Capstone Project - The Battle of the Neighbourhoods - Complete

Topic: Finding the best Mexican Restaurants in New York City

Table of contents

Introduction: Business Problem

"There's no good **Mexican food** in New York City."

That's what most people say, especially the Californians and the Texans, and that's when they're putting it politely. In an early scene of the second episode of **Ugly Delicious**, as David Chang talks about his preferences for tortillas and burritos over tacos, the late Jonathan Gold jokingly replies, "You're from New York. Of course you don't like tacos. You haven't had good tacos."

Gold makes the joke knowing it belies the truth. See, there are two major camps of New York Mexican food haters: Folks from out west visiting New York — well, Manhattan — who are looking for the sit-down Tex-Mex or Cal-Mex places where one can order plates by the number, with sour cream, enchilada gravy, and melted yellow cheese. Then there are the people who only want to go to the kinds of places where one orders in Spanish and the tacos come with onion, cilantro, slices of lime, and fresh, colourful salsas; they claim that there's no good Mexican food in New York because, well, there are not as many **Mexicans** here to make it.

Well to prove this wrong and as a part of my project, I will list and visualize all the major areas of New York City that have best Mexican eat outs.

Target Audience

This project will benefit Mexican expats, aspiring Restaurant owners and Mexican food enthusiasts equally by answering the following questions too:

1. Which is the best place to stay if I prefer Mexican cuisine?
2. What is best location in New York City for Mexican cuisine?
3. Which areas have potential Mexican restaurant market?
4. What areas lack Mexican eateries?

Data

For this project we will require the following data set:

1. New York City data that contains list of Boroughs, Neighbourhoods along with their latitude and longitude. Data source : https://cocl.us/new_york_dataset . This data set contains all the required information related to various neighbourhoods of New York City.
2. Mexican restaurants in each neighbourhood of New York City. Data source: Foursquare API. By using this api we will get all the venues in each New York neighbourhoods. We can filter these venues to get only Mexican restaurants.
3. Geo-Space data. Data source : https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm. By using this geo space data we will get the New York neighbourhood boundaries that will help us visualize choropleth map.

Data Extraction Approach

1. We will collect the New York city data from https://cocl.us/new_york_dataset
2. Using **FourSquare API** we will find all venues for each neighbourhood.
3. We will filter out all venues that are Mexican restaurants.
4. We will find rating, tips and likes count for each Mexican restaurants using **FourSquare API**.
5. By using **rating** for each restaurant, we will sort the data.
6. To get the latitude and longitude coordinates we will use **Geopy** library (Python)
7. We will visualize the ranking of neighbourhoods using **folium** library (python).

Analysis

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

```python
import pandas as pd
import numpy as np
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)
import requests
from bs4 import BeautifulSoup
import geocoder
import os
import folium # map rendering library
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
# Matplotlib and associated plotting modules
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import matplotlib.colors as colors
%matplotlib inline
import math


print('Libraries imported.')
```

```
Libraries imported.
```

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

```python
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 interact 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.

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

    #set variables
    radius=1000
    LIMIT=100

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

    # get all the data
    results = requests.get(url).json()
    venue_data=results["response"]['groups'][0]['items']
    venue_details=[]
    for row in venue_data:
        try:
            venue_id=row['venue']['id']
            venue_name=row['venue']['name']
            venue_category=row['venue']['categories'][0]['name']
            venue_details.append([venue_id,venue_name,venue_category])
        except KeyError:
            pass

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

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

```python
def get_venue_details(venue_id):

    #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 function to get the New York City data such as Boroughs, Neighbourhoods along with their latitude and longitude.

```python
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 function to get the New York city data.

```python
# get new york data
new_york_data=get_new_york_data()
```

```python
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

```python
new_york_data.shape
```

```
(306, 4)
```

So there are total of 306 different Neighbourhoods in New York City.

```python
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Neighborhood for each Borough in New York City')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Neighborhood', fontsize=15)
#giving a bar plot
new_york_data.groupby('Borough')['Neighborhood'].count().plot(kind='bar')
#Legend
plt.legend()
#displays the plot
plt.show()
```

