Where is the best place in NYC Neighborhood to start a Filipino Restaurant?

The Battle of the Neighborhood - Coursera IBM Capstone Project

Author: Julius Benn L. Sabeniano

Background

Restaurants are a notoriously difficult business to own or operate. Not only is it a relatively capital and employee-intensive business, restaurants are also highly regulated, low margin and in most cases have a plethora of competition to deal with. There are more dining establishments and more diners now than during any other time in history; the National Restaurant Association reports over 1 million restaurant locations in the U.S. alone, a particular boon considering more than half the American population visits at least one of them each week. To fill a niche and remain relevant, prospective, as well as established, restaurants have to hedge their bets with well-rounded and well-directed dataset. Data science provides valuable insights regarding market trends and evolving consumer lifestyles so that restaurateurs can better address and meet public demand.

I.Business Problem

No single restaurant is ever going to appeal to everyone. Some people like quiet, intimate settings; others prefer boisterous ones. Some people want to bring their kids along; others want to dine alone. The preferences are as varied as the possible offerings, with specific generational cohorts preferring one thing, "people who like Indian food" preferring another and everyone else liking a million other things in between. My client, a successful Filipino restaurant chain in Philippines is looking to expand operation into North America through New York (NYC).

They are interested in building in an area that meets the following criteria:

- Above average economy. (GDP)

- Average to above average total population.

- Average to above average density. (People/sqkm)

- Average to above average Asian ethnic population.

- Neighborhood with above average Rating Filipino restaurant.

II.Data Requirements and Resources

The necessary information needed by the Client will come from the following sources:

Part 1: NYC Population & Demographic Characteristics

Data source: https://en.wikipedia.org/wiki/New_York_City and
https://en.wikipedia.org/wiki/Demographics_of_New_York_City.

Web scraping techniques was used to get NYC's population density and demographics data
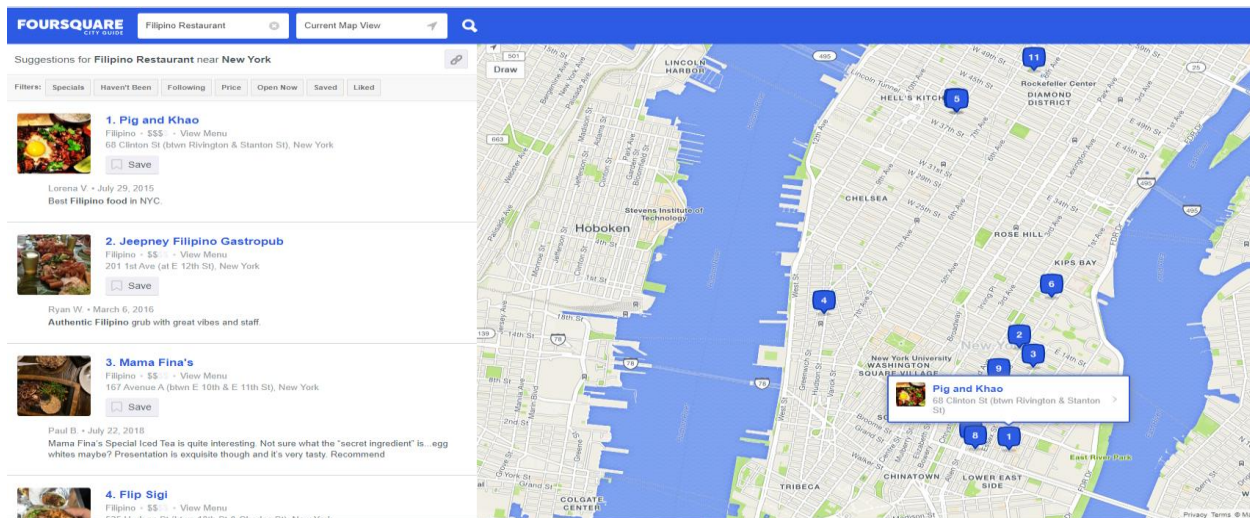from Wikipedia.

Tables from Wikipedia

Part 2: Who are the competitors in that location?

Data source: https://cocl.us/new_york_dataset and Foursquare API. These dataset was used to
explore various neighborhoods and each Filipino restaurants venues in the neighborhood.

