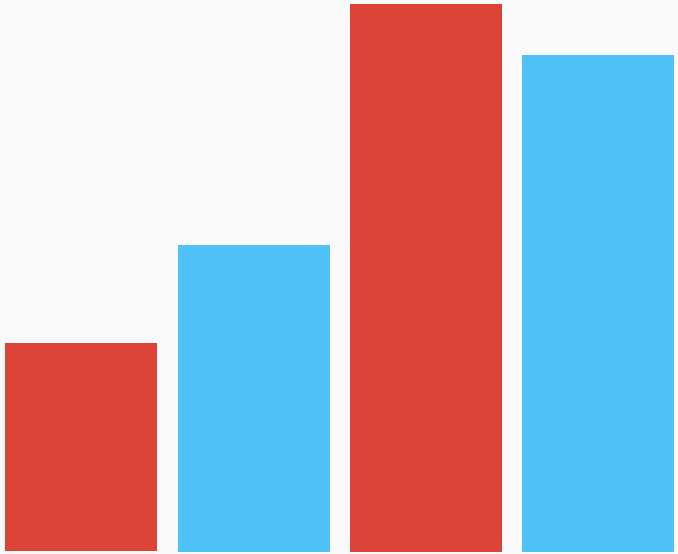
Capstone Project

The Battle of Neighborhoods

Introduction

Introduction

As we known, New York city is main and important financial centre in the world and also described as the cultural, financial, and media capital of the world, significantly influencing commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. To survive in the competitive market in the New York city, ensure a correct location for starting own business is the key of success. I will use my exploratory data analysis to help on suggesting which location is suitable for restaurant opening.



Target Audience

To help the stakeholder who is interested in opening his/ her restaurant in the New York city and suggest them on correct location for his/ her business.

Data

- *Free and existing New York City dataset via* https://geo.nyu.edu/catalog/nyu_2451_34572
- *Foursquare API*

How it works

Step 1

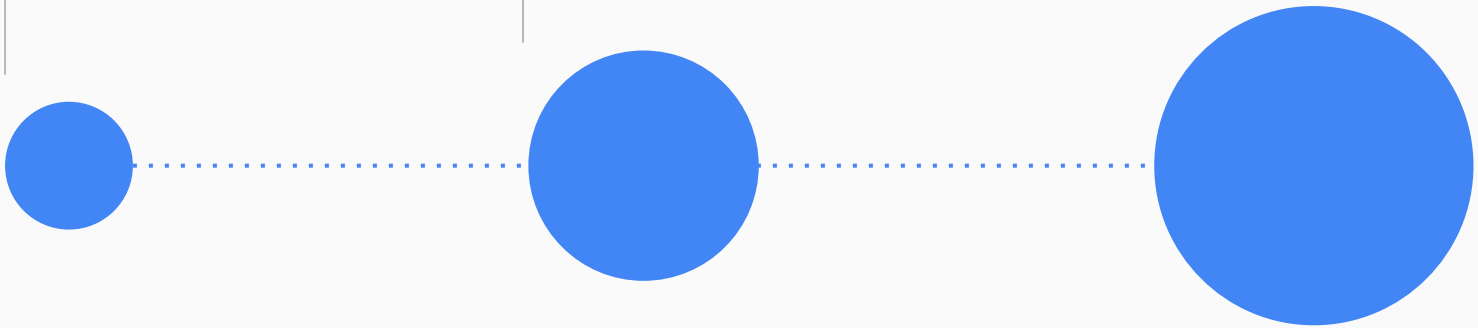
Data Cleansing and Simplifying

Step 2

Segmenting and Clustering the locations in NYC

Step 3

Results of most common visit venues for food and drink



Methodology:

