

# **The Battle of Neighborhood**

Abdul Mohsen Al-Maskeen

January 31, 2021

# Introduction

- The City of New York, is the most populous city in the United States. With an estimated 2018 population of 8,398,748 distributed over a land area of about 302.6 square miles (784 km<sup>2</sup>), New York is also the most densely populated major city in the United States. It started accepting immigrants in 19th century and have since become a melting pot of the diverse languages, people and cultures. In 2019, it was estimated to have a population of 8.3 million which live in 5 main boroughs namely Brooklyn, Bronx, Manhattan, Queens and Staten Island.
- New York is a very busy city, both in terms of its population and tourists. As per the data of 2019, the populations of Asians American is more than a million, which is about 12 % of the population of New York city. Add to it the people from subcontinent i.e. India, Bangladesh and Pakistan, Silence and Nepal and you have a very large Asian and south east Asian population that has come to New York for studying, jobs and businesses.
- A global power city, New York City has been described as the cultural, financial and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports.
- This information shall be useful for anyone who is doing business in New York especially in food restaurants. It shall also be useful for professionals who are looking to change jobs within New York.

# Business Problem Description

Every year hundreds of thousands of immigrants, businessmen, students and professionals visit, migrate to or settle in New York for work, education, livelihood and tourism. Due to the large area, several neighborhoods, income differences, and variations in quality of life from one neighborhood to another.

A business man has recently migrated to New York and is looking to open a business. His idea to open Thai food restaurant because a few Thai restaurants are found in the city. He believes it is good opportunity for a Thai food restaurant as it equally popular among the European and north American tourists who crave exotic food and will pay hefty prices for an upbeat expensive restaurant that provides good ambiance and Asian population who loves aromatic and spicy food at affordable prices.

People in New York is also very diverse in terms of the standard of living and income, so the businessman is thus planning to open two Thai restaurants, an expensive restaurant providing a fine dining experience to wealthy residents and tourists and cheap one for middle class customers.

The business man thus decided to hire a data scientist to suggest to him the location for two restaurants, the expensive restaurant in an locality which offers the opportunity for higher tips and better ratings driven by the social media usage of the rich and wealthy and a location for an cheap restaurant offering Thai food at reasonable prices in a neighborhood which has low number of Thai food restaurants.

# Target Audience

- Tourists of all classes, those looking for a romantic, business or elaborate dinner and those who are on the move and need a fast food on the go at affordable prices.