Number of Neighborhood for each Borough in New York City
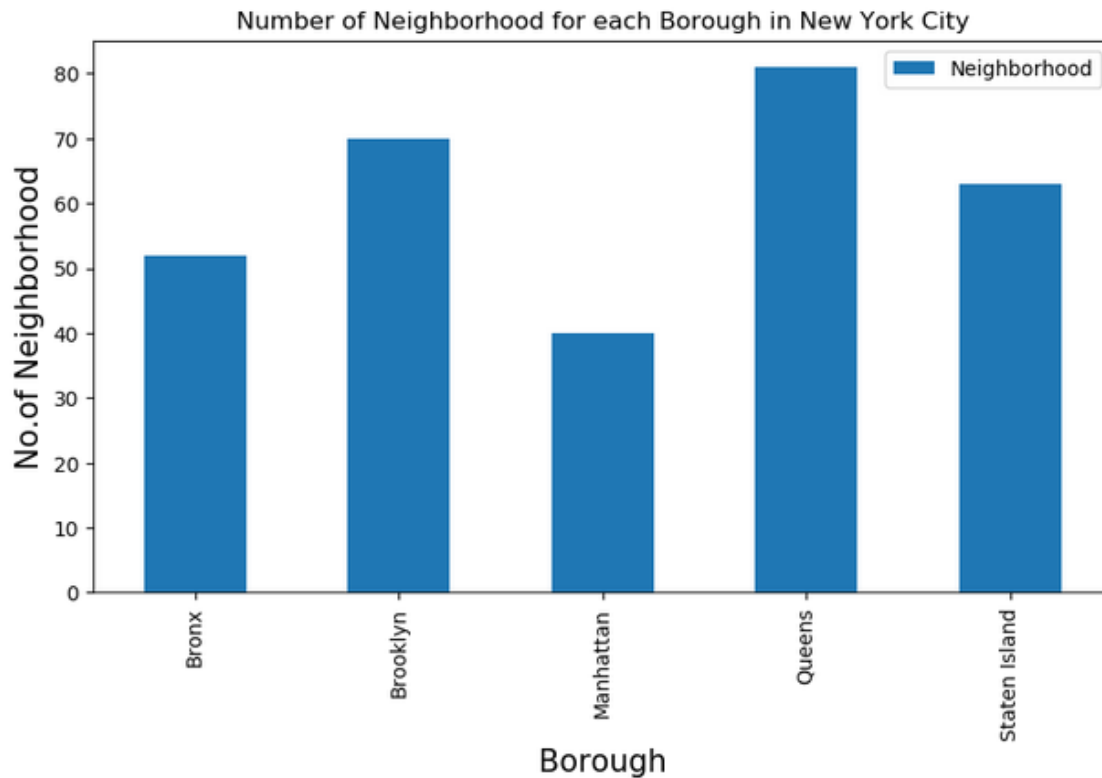
We see that Queens has highest number of neighbourhoods. Now we will collect Mexican restaurants for each Neighbourhood.

```
[29]: # prepare neighborhood list that contains Mexican restaurants
column_names=['Borough', 'Neighborhood', 'ID','Name']
Mexican_rest_ny=pd.DataFrame(columns=column_names)
count=1
for row in new_york_data.values.tolist():
    Borough, Neighborhood, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    Mexican_restaurants=venues[venues['Category']=='Mexican Restaurant']
    print('(',count,'/',len(new_york_data),')','Mexican Restaurants in '+Neighborhood+', '+Borough+':'+str(len(Mexican_restaurants)))
    for restaurant_detail in Mexican_restaurants.values.tolist():
        id, name , category=restaurant_detail
        Mexican_rest_ny = Mexican_rest_ny.append({'Borough': Borough,
                                                  'Neighborhood': Neighborhood,
                                                  'ID': id,
                                                  'Name' : name
                                                 }, ignore_index=True)
    count+=1
```

```
( 1 / 306 ) Mexican Restaurants in Wakefield, Bronx:0
( 2 / 306 ) Mexican Restaurants in Co-op City, Bronx:2
( 3 / 306 ) Mexican Restaurants in Eastchester, Bronx:0
( 4 / 306 ) Mexican Restaurants in Fieldston, Bronx:2
( 5 / 306 ) Mexican Restaurants in Riverdale, Bronx:2
( 6 / 306 ) Mexican Restaurants in Kingsbridge, Bronx:6
( 7 / 306 ) Mexican Restaurants in Marble Hill, Manhattan:4
( 8 / 306 ) Mexican Restaurants in Woodlawn, Bronx:0
( 9 / 306 ) Mexican Restaurants in Norwood, Bronx:3
```

