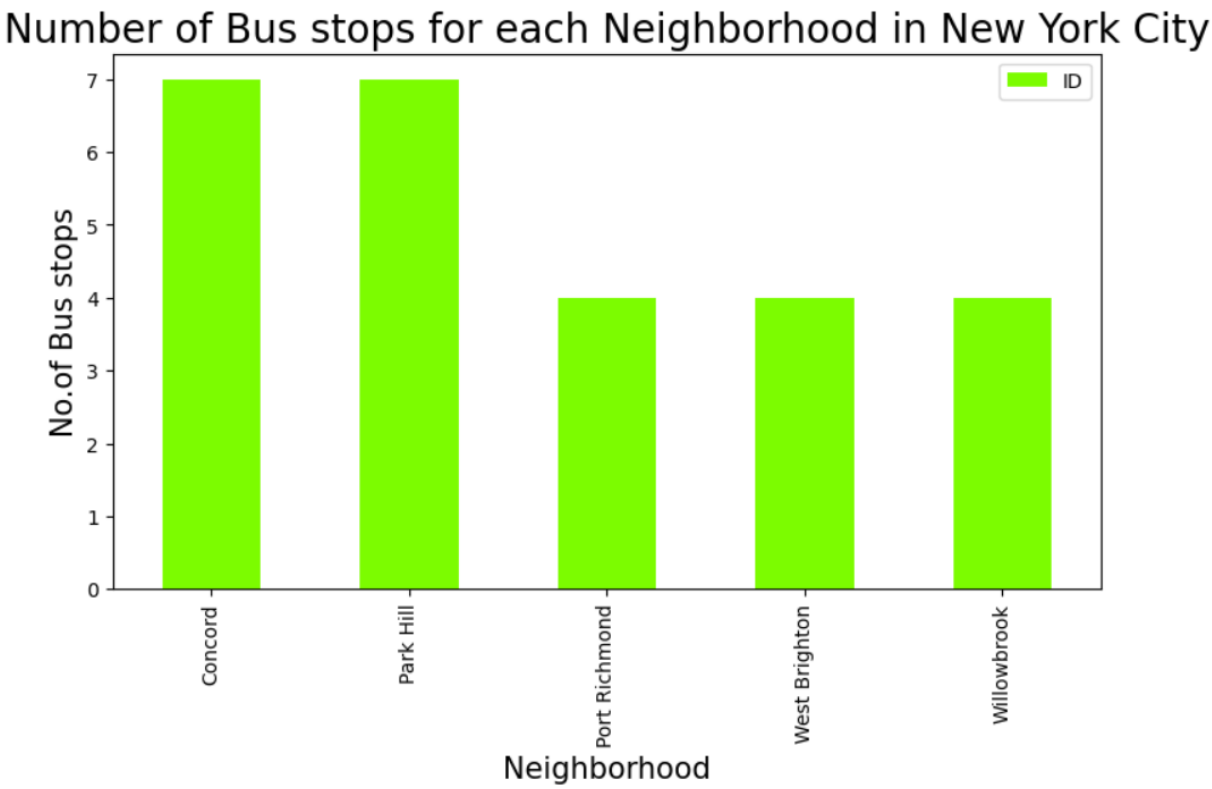
10. Bar Plot presenting showing number of Indian restaurants in Borough and Neighborhood of NYC.