# Data Sources

- Geospatial data of the boroughs
- New York population is distributed into 5 boroughs and 306 neighbourhoods. To explore the data, we need to get the access to the data containing the boroughs and their geospatial coordinates. I downloaded the data freely available from the website
- [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- in geojson format. This data will be transformed into Pandas data frame for easy data analysis and visualization.

```
new_york_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

## Data Sources (Continued)

- Additionally from FourSquareAPI , we get Neighborhood names, ID, restaurant names, lat/long and restaurant categories.
- The data was retrieved from Foursquare API that required a user account and the secret key and Client ID which allows to extract required data from Foursquare API

	Neighborhood	Id	Name	Latitude	Longitude	Category
0	Marble Hill	4a739e29f964a520f5dc1fe3	Siam Square	40.878796	-73.916701	Thai Restaurant
1	Chinatown	5bbea2ad9411f2002c2c8562	Noree Thai Bazaar	40.717900	-73.992966	Thai Restaurant
2	Chinatown	5cc4e9d0c876c8002c3010cb	Wayla	40.718291	-73.992584	Thai Restaurant
3	Chinatown	598b97d559fe5c1d37565107	Jia	40.715454	-73.990036	Thai Restaurant
4	Chinatown	57e0890e498ed6d471c6fe92	Thailicious NYC	40.716310	-73.999944	Thai Restaurant

# Methodology – Data Retrieval

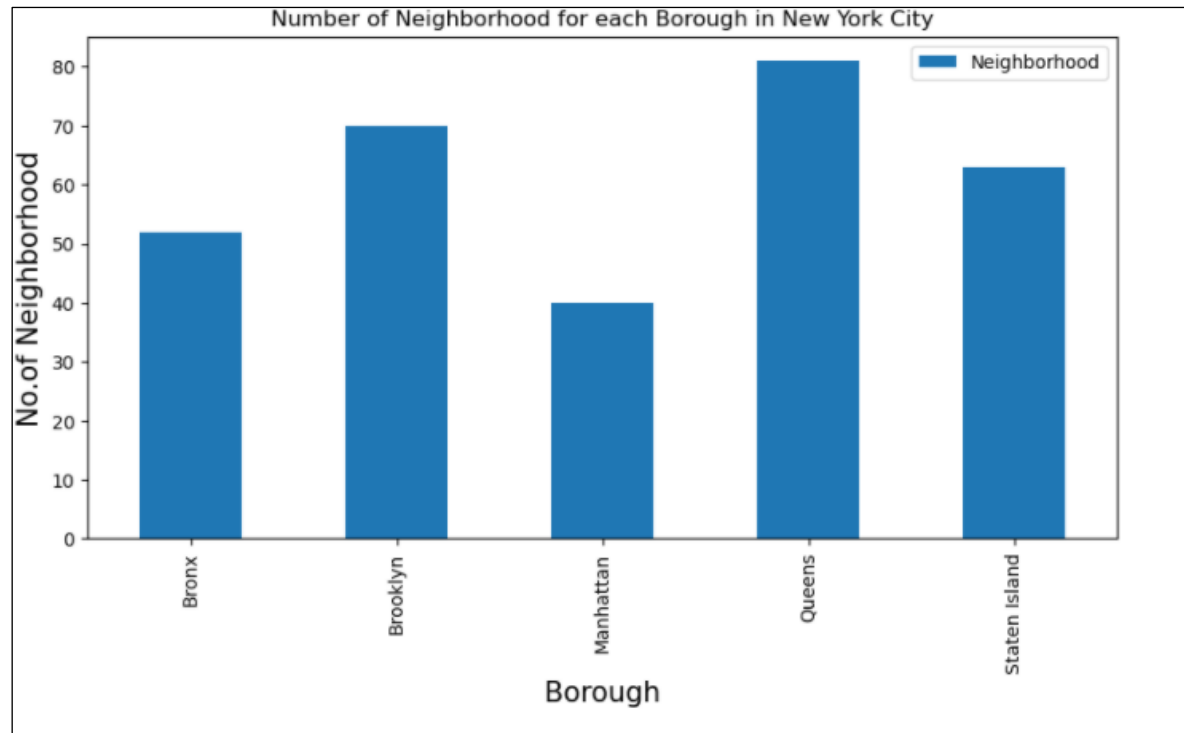
- Data was retrieved from
- [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)
- The data is automatically downloaded to a file called “nyu\_2451\_34572-geojson.json”. The file was renamed to newyork.Json for easy referral.

```
with open('newyork.json') as json_data:
    nydata = json.load(json_data)
nydata

{'type': 'FeatureCollection',
 'totalFeatures': 306,
 'features': [{'type': 'Feature',
  'id': 'nyu_2451_34572.1',
  'geometry': {'type': 'Point',
   'coordinates': [-73.84720052054902, 40.89470517661]},
  'geometry_name': 'geom',
  'properties': {'name': 'Wakefield',
   'stacked': 1,
   'annoline1': 'Wakefield',
   'annoline2': None,
   'annoline3': None,
   'annoangle': 0.0,
   'borough': 'Bronx',
   'bbox': [-73.84720052054902,
    40.89470517661,
    -73.84720052054902,
    40.89470517661]}},
```

# Data Visualization

- Plotting the bar graph of the borough, provides a visually easy to interpret the method of having a quick look of how neighborhood are distributed among boroughs.





# Data Transformation

- Data was transformed from .json format to pandas dataframe called “neighborhoods”. The transformation into pandas dataframe provides the opportunity to analyze, clean, and format the data into an analysis and visualization friendly format.

