Capstone Project - The Battle of Neighborhoods (Week 2). Chinese restaurants in New York.

The New York metropolitan area is home to the largest and most prominent ethnic Chinese population outside of Asia. New York City itself contains by far the highest ethnic Chinese population of any individual city outside Asia, estimated at 628,763 as of 2017. Between 2000 and 2015, the population of foreign-born Chinese people in NYC grew by nearly 50 percent, the second-fastest-growing immigrant group, according to government data. At New York University alone, the population of Chinese international students more than doubled between 2012 and 2016, something that's already impacted the East Village dining scene. New York City is experiencing a Chinese food renaissance. Never before have the city's offerings been so diverse; not only are multiple regions represented, but price points range, too. A huge audience of discerning Chinese expats who seek flavors from home is fueling a highly competitive market of contemporary regional Chinese restaurants — one that means all of New York has access to a restaurant scene that's better than it's ever been. Despite the fact that the COVID-19 pandemic is far from over, chinese restaurants have reopened faster than any other category of independent restaurant. On March 30, 94% of Chinese restaurants were closed, compared with 61% of the city's other restaurants. By June 20, the number of closures had fallen to 10%. As follows from the above, New York City attracts many to start their business in the Chinese food industry. Before opening your business, you need to correctly determine the location, for which you need to analyze the situation with existing restaurants, their specifics and ratings in different areas of New York.

Data.

New York City data that contains list of Boroughs and Neighborhoods along with their latitude and longitude.
Data source : https://cocl.us/new_york_dataset
Description : This data set contains the required information. And we will use this data set to explore various neighborhoods of New York City.

Chinese restaurants in each neighborhood of new york city.
Data source : Fousquare API
Description : By using this api we will get all the venues in each neighborhood. We can filter these venues to get only chinese resturants.

GeoSpace data
Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm

Approach:
Collect the New York City data from https://cocl.us/new_york_dataset
Using FourSquare API we will find all venues for each neighborhood.
Filter out all venues that are Chinese Resturants.
Find rating , tips and like count for each Chinese Resturants using FourSquare API.
Using rating for each resturant , we will sort that data.
Visualize the Ranking of neighborhoods using folium library

Methodology

Required Libraries
pandas and numpy for handling data.
request module for using FourSquare API.
geopy to get co-ordinates of City of New York.
folium to visualize the results on a map

```
[1]: import pandas as pd
     import numpy as np

     pd.set_option('display.max_columns', None)
     pd.set_option('display.max_rows', None)

     import json # library to handle JSON files

     import requests
     from pandas.io.json import json_normalize # tranform JSON file into a pandas dataframe

     from urllib.request import urlopen

     !pip install bs4
     from bs4 import BeautifulSoup

     !conda config --add channels conda-forge

     !conda install -c conda-forge geopy --yes
     from geopy.geocoders import Nominatim

     !conda install -c conda-forge geocoder --yes
     import geocoder
     import os

     !conda install -c conda-forge folium --yes
     import folium # map rendering library

     import matplotlib.pyplot as plt
     import matplotlib.cm as cm
```

Defining functions
Now we need to define a function to get the geocodes i.e latitude and longitude of a given location using geopy.

```
[ ]: def geo_location(address):
         # get geo location of address
         geolocator = Nominatim(user_agent="ny_explorer")
         location = geolocator.geocode(address)
         latitude = location.latitude
         longitude = location.longitude
         return latitude,longitude
```

We define a function to intract with FourSquare API and get top 100 venues within a radius of 1000 metres for a given latitude and longitude. Below function will return us the venue id , venue name and category.

```
[ ]: def get_venues(lat,lng):

       #set variables
       radius=1000
       LIMIT=100
       CLIENT_ID =
       CLIENT_SECRET =
       VERSION = '20180605'

       #url to fetch data from foursquare api
       url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={
               CLIENT_ID,
               CLIENT_SECRET,
               VERSION,
               lat,
               lng,
```

Then we will define a function to get venue details like like count , rating , tip counts for a given venue id. This will be used for ranking.

```
[ ]: def get_venue_details(venue_id):

       CLIENT_ID =
       CLIENT_SECRET =
       VERSION = '20180605' # Foursquare API version

       #url to fetch data from foursquare api
       url = 'https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}'.format(
               venue_id,
               CLIENT_ID,
               CLIENT_SECRET,
               VERSION)

       # get all the data
       results = requests.get(url).json()
       venue_data=results['response']['venue']
       venue_details=[]
       try:
           venue_id=venue_data['id']
           venue_name=venue_data['name']
           venue_likes=venue_data['likes']['count']
           venue_rating=venue_data['rating']
           venue_tips=venue_data['tips']['count']
           venue_details.append([venue_id,venue_name,venue_likes,venue_rating,venue_tips])
       except KeyError:
           pass

       column_names=['ID','Name','Likes','Rating','Tips']
       df = pd.DataFrame(venue_details,columns=column_names)
       return df
```