Screenshot of a Foursquare Website

III. Methodology

In order to establish the targeted neighborhood(s), we will explore the demographics of the neighborhoods in the city of New York by segmenting the data and conducting descriptive analysis using Panda. Additional data will be extracted by web scraping and API will be used to generate data.

Data Group 1 (Population and Demographic Data)

1. Data was pulled from Wikipedia by web scraping and using python library beautiful soup.

```
In [5]:  #web scrapping data table from wikipedia using bs4
         response_obj = requests.get('https://en.wikipedia.org/wiki/New_York_City').text
         soup = BeautifulSoup(response_obj,'lxml')
         Neighborhoods_NYC_Table = soup.find('table', {'class':'wikitable sortable'})
```

```
In [10]:  #checking website table from wikipedia
          rows = Neighborhoods_NYC_Table.select("tbody > tr")[3:8]

          boroughs = []
          for row in rows:
              borough = {}
              tds = row.select('td')
              borough["borough"] = tds[0].text.strip()
              borough["county"] = tds[1].text.strip()
              borough["population"] = float(tds[2].text.strip().replace(",",""))
              borough["gdp_billions"] = float(tds[3].text.strip().replace(",",""))
              borough["gdp_per_capita"] = float(tds[4].text.strip().replace(",",""))
              borough["land_sqm"] = float(tds[5].text.strip().replace(",",""))
              borough["land_sqkm"] = float(tds[6].text.strip().replace(",",""))
              borough["persons_sqm"] = float(tds[7].text.strip().replace(",",""))
              borough["persons_sqkm"] = float(tds[8].text.strip().replace(",",""))

              boroughs.append(borough)

          print(boroughs)

          df = pd.DataFrame(boroughs, columns=["borough","county", "population", "gdp_per_capita", "persons_sqkm"])
          df
```

```
[{'borough': 'The Bronx', 'county': 'Bronx', 'population': 1432132.0, 'gdp_billions': 42.695, 'gdp_per_capita': 29200.0, 'land_
sqm': 42.1, 'land_sqkm': 109.04, 'persons_sqm': 34653.0, 'persons_sqkm': 13231.0}, {'borough': 'Brooklyn', 'county': 'Kings',
'population': 2582830.0, 'gdp_billions': 91.559, 'gdp_per_capita': 34600.0, 'land_sqm': 70.82, 'land_sqkm': 183.42, 'persons_sq
m': 37137.0, 'persons_sqkm': 14649.0}, {'borough': 'Manhattan', 'county': 'New York', 'population': 1628701.0, 'gdp_billions':
600.244, 'gdp_per_capita': 360900.0, 'land_sqm': 22.83, 'land_sqkm': 59.13, 'persons_sqm': 72033.0, 'persons_sqkm': 27826.0},
{'borough': 'Queens', 'county': 'Queens', 'population': 2278906.0, 'gdp_billions': 93.31, 'gdp_per_capita': 39600.0, 'land_sq
m': 108.53, 'land_sqkm': 281.09, 'persons_sqm': 21460.0, 'persons_sqkm': 8354.0}, {'borough': 'Staten Island', 'county': 'Richm
ond', 'population': 476179.0, 'gdp_billions': 14.514, 'gdp_per_capita': 30300.0, 'land_sqm': 58.37, 'land_sqkm': 151.18, 'perso
ns_sqm': 8112.0, 'persons_sqkm': 3132.0}]
```

Out[10]:

	borough	county	population	gdp_per_capita	persons_sqkm
0	The Bronx	Bronx	1432132.0	29200.0	13231.0
1	Brooklyn	Kings	2582830.0	34600.0	14649.0
2	Manhattan	New York	1628701.0	360900.0	27826.0
3	Queens	Queens	2278906.0	39600.0	8354.0
4	Staten Island	Richmond	476179.0	30300.0	3132.0

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

Data Group 2 (NYC Geographical Neighbourhood Data)

1. Retrieved Borough, Neighbourhood and Coordinates data from
   https://cocl.us/new_york_dataset , transformed to data frame and plot it using Bar
   chart.

There is a total of 306 Neighborhoods in New York City.