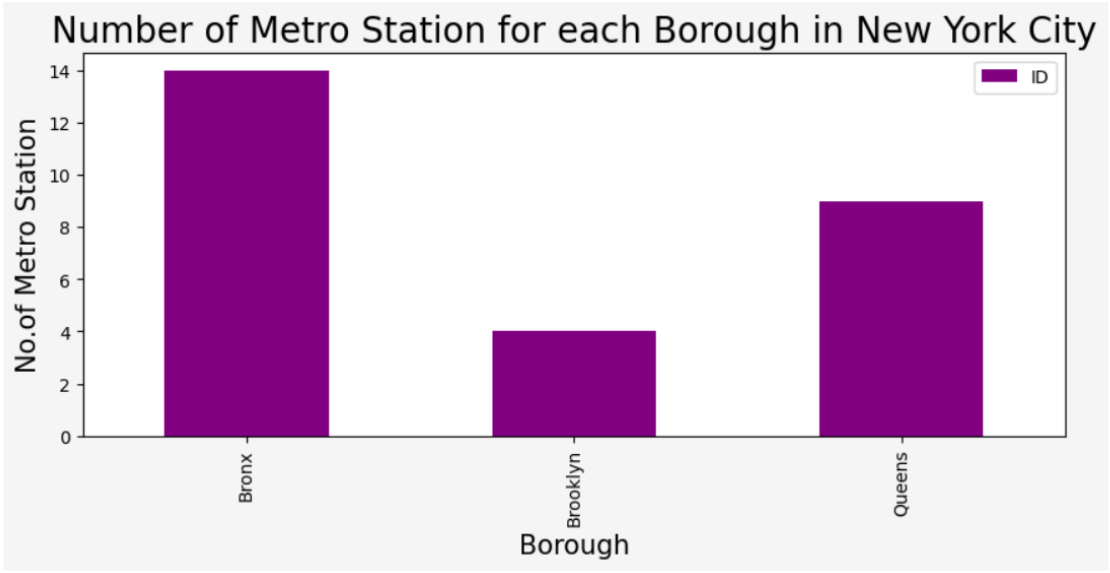


11. Bar Plot presenting showing number of Bus Stops in Borough and Neighborhood of NYC.

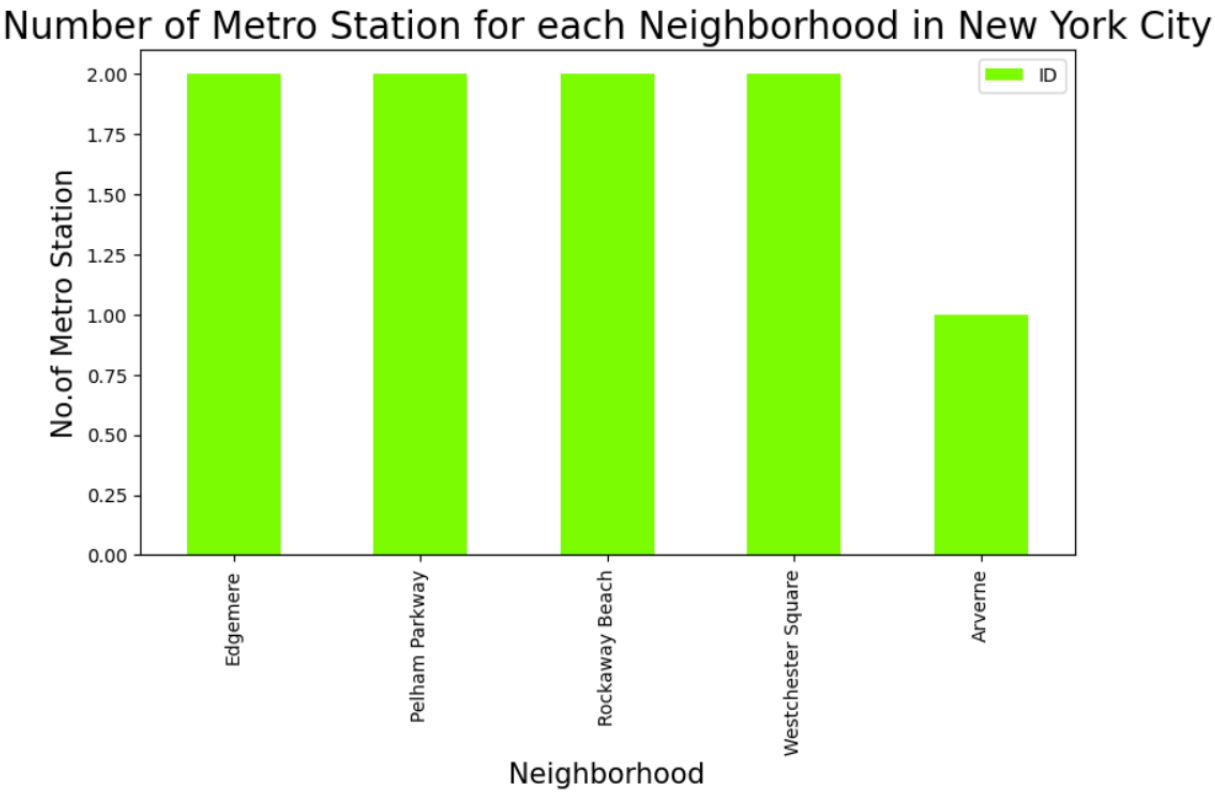




12.Bar Plot presenting showing number of Metro Stations in Borough and Neighborhood of NYC.