K-Means Clustering Algorithm

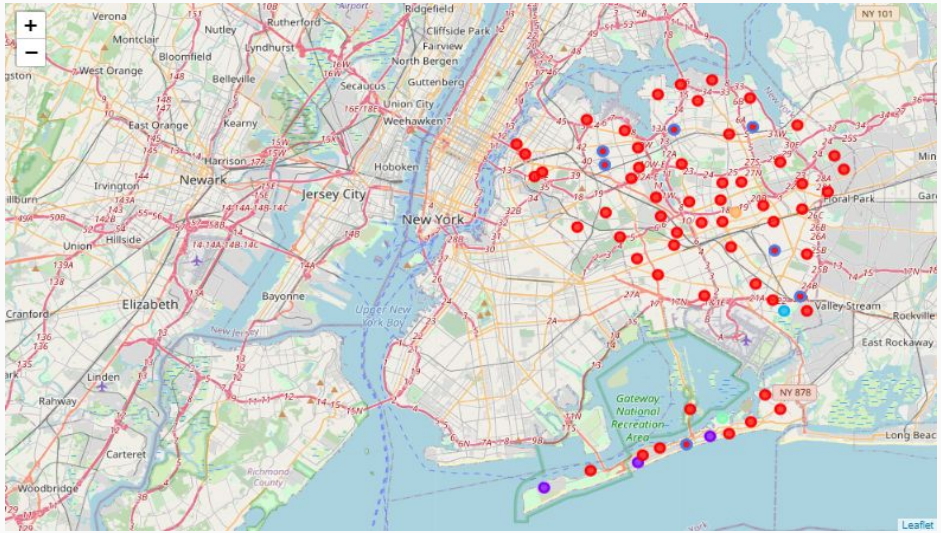
Results

- Queens has the most neighborhoods among other borough.
- Inside the Queens, there are total 81 Neighborhoods with 273 Venues included 76 of Restaurants, Cafe and Bars.
- The most visited venues are Cluster 1 and also covered with light blue color.

```
top_10_neigh = []
for neigh in queens_grouped_sorted['Neighborhood'].head(10):
    top_10_neigh.append(neigh)

top_10_neigh
```

```
['Elmhurst',
 'Murray Hill',
 'Ravenswood',
 'Astoria',
 'Jackson Heights',
 'Laurelton',
 'Bayside',
 'Rockaway Beach',
 'Flushing',
 'St. Albans']
```



Results

Top 10 Neighborhoods has been defined by using frequency of visit of each venues for food and drinks:

- Elmhurst freq=0.787879,
- Murray Hill freq=0.595745,
- Ravenswood freq=0.576923,
- Astoria freq=0.550000,
- Jackson Heights freq=0.525000,
- Laurelton freq=0.500000,
- Bayside freq=0.473684,
- Rockaway Beach freq=0.465116,
- Flushing freq=0.450000,
- St. Albans freq=0.444444

Get Top 10 highest visited neighborhoods for food and drink by sorting dataframe in descending order

```
queens_grouped_sorted = queens_grouped.sort_values(ascending=False,by=['Total Visited Frequecy']).reset_index(drop=True)
print(queens_grouped_sorted.shape)
queens_grouped_sorted.head(10)
```

```
(81, 274)
```

	Neighborhood	Yoga Studio	Accessories Store	Afghan Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Autc
0	Elmhurst	0.000000	0.0	0.000000	0.000000	0.000000	0.030303	0.0	0.0	0.0	0.0	0.000000	0.0	0.
1	Murray Hill	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.021277	0.0	0.
2	Ravenswood	0.000000	0.0	0.038462	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.
3	Astoria	0.000000	0.0	0.000000	0.010000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.
4	Jackson Heights	0.000000	0.0	0.000000	0.000000	0.000000	0.012500	0.0	0.0	0.0	0.0	0.012500	0.0	0.
5	Laurelton	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.
6	Bayside	0.013158	0.0	0.000000	0.039474	0.000000	0.000000	0.0	0.0	0.0	0.0	0.013158	0.0	0.
7	Rockaway Beach	0.000000	0.0	0.000000	0.000000	0.046512	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.
8	Flushing	0.016667	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.016667	0.0	0.
9	St. Albans	0.000000	0.0	0.000000	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.


Thank You!