Now that we have got all the Mexican restaurants in New York City, we will analyse it.

```
Mexican_rest_ny.head()
```

	Borough	Neighborhood	ID	Name
0	Bronx	Co-op City	5c47a4b21953f3002c558218	Chipotle Mexican Grill
1	Bronx	Co-op City	58582b156431e554e165d909	Moe's Southwest Grill
2	Bronx	Fieldston	523f6e14498ef6fdefc0d690	Burrito Shop
3	Bronx	Fieldston	4b78ba71f964a5209fdf2ee3	Santa Fe Grill & Bar
4	Bronx	Riverdale	590e5d2ce96d0c61de2dcf1d	Cocina Chente

```
Mexican_rest_ny.shape
```

```
(318, 4)
```

We got 318 Mexican Restaurants across New York City.

```python
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Mexican Restaurants for each Borough in New York City')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Mexican Resturants', fontsize=15)
#giving a bar plot
Mexican_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')
#Legend
plt.legend()
#displays the plot
plt.show()
```
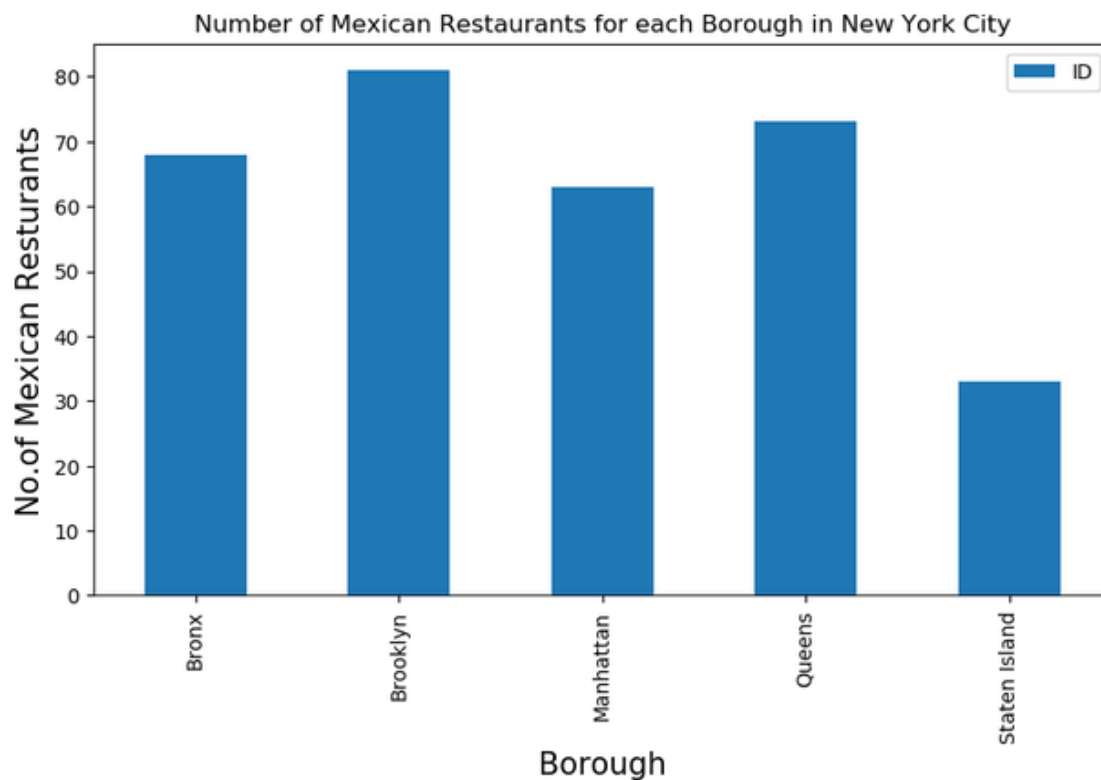