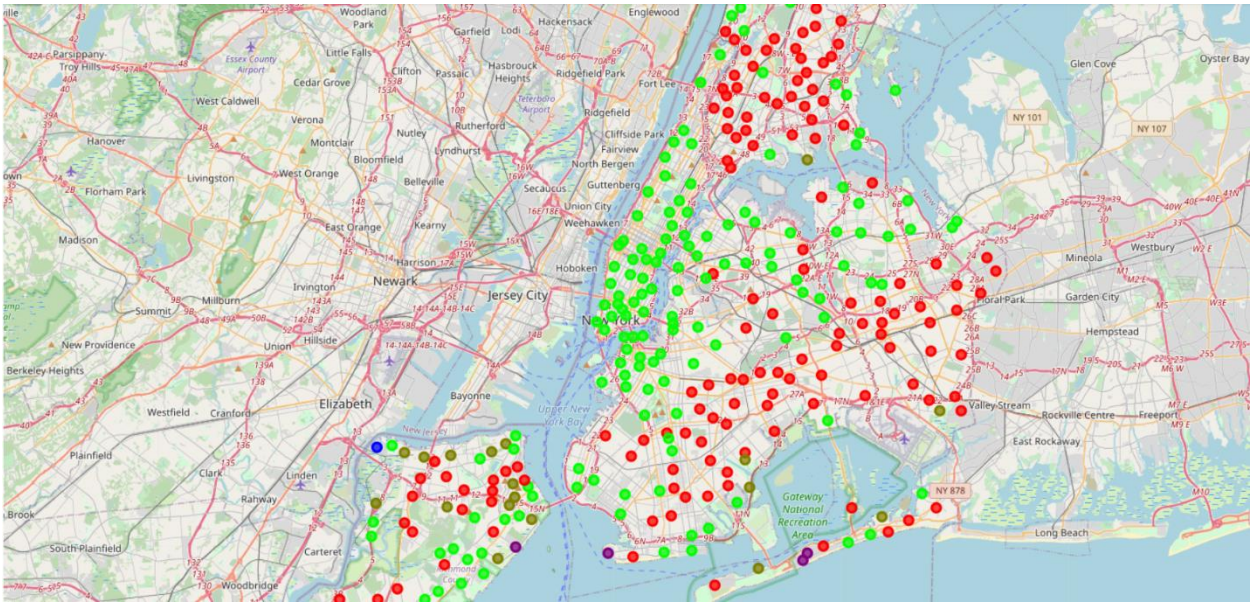


As the results shows Bronx has the highest number of Metro Stations.



Observations and Recommendations:

Which neighbourhoods can be considered similar in NYC, show them in NYC map?



Various types of venues in NYC?

```
[ ] print('There are {} uniques categories.'.format(len(NYC_venues['Venue Category'].unique())))

    There are 458 uniques categories.

[ ] NYC_venues['Venue Category'].unique()

    array(['Dessert Shop', 'Ice Cream Shop', 'Caribbean Restaurant',
           'Pharmacy', 'Donut Shop', 'Gas Station', 'Sandwich Place',
           'Supermarket', 'Fried Chicken Joint', 'Deli / Bodega',
           'Pizza Place', 'Fast Food Restaurant', 'Bakery', 'Food Truck',
           'Gift Shop', 'Bagel Shop', 'Dumpling Restaurant',
           'Spanish Restaurant', 'Discount Store', 'Seafood Restaurant',
           'Mexican Restaurant', 'Post Office', 'Shopping Mall',
           'Mobile Phone Shop', 'Department Store', 'Grocery Store',
           'Convenience Store', 'Mattress Store',
           'Paper / Office Supplies Store', 'Restaurant', 'Bus Station',
           'Chinese Restaurant', 'Hotel', 'Basketball Court', "Men's Store",
           'Baseball Field', 'Park', 'Harbor / Marina', 'Accessories Store',
           'Electronics Store', 'Laundromat', 'Music Venue',
           'Other Great Outdoors', 'Print Shop', 'Diner',
           'Furniture / Home Store', 'Historic Site', 'Burger Joint',
           'Bowling Alley', 'Breakfast Spot', 'Cocktail Bar',
           'Asian Restaurant', 'Nightclub', "Women's Store",
           'Food & Drink Shop', 'Sporting Goods Shop', 'Plaza', 'Coffee Shop',
           'River', 'Art Gallery', 'Playground', 'Athletics & Sports',
           'Japanese Restaurant', 'Tapas Restaurant', 'Greek Restaurant',
           'Bar', 'Café', 'Bank', 'Health & Beauty Service',
           'Thai Restaurant', 'Italian Restaurant', 'Wine Shop',
           'Frozen Yogurt Shop', 'Medical Supply Store', 'Gourmet Shop',
           'Latin American Restaurant', 'Pub', 'Beer Bar', 'Warehouse Store',
           'Trail', 'Wings Joint', 'Yoga Studio', 'Thrift / Vintage Store',
           'Candy Store', 'Gym', 'Rental Car Location', 'Supplement Shop',
           'Miscellaneous Shop', 'American Restaurant', 'Pet Store',
           'Steakhouse', 'Shopping Plaza', 'Metro Station', 'Jewelry Store',
           'Outdoor Sculpture', 'Bus Stop', 'Scenic Lookout',
           'Tennis Stadium', 'Video Game Store', 'Big Box Store',
           'Kids Store', 'College Stadium', 'Tennis Court',
```