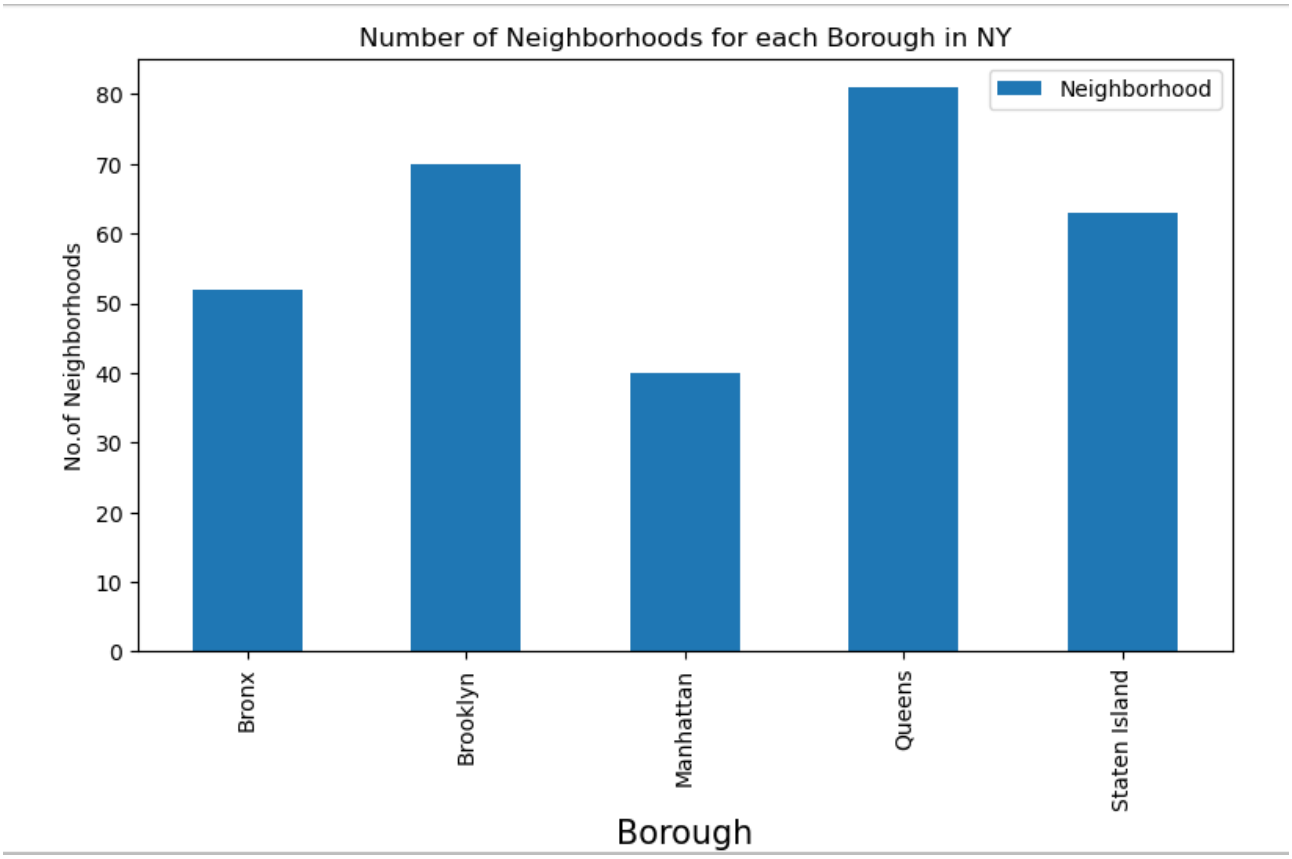
Now we define a funtion to get the New York city data such as Boroughs, Neighborhoods along with their latitude and longitude.

```
[ ]:
def get_new_york_data():
    url='https://cocl.us/new_york_dataset'
    resp=requests.get(url).json()
    # all data is present in features label
    features=resp['features']

    # define the dataframe columns
    column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']
    # instantiate the dataframe
    new_york_data = pd.DataFrame(columns=column_names)

    for data in features:
        borough = data['properties']['borough']
        neighborhood_name = data['properties']['name']

        neighborhood_latlon = data['geometry']['coordinates']
        neighborhood_lat = neighborhood_latlon[1]
        neighborhood_lon = neighborhood_latlon[0]

        new_york_data = new_york_data.append({'Borough': borough,
                                              'Neighborhood': neighborhood_name,
                                              'Latitude': neighborhood_lat,
                                              'Longitude': neighborhood_lon}, ignore_index=True)

    return new_york_data
```