```
neighborhoods.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

# Methodology –Geocoder Locations

- Geopypackage helps to retrieve the location data from any selected area such as neighborhood or borrows.

## Get Geographical Coordinates of Newyork city

```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of New York City are 40.7127281, -74.0060152.

# Retrieve Geospatial Data

- Geopy package was used to retrieve the location data for neighborhood and boroughs.

## Get Geographical Coordinates of Newyork city

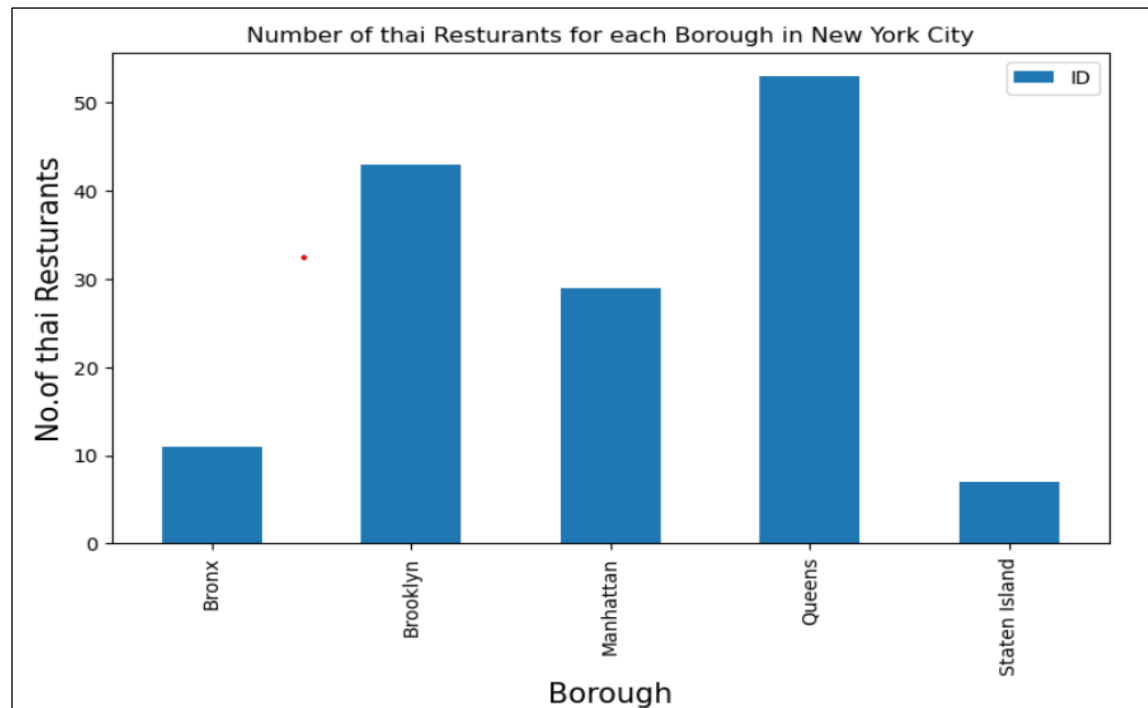
```
address = 'New York City, NY'