We see that Brooklyn has the largest number of Mexican restaurants

```python
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Mexican Restaurants for each Neighborhood in New York City')
#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)
#On y-axis
plt.ylabel('No.of Mexican Restaurants', fontsize=15)
#giving a bar plot
Mexican_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```

Number of Mexican Restaurants for each Neighborhood in New York City
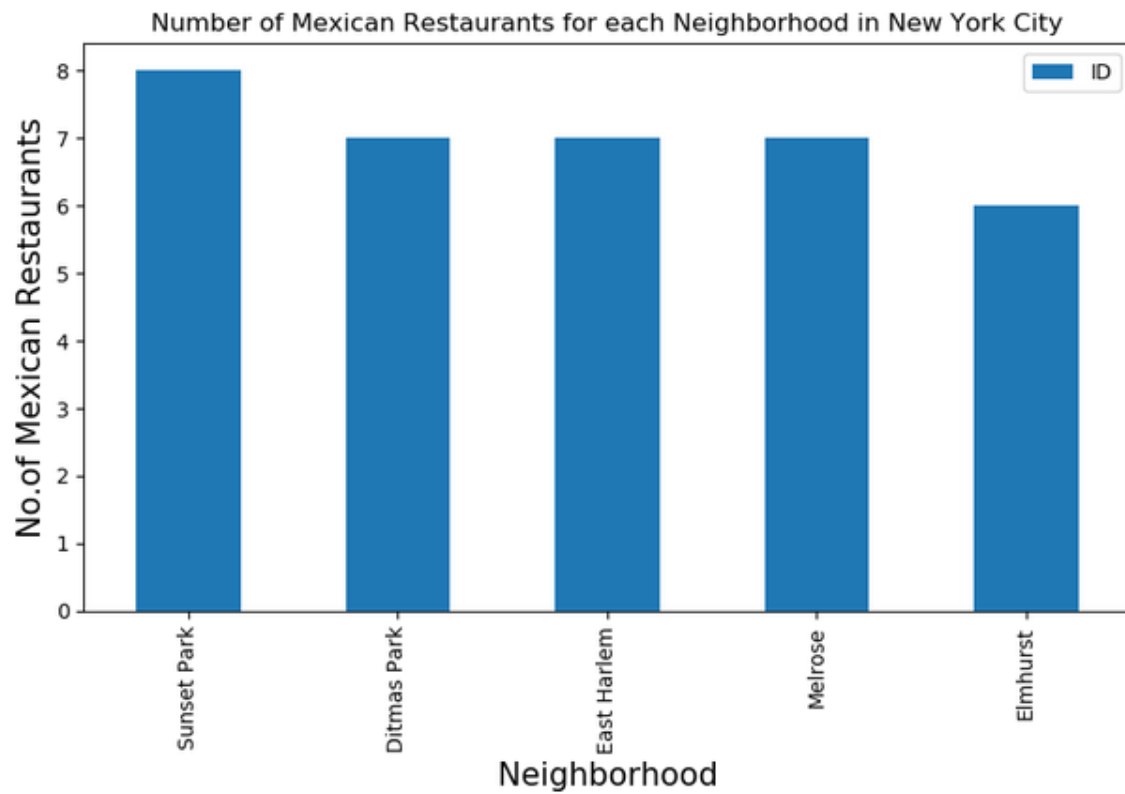

```
[34]: Mexican_rest_ny[Mexican_rest_ny['Neighborhood']=='Sunset Park']
```

[34]:

	Borough	Neighborhood	ID	Name
66	Brooklyn	Sunset Park	49bc364ff964a5201f541fe3	Tacos Matamoros
67	Brooklyn	Sunset Park	4b5415ecf964a520d3b127e3	Ricos Tacos
68	Brooklyn	Sunset Park	4b5cff55f964a520424e29e3	Maria's Bistro Mexicano
69	Brooklyn	Sunset Park	4c9521df82b56dcbc5fbd5aa	Tacos El Bronco
70	Brooklyn	Sunset Park	4bc25f7c461576b0fd5d7d32	Puebla Mini Market / Tortas a la Plancha Don Pepe
71	Brooklyn	Sunset Park	55a85c0f498ee3c735f1fa8d	Jalapeño Restaurant
72	Brooklyn	Sunset Park	4aa30695f964a520094320e3	Casa Vieja Restaurant
73	Brooklyn	Sunset Park	4d6d976d792bb60cbdd156be	Tulcingo Restaurant

So Sunset Park in Brooklyn has the highest number of Mexican Restaurants with a total count of 8.
Now we will get the ranking of each restaurant for further analysis.