In [35]:
```python
from matplotlib import pyplot as plt
plt.style.use('ggplot')

plt.figure(figsize=(9,5), dpi = 80)
# title
plt.title('Number of Neighbourhood in NYC by Borough')
#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()
plt.show()
```
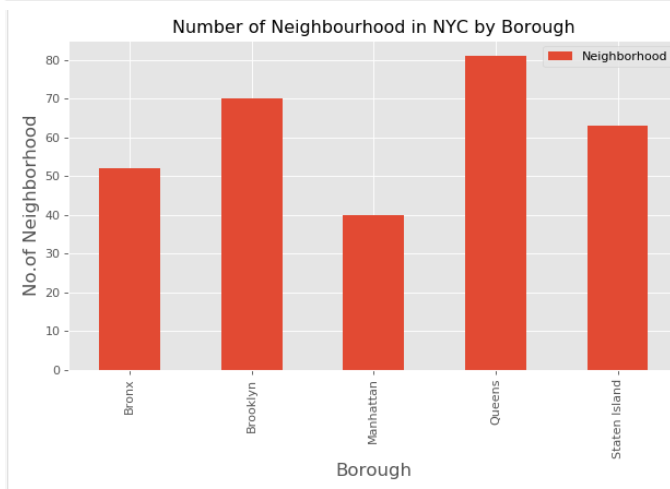

Queens is the most densely populated borough in New York City with 80 neighborhoods

2. Used Foursquare API to retrieve the number of Filipino restaurant in NYC Borough.

In [36]:
```python
# prepare neighborhood list that contains filipino restaurant
column_names=['Borough', 'Neighborhood', 'ID','Name']
filip_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)
    filip_resturants=venues[venues['Category']=='Filipino Restaurant']
    print('(',count,'/',len(new_york_data),')','Filipino Restaurants in '+Neighborhood+', '+Borough+':'+str(len(filip_resturants
    for resturant_detail in filip_resturants.values.tolist():
        id, name , category=resturant_detail
        filip_rest_ny = filip_rest_ny.append({'Borough': Borough,
                                              'Neighborhood': Neighborhood,
                                              'ID': id,
                                              'Name' : name
                                             }, ignore_index=True)
    count+=1
```

```
( 182 / 306 ) Filipino Restaurants in Floral Park, Queens:0
( 183 / 306 ) Filipino Restaurants in Holliswood, Queens:0
( 184 / 306 ) Filipino Restaurants in Jamaica Estates, Queens:0
( 185 / 306 ) Filipino Restaurants in Queensboro Hill, Queens:0
( 186 / 306 ) Filipino Restaurants in Hillcrest, Queens:0
( 187 / 306 ) Filipino Restaurants in Ravenswood, Queens:0
( 188 / 306 ) Filipino Restaurants in Lindenwood, Queens:0
( 189 / 306 ) Filipino Restaurants in Laurelton, Queens:0
( 190 / 306 ) Filipino Restaurants in Lefrak City, Queens:0
( 191 / 306 ) Filipino Restaurants in Belle Harbor, Queens:0
( 192 / 306 ) Filipino Restaurants in Rockaway Park, Queens:0
( 193 / 306 ) Filipino Restaurants in Somerville, Queens:0
( 194 / 306 ) Filipino Restaurants in Brookville, Queens:0
( 195 / 306 ) Filipino Restaurants in Bellaire, Queens:0
( 196 / 306 ) Filipino Restaurants in North Corona, Queens:0
( 197 / 306 ) Filipino Restaurants in Forest Hills Gardens, Queens:0
( 198 / 306 ) Filipino Restaurants in St. George, Staten Island:0
( 199 / 306 ) Filipino Restaurants in New Brighton, Staten Island:0
( 200 / 306 ) Filipino Restaurants in Stapleton, Staten Island:0
( 201 / 306 ) Filipino Restaurants in Rosebank, Staten Island:1
```

In [37]: `filip_rest_ny.shape`

Out[37]: (20, 4)