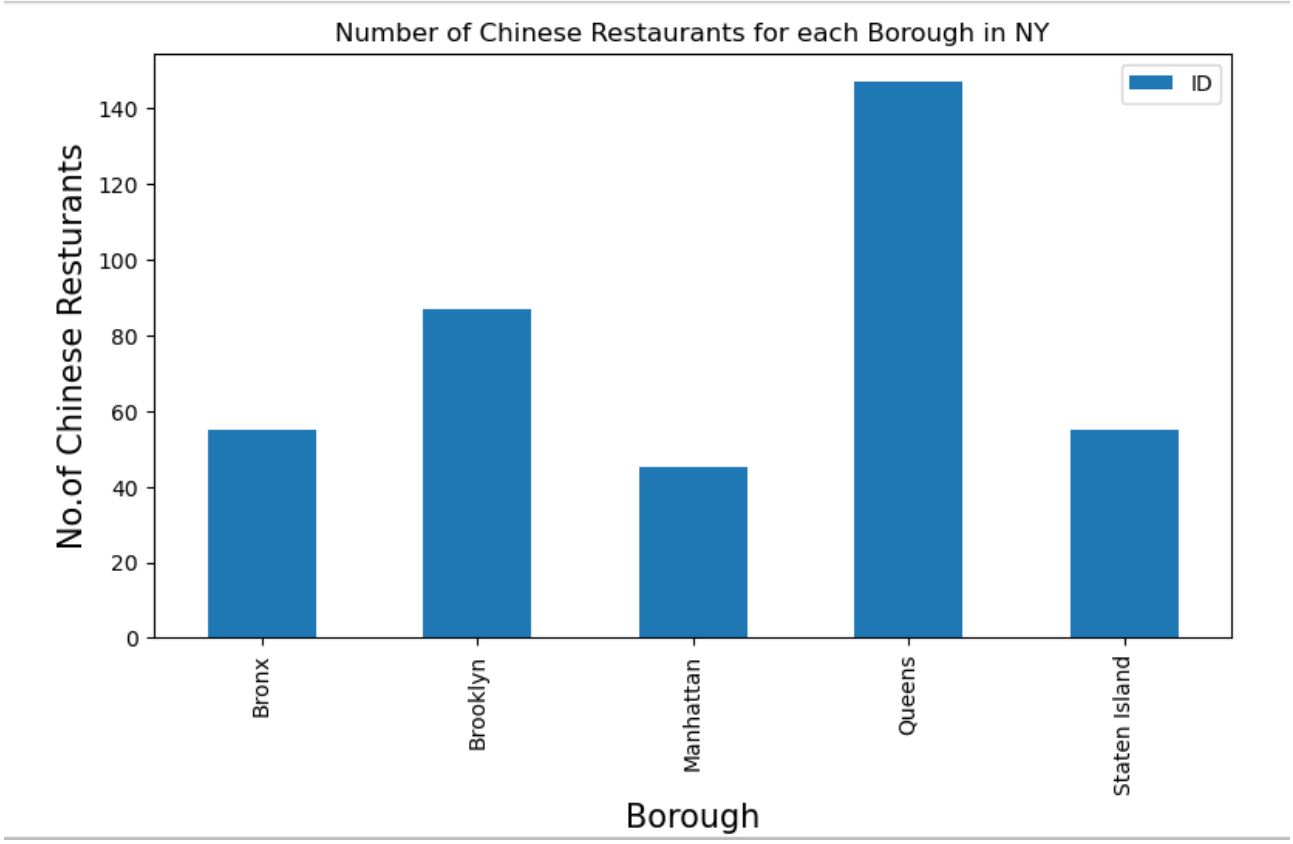
We will call the above funtion to get the New York city data.