geolocator = Nominatim(user_agent="ny_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of New York City are {}, {}'.format(latitude, longitude))
```

The geograpical coordinate of New York City are 40.7127281, -74.0060152.

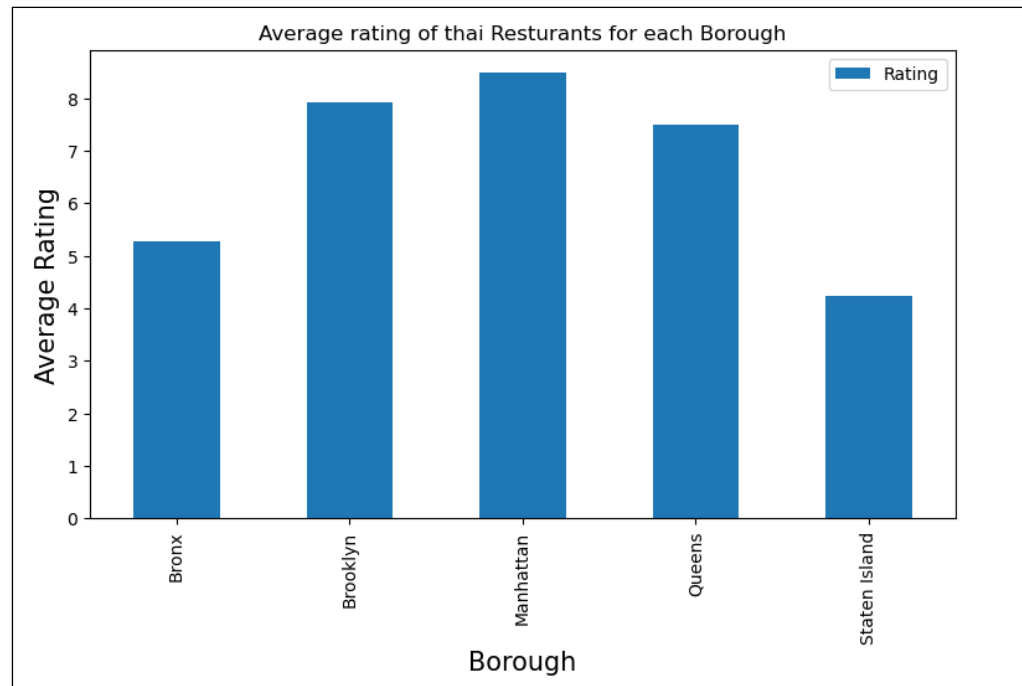
# Thai Restaurants for each Borough in New York City

```
plt.figure(figsize=(9,5), dpi = 100)
plt.title('Number of thai Resturants for each Borough in New York City')
plt.xlabel('Borough', fontsize = 15)
plt.ylabel('No.of thai Resturants', fontsize=15)
thai_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')
plt.legend()
plt.show()
```



# Average Rating of Thai restaurants for each Borough