```
[35]: # prepare neighborhood List that contains Mexican restaurants
column_names=['Borough', 'Neighborhood', 'ID','Name','Likes','Rating','Tips']
Mexican_rest_stats_ny=pd.DataFrame(columns=column_names)
count=1


for row in Mexican_rest_ny.values.tolist():
    Borough,Neighborhood,ID,Name=row
    try:
        venue_details=get_venue_details(ID)
        print(venue_details)
        id,name,likes,rating,tips=venue_details.values.tolist()[0]
    except IndexError:
        print('No data available for id=',ID)
        # we will assign 0 value for these restaurants as they may have been
        #recently opened or details does not exist in FourSquare Database
        id,name,likes,rating,tips=[0]*5
    print('(',count,'/',len(Mexican_rest_ny),')','processed')
    Mexican_rest_stats_ny = Mexican_rest_stats_ny.append({'Borough': Borough,
                                                          'Neighborhood': Neighborhood,
                                                          'ID': id,
                                                          'Name' : name,
                                                          'Likes' : likes,
                                                          'Rating' : rating,
                                                          'Tips' : tips
                                                          }, ignore_index=True)
    count+=1

                   ID             Name  Likes  Rating  Tips
0  5c47a4b21953f3002c558218  Chipotle Mexican Grill      0     7.5     0
( 1 / 318 ) processed
                   ID             Name  Likes  Rating  Tips
0  58582b156431e554e165d909  Moe's Southwest Grill      1     6.6     0
( 2 / 318 ) processed
                   ID             Name  Likes  Rating  Tips
```

```
[36]: Mexican_rest_stats_ny.head()
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Co-op City	5c47a4b21953f3002c558218	Chipotle Mexican Grill	0	7.5	0
1	Bronx	Co-op City	58582b156431e554e165d909	Moe's Southwest Grill	1	6.6	0
2	Bronx	Fieldston	523f6e14498ef6fdefc0d690	Burrito Shop	10	7.0	4
3	Bronx	Fieldston	4b78ba71f964a5209fdf2ee3	Santa Fe Grill & Bar	11	6.3	17
4	Bronx	Riverdale	590e5d2ce96d0c61de2dcf1d	Cocina Chente	13	8.8	5

```
[37]: Mexican_rest_stats_ny.shape
```

```
[37]: (318, 7)
```

```
[38]: Mexican_rest_ny.shape
```

```
[38]: (318, 4)
```

So we got data for all restaurants Now let's save this data to a csv sheet.

```
[39]: Mexican_rest_stats_ny.to_csv('Mexican_rest_stats_ny.csv', index=False)
```

```
[40]: Mexican_rest_stats_ny_csv=pd.read_csv('Mexican_rest_stats_ny.csv')
```

```
[41]: Mexican_rest_stats_ny_csv.shape
```

```
[41]: (318, 7)
```

```
[42]: Mexican_rest_stats_ny_csv.head()
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
0	Bronx	Co-op City	5c47a4b21953f3002c558218	Chipotle Mexican Grill	0	7.5	0
1	Bronx	Co-op City	58582b156431e554e165d909	Moe's Southwest Grill	1	6.6	0
2	Bronx	Fieldston	523f6e14498ef6fdefc0d690	Burrito Shop	10	7.0	4
3	Bronx	Fieldston	4b78ba71f964a5209fdf2ee3	Santa Fe Grill & Bar	11	6.3	17
4	Bronx	Riverdale	590e5d2ce96d0c61de2dcf1d	Cocina Chente	13	8.8	5

```
[43]: Mexican_rest_stats_ny.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 318 entries, 0 to 317
Data columns (total 7 columns):
Borough         318 non-null object
Neighborhood    318 non-null object
ID              318 non-null object
Name            318 non-null object
Likes           318 non-null object
Rating          318 non-null float64
Tips            318 non-null object
dtypes: float64(1), object(6)
memory usage: 17.5+ KB
```

We see that values like Likes, Tips are string values. We would need to convert them into float for further analysis.