Results

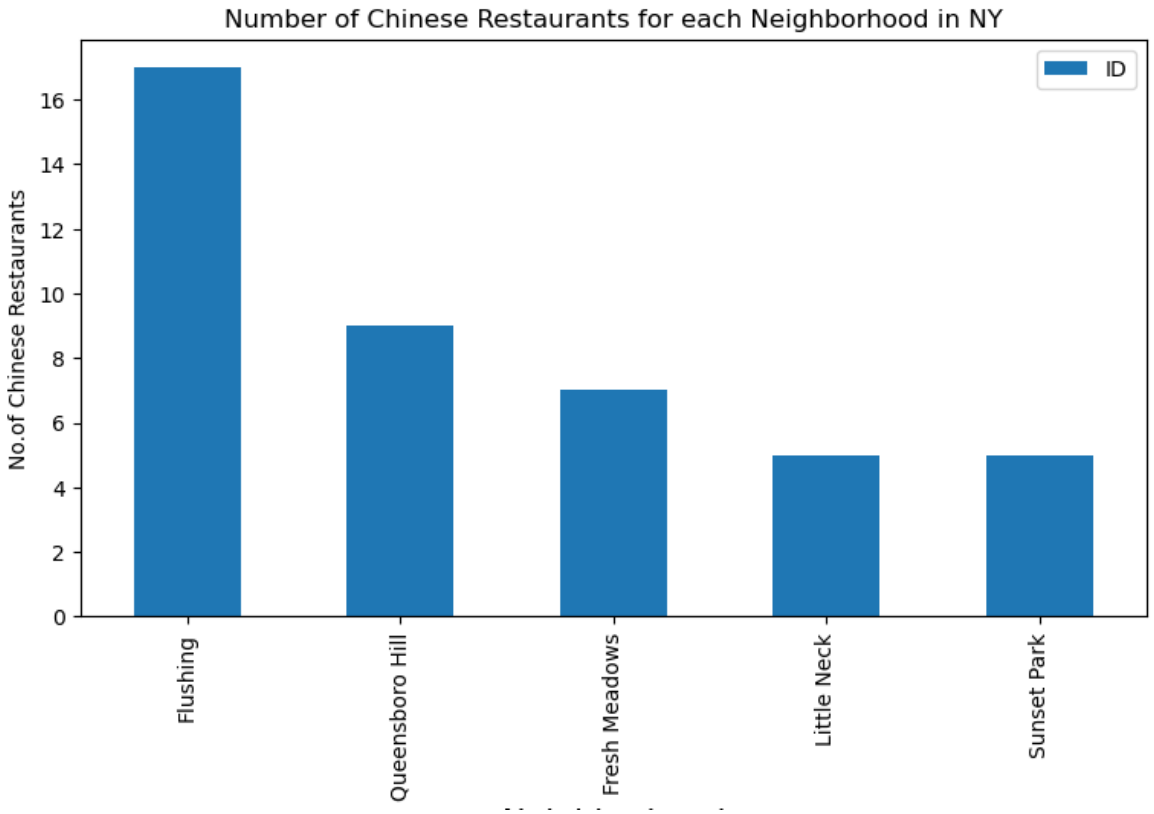
We see that Queens has the highest number of Neighbourhoods.