For answer next phase of questions please read below paragraph-

According to the economics of the market, the larger the competition between the seller in the market, the greater the selection the consumer gets. The greater selection typically causes lower prices for the products, compared to what the price would be if there was no competition (monopoly) or little competition (oligopoly). Competition bolsters product differentiation as businesses try to innovate and entice consumers to gain a higher market share. It helps in improving the processes and productivity as businesses strive to perform better than competitors with limited resources. Considering this fact, the larger the number of restaurants, will result in better competition between restaurants. This will result in better quality and taste of food at cheaper rates.

Hence, the person preferring a specific cuisine should visit the neighbourhood consisting of the largest number of restaurants of that specific cuisine. This result will also be valid during picking the restaurant for dining. We can directly figure out the neighbourhood to select for dining purpose.

Number of Italian restaurants in NYC?

Ans: -429

what is best location in New York City for Italian Cuisine?

Ans: - Belmont in Bronx has the highest number of Italian Restaurants with total of 15. Hence this area can be considered the ideal location for Italian Cuisine.

Number of Chinese restaurants in NYC?

Ans: -345

what is best location in New York City for Chinese Cuisine?

Ans: - Flushing in Queens has the highest number of Chinese Restaurants with total of 12. Hence this area can be considered the ideal location for Chinese Cuisine.

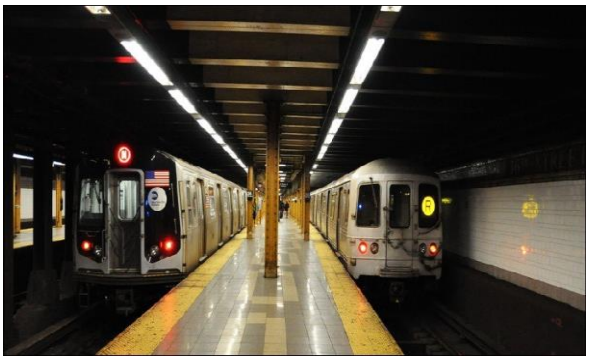
Number of Indian restaurants in NYC?

Ans: -103

what is best location in New York City for Indian Cuisine?

Ans: - Floral Park in Queens has the highest number of Indian Restaurants with total of 8. Hence this area can be considered the ideal location for Indian Cuisine.

For answer next phase of questions please read below paragraph-



Similarly, for the case of commute larger the stops/stations easier will it be to access the transportation medium.

what is best location in New York City considering Buses as the mode of transport?

Ans: - Concord and Park Hill neighbourhood of Staten Island has highest number of Bus stop each having 7 bus stops. Hence these areas can be considered the ideal location if our search is based on bus stops.

what is best location in New York City considering Metros as the mode of transport?

Ans: - Edgemere, Pelham Parkway, Rockaway Beach and Westchester Square, all have 2 metro station each. In case of metros recommending a location based on largest number of stations will not be advisable. Rather the neighbourhood which does not consists any metro stations will not be advisable. Hence if a neighbourhood has even 1 station it's good to go. As we have selected a radius of 750 meters.

Various other parameters like rental cost are ignored. For better results, these parameters should be considered. Some parameters like ratings, tips and likes of restaurants can be considered for better results. But due to unavailability of premium Foursquare account, the above features for such large neighborhood cannot be accessed.

There is always room for improvement and hence the above solution I have provided can also be improved for best results depending upon the data we have. I have only analyzed some aspects of NYC various other parameters can be analyzed both with breadth and depth.

CONCLUSION:

An attempt has been made to analyze, discover and describe the various aspects of the neighborhood of New York City. All the questions presented in the section ‘Problem Description’ are answered, results are provided with proper visualization and specific observations are drawn out.