```
[44]: Mexican_rest_stats_ny['Likes']=Mexican_rest_stats_ny['Likes'].astype('float64')
```

```
[45]: Mexican_rest_stats_ny['Tips']=Mexican_rest_stats_ny['Tips'].astype('float64')
```

```
[46]: Mexican_rest_stats_ny.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 318 entries, 0 to 317
Data columns (total 7 columns):
Borough         318 non-null object
Neighborhood    318 non-null object
ID              318 non-null object
Name            318 non-null object
Likes           318 non-null float64
Rating          318 non-null float64
Tips            318 non-null float64
dtypes: float64(3), object(4)
memory usage: 17.5+ KB
```

```
[48]: # Resturant with maximum Likes
      Mexican_rest_stats_ny.iloc[Mexican_rest_stats_ny['Likes'].idxmax()]
```

```
[48]: Borough                        Manhattan
      Neighborhood                 Little Italy
      ID              431e2d80f964a52079271fe3
      Name                          La Esquina
      Likes                               1978
      Rating                                 9
      Tips                                 720
      Name: 168, dtype: object
```

```
[49]:  # Resturant with maximum Rating
       Mexican_rest_stats_ny.iloc[Mexican_rest_stats_ny['Rating'].idxmax()]
```

```
[49]:  Borough                      Brooklyn
       Neighborhood                  Bushwick
       ID              5a38273a005ac129b7a93010
       Name                     La Loncheria
       Likes                             51
       Rating                           9.2
       Tips                              15
       Name: 93, dtype: object
```

```
[50]:  # Resturant with maximum Tips
       Mexican_rest_stats_ny.iloc[Mexican_rest_stats_ny['Tips'].idxmax()]
```

```
[50]:  Borough                     Manhattan
       Neighborhood               Little Italy
       ID              431e2d80f964a52079271fe3
       Name                       La Esquina
       Likes                           1978
       Rating                             9
       Tips                             720
       Name: 168, dtype: object
```

Now let's visualize neighbourhood with maximum average rating of restaurants.

```
[51]:  ny_neighborhood_stats=Mexican_rest_stats_ny.groupby('Neighborhood',as_index=False).mean()[['Neighborhood','Rating']]
       ny_neighborhood_stats.columns=['Neighborhood','Average Rating']
```

```
[52]:  ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

[52]:

	Neighborhood	Average Rating
78	Little Italy	9.000000
98	Noho	9.000000
40	East Village	9.000000
127	Soho	9.000000
48	Flatiron	8.900000
151	Windsor Terrace	8.800000
21	Chinatown	8.800000
60	Greenpoint	8.800000
58	Gramercy	8.766667
142	Upper East Side	8.700000

Above are the top neighbourhoods with top average rating of Mexican restaurants.

```
[53]:  ny_borough_stats=Mexican_rest_stats_ny.groupby('Borough',as_index=False).mean()[['Borough','Rating']]
       ny_borough_stats.columns=['Borough','Average Rating']
```

```
[54]:  ny_borough_stats.sort_values(['Average Rating'],ascending=False).head()
```

[54]:

	Borough	Average Rating
2	Manhattan	8.069841
1	Brooklyn	7.453086
3	Queens	7.439726
0	Bronx	6.751471
4	Staten Island	6.315152

Similarly these are the average rating of Mexican Restaurants for each Borough. Let's visualize it.

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Average rating of Mexican Restaurants for each Borough')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('Average Rating', fontsize=15)
#giving a bar plot
Mexican_rest_stats_ny.groupby('Borough').mean()['Rating'].plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```