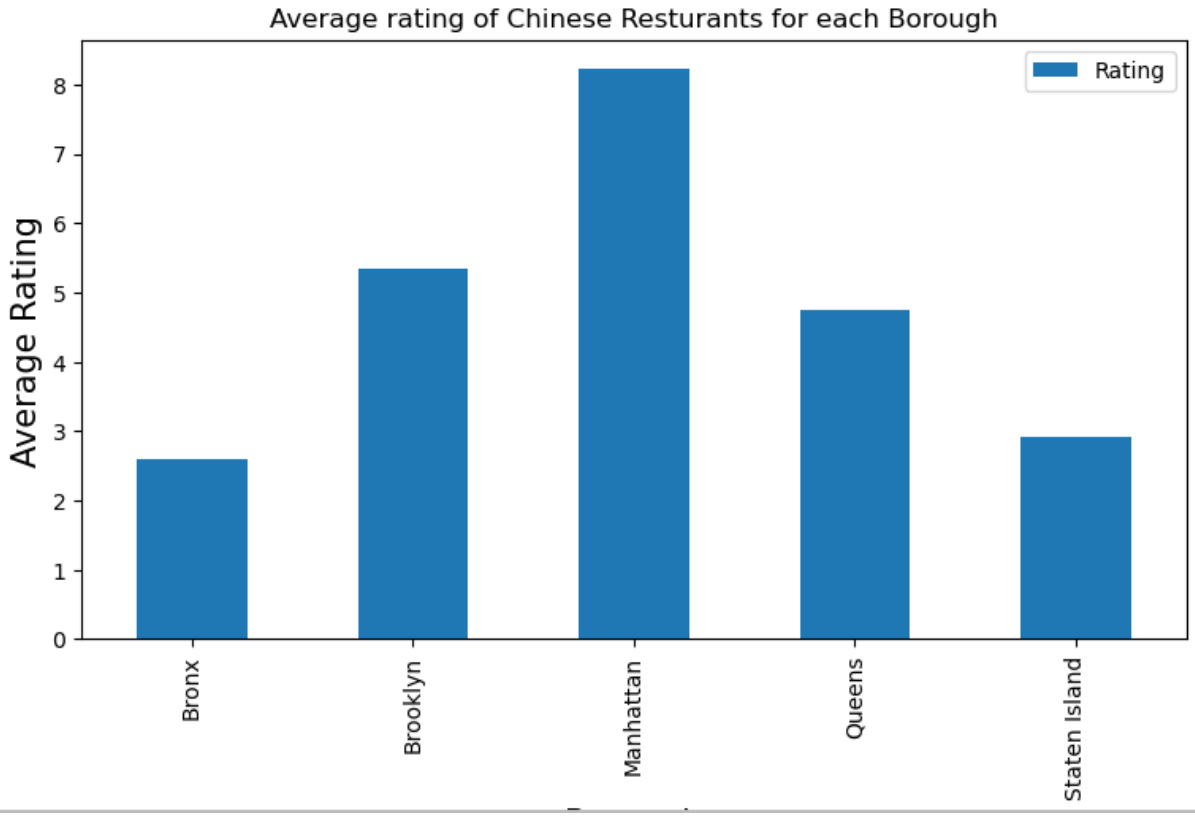
Also Queens has the highest number of Chinese Restaurants.



Flushing in Queens has the highest number of Chinese Restaurants with a total count of 17.



Manhattan has the highest average rating for Chinese Restaurants



And, finally, Chinese restaurants in New York that has maximum likes, rating and tips.

```
[24]: # Resturant with maximum Likes
chinese_rest_stats_ny.iloc[chinese_rest_stats_ny['Likes'].idxmax()]
```

```
[24]: Borough                        Manhattan
Neighborhood                       Chelsea
ID              44e9b421f964a520a5371fe3
Name                             Buddakan
Likes                                1489
Rating                                8.8
Tips                                  521
Name: 135, dtype: object
```

```
[25]: # Resturant with maximum Rating
chinese_rest_stats_ny.iloc[chinese_rest_stats_ny['Rating'].idxmax()]
```

```
[25]: Borough                        Manhattan
Neighborhood                 Lower East Side
ID              5b380f649deb7d00399fdf9d
Name                 Kings County Imperial
Likes                                   67
Rating                                   9
Tips                                    10
Name: 137, dtype: object
```

```
[26]: # Resturant with maximum Tips
chinese_rest_stats_ny.iloc[chinese_rest_stats_ny['Tips'].idxmax()]
```

```
[26]: Borough                          Manhattan
      Neighborhood                       Chelsea
      ID              44e9b421f964a520a5371fe3
      Name                             Buddakan
      Likes                                1489
      Rating                                8.8
      Tips                                  521
      Name: 135, dtype: object
```

We will consider all the neighborhoods with average rating greater or equal 8.0 to visualize on map
We will join this dataset to original new york data to get lonitude and latitude
Now we will show this data on a map

Discussion.

Questions that can be asked using the above mentioned datasets

What is best location in New York City for Chinese Cuisine ?
Which areas have potential Chinese Resturant Market ?
Which all areas lack Chinese Resturants ?
Which is the best place to stay if I prefer Chinese Cuisine ?

Conclusion

Queens borough has the largest amount of chinese reatuarants in NY, far ahead of others. Bronx, Manhattan and Staten island have not large amount, relatively equal in size.
Manhattan has an excellent rating of chinese restaraunts, while Brooklyn and Queens have a moderate one, Bronx and Staten island with poorest rating.
Buddakan (Chelsea, Manhattan) is the most popular chinese restaurant in NY, and Kings County Imperial (Lower East Side, Manhattan) is the most respectable and refined.
Manhattan has significant potential for the opening of new restaurants.
Since Queens is a popular area and not very highly rated, its potential for chinese restaurant market requires further study. It also colud be recommended for for visitors who are not too demanding on the quality of the establishment.
Brooklyn is obviously a much better borough based on the price / quality criterion, than Bronx and Staten Island.

Limitations
The ranking is purely on basis of rating of resturants
The accuracy of data depends purely depends on the data provided by FourSquare