In [40]:
```python
from matplotlib import pyplot as plt
plt.style.use('ggplot')

plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Filipino Restaurants in NYC by Borough')
#On x-axis
plt.xlabel('Borough', fontsize = 15)
#On y-axis
plt.ylabel('No.of Filipino Restaurant', fontsize=15)
#giving a bar plot
filip_rest_ny.groupby('Borough')['ID'].count().plot(kind='bar')
#legend
plt.legend()
#displays the plot
plt.show()
```
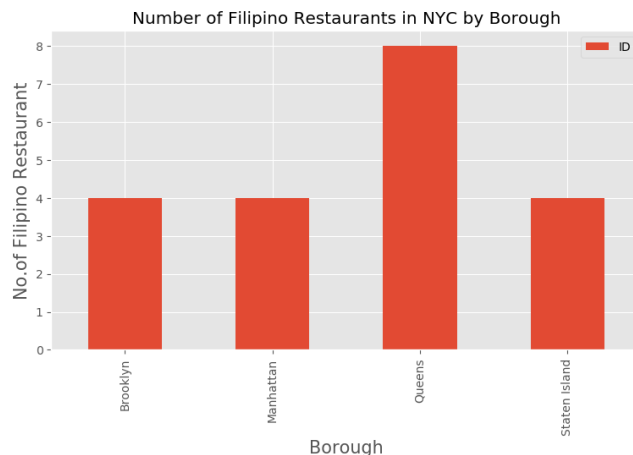

Queens has the highest number of Filipino Restaurant.

3. Retrieved the number of Filipino restaurants in NYC by Neighboorhood.

```
In [44]: from matplotlib import pyplot as plt
         plt.style.use('ggplot')

         plt.figure(figsize=(9,5), dpi = 100)
         # title
         plt.title('Number of Filipino Restaurants in NYC by Neighbourhood')
         #On x-axis
         plt.xlabel('Neighborhood', fontsize = 15)
         #On y-axis
         plt.ylabel('No.of Filipino Restaurants', fontsize=15)
         #giving a bar plot
         filip_rest_ny.groupby('Neighborhood')['ID'].count().nlargest(5).plot(kind='bar')
         #Legend
         plt.legend()
         #displays the plot
         plt.show()
```
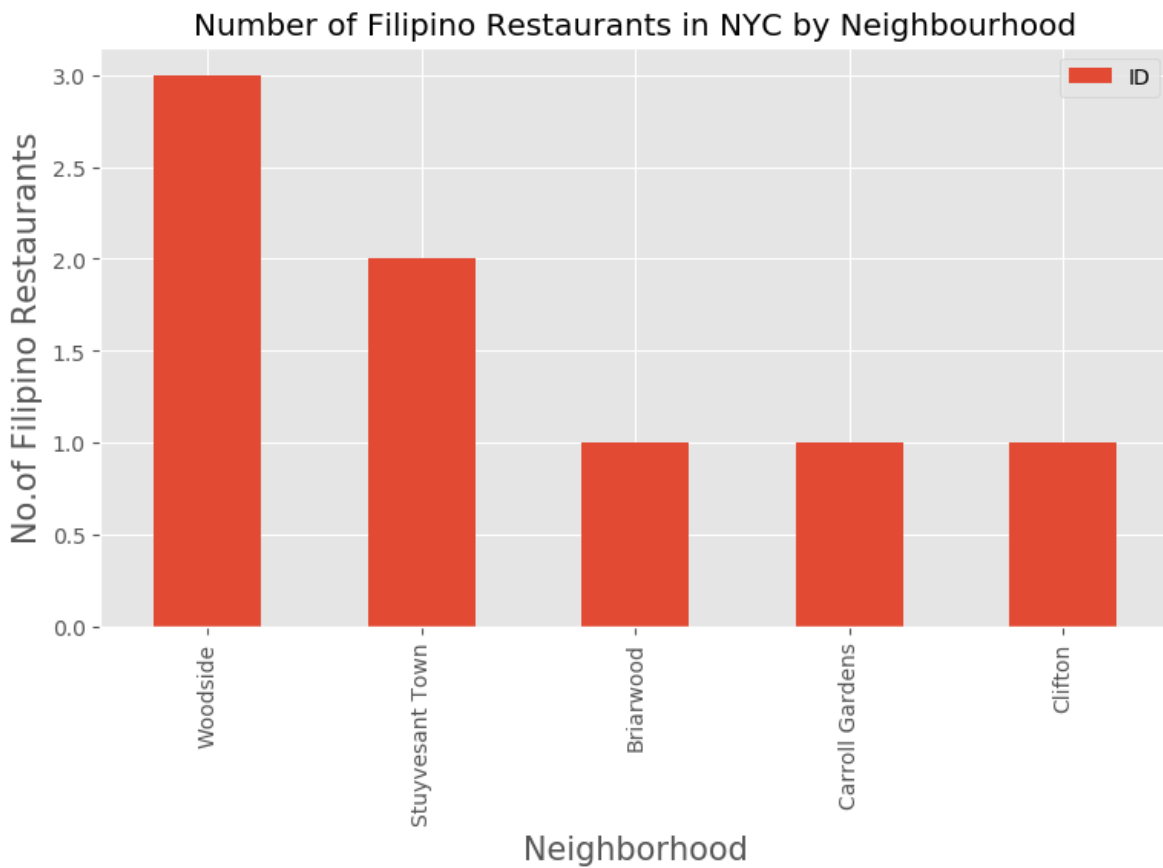
Number of Filipino Restaurants in NYC by Neighbourhood

Woodside has the most Filipino Restaurant with a total of 3.