Average rating of Mexican Restaurants for each Borough
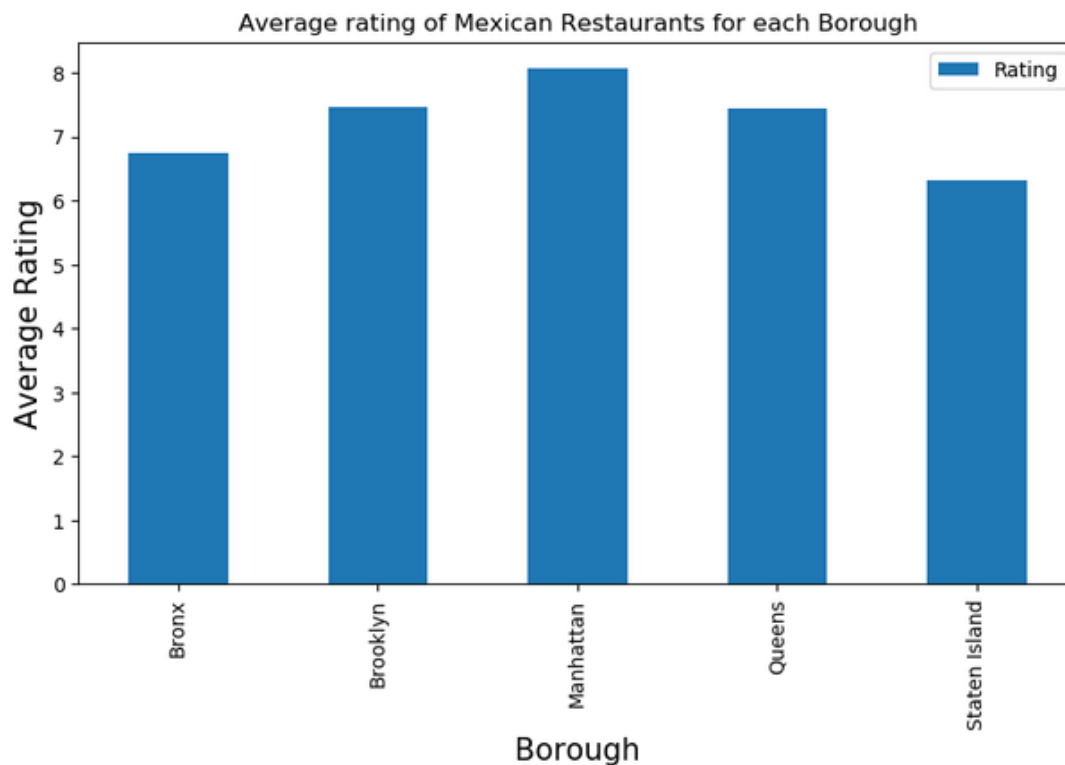
We will consider all the neighbourhoods with average rating greater or equal 9.0 to visualize on map.

```
[56]: ny_neighborhood_stats=ny_neighborhood_stats[ny_neighborhood_stats['Average Rating']>=9.0]
```

```
[57]: ny_neighborhood_stats
```

[57]:

	Neighborhood	Average Rating
40	East Village	9.0
78	Little Italy	9.0
98	Noho	9.0
127	Soho	9.0

```
[58]: ny_neighborhood_stats=pd.merge(ny_neighborhood_stats,new_york_data, on='Neighborhood')
```

```
[59]: ny_neighborhood_stats=ny_neighborhood_stats[['Borough','Neighborhood','Latitude','Longitude','Average Rating']]
```

```
[60]: ny_neighborhood_stats
```

[60]:

|	Borough	Neighborhood	Latitude	Longitude	Average Rating
0	Manhattan	East Village	40.727847	-73.982226	9.0
1	Manhattan	Little Italy	40.719324	-73.997305	9.0
2	Manhattan	Noho	40.723259	-73.988434	9.0
3	Manhattan	Soho	40.722184	-74.000657	9.0