```
plt.figure(figsize=(9,5), dpi = 100)
plt.title('Average rating of thai Resturants for each Borough')
plt.xlabel('Borough', fontsize = 15)
plt.ylabel('Average Rating', fontsize=15)
thai_rest_stats_ny.groupby('Borough').mean()['Rating'].plot(kind='bar')
plt.legend()
plt.show()
```

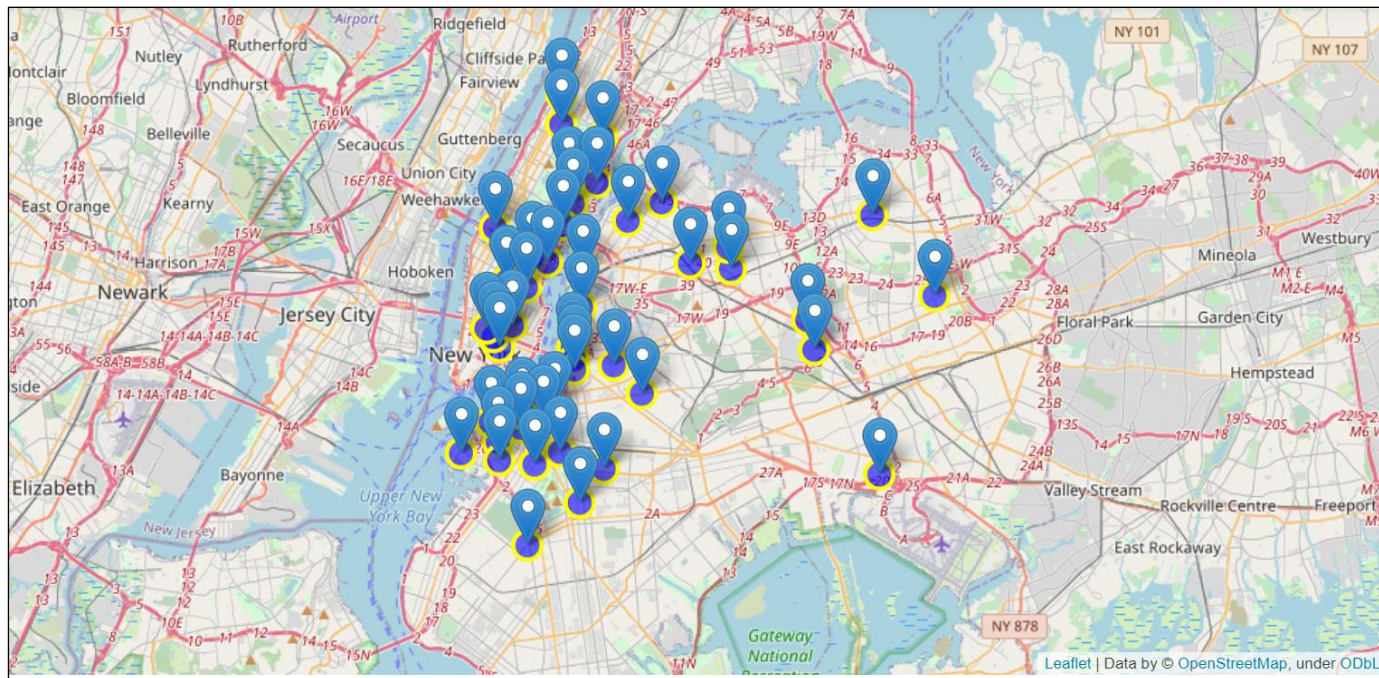


# Mapping and Data Visualization

```
ny_map = folium.Map(location=geo_location('New York'), zoom_start=12)
incidents = folium.map.FeatureGroup()

for lat, lng, in ny_neighborhood_stats[['Latitude','Longitude']].values:
    incidents.add_child(
        folium.CircleMarker(
            [lat, lng],
            radius=10,
            color='yellow',
            fill=True,
            fill_color='blue',
            fill_opacity=0.6
        )
    )

ny_neighborhood_stats['Label']=ny_neighborhood_stats['Neighborhood']+', '+ny_neighborhood_stats['Borough']+', '+ny_neighborhood_stats['City']
for lat, lng, label in ny_neighborhood_stats[['Latitude','Longitude','Label']].values:
    folium.Marker([lat, lng], popup=label).add_to(ny_map)
ny_map.add_child(incidents)
```



# Results

- Data science methodology proved effective in recommending areas for opening a Thai Restaurant with high probability of good return on investment.
- The exercise showed how data can be scraped from a website and used in python environment for data analysis, visualization and applying machine learning
- Data visualization provided excellent methods of graphically representing the data and using seaborn library and geospatial data was effectively visualized using the folium library.

# Discussion

- There is room for the improvement as other features such as restaurant ratings, areas with best tips and user likes from foursquare would provide more data and better clustering based on multi-attribute analysis and clustering.
- The exercise showcased the power of data science methodology and practice as recommender system and in visualization and data wrangling domains



# Conclusion

- The exercise provided good opportunity to help recommend a best place / places for opening a restaurant in Manhattan, New York. This methodology can be applied to variety of similar problems requiring clustering and recommendations using unsupervised machine learning.
- We were able to predict the best location to start Thai restaurant while ensuring the high rate of return and safe investment.

**END**

**Abdul Mohsen Al-Maskeen**  
**January 2021**