4. Called again to get the Neighbourhood's Name, Likes, Rating and Tips.

```
In [47]: # prepare neighborhood list that contains Filipino resturants
         column_names=['Borough', 'Neighborhood', 'ID','Name','Likes','Rating','Tips']
         filip_rest_stats_ny=pd.DataFrame(columns=column_names)
         count=1


         for row in filip_rest_ny.values.tolist():
             Borough,Neighborhood,ID,Name=row

             # prepare neighborhood list that contains indian resturants
         column_names=['Borough', 'Neighborhood', 'ID','Name','Likes','Rating','Tips']
         filip_rest_stats_ny=pd.DataFrame(columns=column_names)
         count=1


         for row in filip_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, KeyError) as e:
                 print('No data available for id=',ID)
                 # we will assign 0 value for these resturants 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(filip_rest_ny),')','processed')
             filip_rest_stats_ny = filip_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   4afa2432f964a520081722e3  Purple Yam    137     7.9    76
( 1 / 20 ) processed
                          ID           Name  Likes  Rating  Tips
0   591e55677247502fb9514269  Fob Restaurant     55     7.8    28
( 2 / 20 ) processed
                          ID        Name   Likes   Rating   Tips
```

```
                          ID                         Name   Likes   Rating   \
0   4aecab47f964a52026ca21e3  Tito Rad's Grill & Restaurant    183      8.9

    Tips
0     74
( 9 / 20 ) processed
                          ID             Name   Likes   Rating   Tips
0   4c22a0bc9085d13a53af86cc  Asian Food Ltd.     13      6.3      6
( 10 / 20 ) processed
                          ID            Name   Likes   Rating   Tips
0   54c43136498ee8c2a685861e  Phil-Am Kusina     17      7.6      9
( 11 / 20 ) processed
                          ID         Name   Likes   Rating   Tips
0   4afa2432f964a520081722e3  Purple Yam     137     7.9     76
( 12 / 20 ) processed
                          ID            Name   Likes   Rating   Tips
0   54c43136498ee8c2a685861e  Phil-Am Kusina     17      7.6      9
( 13 / 20 ) processed
                          ID           Name   Likes   Rating   Tips
0   50588bc6526260483ab2860e  Pig and Khao    1081     9.1    301
( 14 / 20 ) processed
                          ID            Name   Likes   Rating   Tips
0   54c43136498ee8c2a685861e  Phil-Am Kusina     17      7.6      9
( 15 / 20 ) processed
                          ID            Name   Likes   Rating   Tips
0   54c43136498ee8c2a685861e  Phil-Am Kusina     17      7.6      9
( 16 / 20 ) processed
                          ID             Name   Likes   Rating   Tips
0   4c22a0bc9085d13a53af86cc  Asian Food Ltd.     13      6.3      6
( 17 / 20 ) processed
                          ID         Name   Likes   Rating   Tips
0   5a73b1f49ef8ef531a1c08df  Mama Fina's     44      8.2     14
( 18 / 20 ) processed
                          ID                       Name   Likes   Rating   Tips
0   503f293c19a9def238a153b3  Jeepney Filipino Gastropub    574      8.5    170
( 19 / 20 ) processed
                          ID                         Name   Likes   Rating   \
0   4aecab47f964a52026ca21e3  Tito Rad's Grill & Restaurant    183      8.9

    Tips
0     74
( 20 / 20 ) processed
```

```
In [72]: plt.figure(figsize=(9,5), dpi = 100)
         # title
         plt.title('Average rating of Filipino Restaurant in each  NYC Borough')
         #On x-axis
         plt.xlabel('Borough', fontsize = 15)
         #On y-axis
         plt.ylabel('Average Rating', fontsize=15)
         #giving a bar plot
         filip_rest_stats_ny.groupby('Borough').mean()['Rating'].plot(kind='bar')
         #Legend
         plt.legend()
         #displays the plot
         plt.show()
```
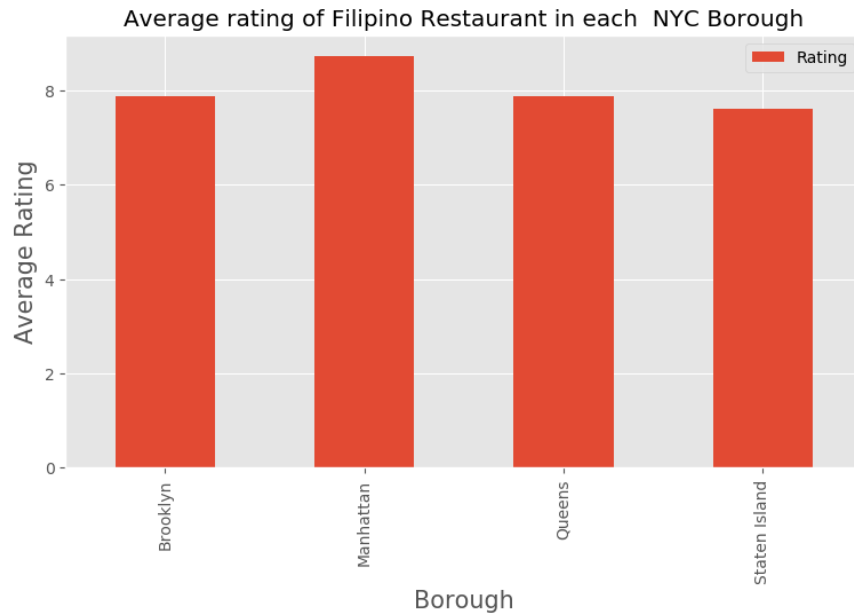