Now we will show this data on a map.

```
[61]:  # create map and display it
       ny_map = folium.Map(location=geo_location('New York'), zoom_start=12)
```

```
[62]:  # instantiate a feature group for the incidents in the dataframe
       incidents = folium.map.FeatureGroup()

       # Loop through the 100 crimes and add each to the incidents feature group
       for lat, lng, in ny_neighborhood_stats[['Latitude','Longitude']].values:
           incidents.add_child(
               folium.CircleMarker(
                   [lat, lng],
                   radius=10, # define how big you want the circle markers to be
                   color='yellow',
                   fill=True,
                   fill_color='blue',
                   fill_opacity=0.6
               )
           )
```

```
[63]:  ny_neighborhood_stats['Label']=ny_neighborhood_stats['Neighborhood']+', '+ny_neighborhood_stats['Borough']+'('+ny_neighborhood_stats['
```

```
[64]:  # add pop-up text to each marker on the map
       for lat, lng, label in ny_neighborhood_stats[['Latitude','Longitude','Label']].values:
           folium.Marker([lat, lng], popup=label).add_to(ny_map)
       # add incidents to map
       ny_map.add_child(incidents)
```
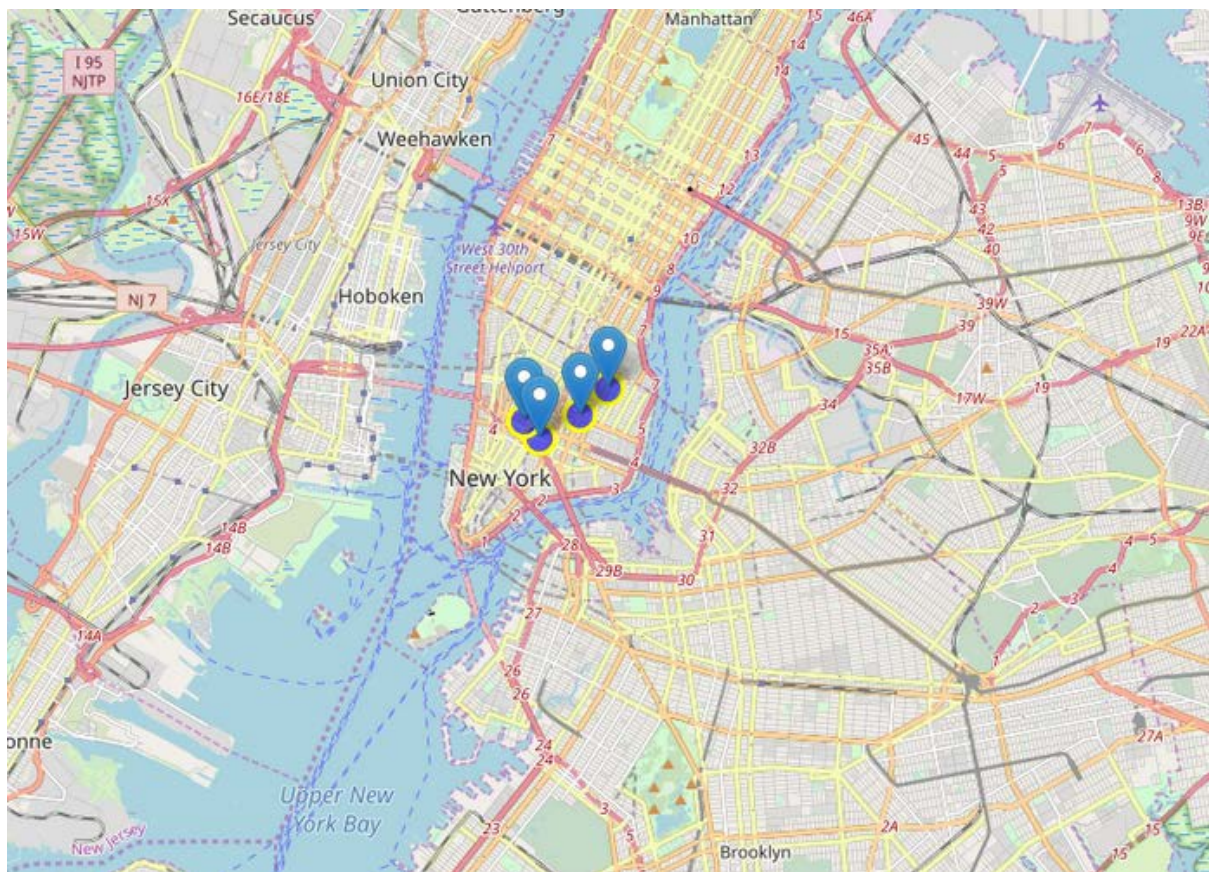

Now that we have visualized the Neighbourhoods. Let's Visualize Boroughs based on average Rating.

```
[3]: ny_map = folium.Map(location=geo_location('New York'), zoom_start=12)
     ny_geo = r'Borough Boundaries.geojson'

     ny_map.choropleth(
         geo_data=ny_geo,
         data=ny_borough_stats,
         columns=['Borough', 'Average Rating'],
         key_on='feature.properties.boro_name',
         fill_color='YlOrRd',
         fill_opacity=0.7,
         line_opacity=0.2,
         legend_name='Average Rating'
     )

     # display map
     # as this is huge map data , we will save it to a file
     ny_map.save('borough_rating.html')
```

The saved image can be downloaded at : https://github.com/Qaisar-prog/Coursera/blob/master/borough_rating.html

Conclusion

1. East Village, Little Italy, and Noho in Manhattan are some of the best places for Mexican cuisine.

2. Manhattan have potential Mexican Restaurant Market.

3. Staten Island ranks last in average rating of Mexican Restaurants.

4. Manhattan is the best place to stay if you prefer Mexican Cuisine.