Manhattan has the average top rating Filipino Retaurant in NYC Borough.

Let us visualized in the map, All the neighborhoods with average rating greater or equal 7.0.

```
In [75]: ny_neighborhood_stats=ny_neighborhood_stats[ny_neighborhood_stats['Average Rating']>=7.0]
```

```
In [76]: ny_neighborhood_stats
```

Out[76]:

	Neighborhood	Average Rating
4	Elmhurst	8.000000
7	Lower East Side	9.100000
8	Noho	9.100000
13	Stuyvesant Town	8.350000
14	Sunnyside	8.900000
15	Sunnyside Gardens	8.900000
16	Woodside	8.166667

```
In [77]: ny_neighborhood_stats=pd.merge(ny_neighborhood_stats,new_york_data, on='Neighborhood')
```

```
In [78]:
         ny_neighborhood_stats=ny_neighborhood_stats[['Borough','Neighborhood','Latitude','Longitude','Average Rating']]
```

```
In [79]: ny_neighborhood_stats.sort_values(['Average Rating'],ascending=False).head(10)
```

Out[79]:

	Borough	Neighborhood	Latitude	Longitude	Average Rating
1	Manhattan	Lower East Side	40.717807	-73.980890	9.100000
2	Manhattan	Noho	40.723259	-73.988434	9.100000
4	Queens	Sunnyside	40.740176	-73.926916	8.900000
5	Staten Island	Sunnyside	40.612760	-74.097126	8.900000
6	Queens	Sunnyside Gardens	40.745652	-73.918193	8.900000
3	Manhattan	Stuyvesant Town	40.731000	-73.974052	8.350000
7	Queens	Woodside	40.746349	-73.901842	8.166667
0	Queens	Elmhurst	40.744049	-73.881656	8.000000

5. Analyzed the Data Group 1/Group 2 using basic statistics and plot the final Map.

```
In [80]: # create map and display it
         ny_map = folium.Map(location=(40.693943, -73.985880), zoom_start=12)
```

```
In [81]: # 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
                 )
             )
```

```
In [82]: ny_neighborhood_stats['Label']=ny_neighborhood_stats['Neighborhood']+', '+ny_neighborhood_stats['Borough']+'('+ny_neighborhood_s
```

Best neighborhoods in New York that has highest average rating for Filipino Restaurants

```
In [86]: # 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)
```
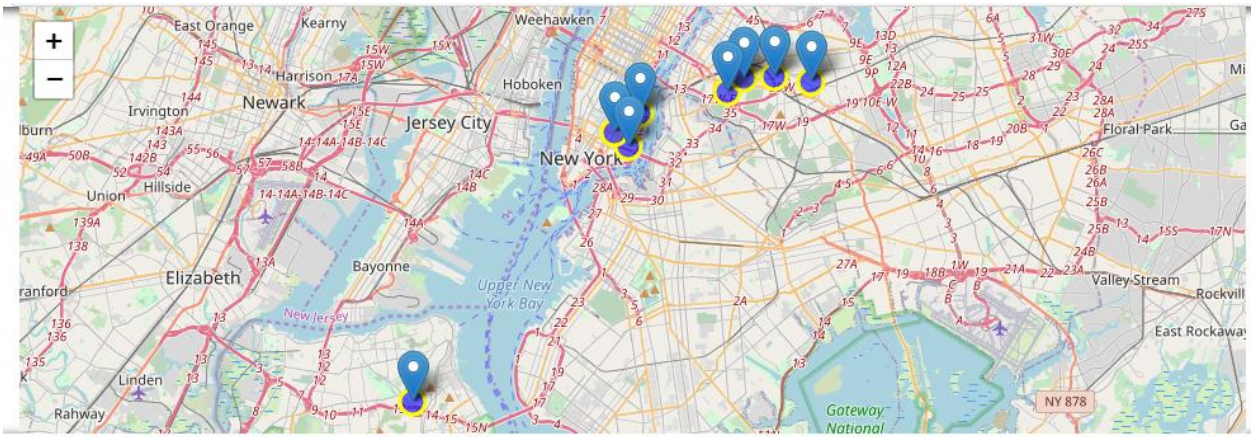

Who are the competitors in Manhattan?

```
filip_rest_stats_ny[filip_rest_stats_ny['Borough']=='Manhattan']
```

	Borough	Neighborhood	ID	Name	Likes	Rating	Tips
3	Manhattan	Lower East Side	50588bc6526260483ab2860e	Pig and Khao	1081.0	9.1	301.0
13	Manhattan	Noho	50588bc6526260483ab2860e	Pig and Khao	1081.0	9.1	301.0
17	Manhattan	Stuyvesant Town	5a73b1f49ef8ef531a1c08df	Mama Fina's	44.0	8.2	14.0
18	Manhattan	Stuyvesant Town	503f293c19a9def238a153b3	Jeepney Filipino Gastropub	574.0	8.5	170.0

6. Validated Findings/Observations and Hypothesis.

7. Constructed Conclusion and Hypothesis.

V.Discussions and Results

Part 1: NYC Population & Demographic Characteristics

- **Findings and Observations**

    1. Brooklyn is the most populous Borough but has the lowest gdp_per_capita among the Boroughs.

    2. Queen is the second most populous Borough and has higher gdp_per_capita than Brooklyn but smaller population density. Highest in-terms of asian ethnic minority population.

    3. Manhattan is the second asian ethnic minority population, first in gdp_per_capita and population density. High person per km means people are living in a high rise building like appartments, hotels and condominiums.

- **Hypothesis 1**

Manhattan could be the best place to start a Filipino Restaurant given that;

    1. Good Economy (highest gdp_per_capita).

    2. High Population density (person per km can be relate to people who are living in a high rise building

    3. Second in Asian ethnic minority population.

	jurisdiction	%_white	%_black_or_african_amercian	%_Asian	%_other	%_mixed_race	%_hispanic_latino_of_other_race
0	Queens	44.1	20.0	17.6	12.3	6.1	25.0
1	Manhattan	54.4	17.4	9.4	14.7	4.1	27.2
2	Bronx	29.9	35.6	3.0	25.7	5.8	48.4
3	Staten Island	77.6	9.7	5.7	4.3	2.7	12.1
4	NYC Total	44.7	26.6	9.8	14.0	4.9	27.0

Part 2: Where is the best location and Who are the competitors in that location?

- **Findings and Observations**

1. Lower East Side, Noho, Stuyvesant in Manhattan are some of the best neighborhoods for Filipino cuisine.

2. Staten Island Neighborhood has the lowest rated Filipino Restaurants in NYC.

3. Manhattan Borough is the best place to stay if you prefer Filipino Cuisine.

4. Bronx Borough does not have Filipino Restaurant at the moment.

5. Pig and Khao Restaurant will be the tough competitor. It has 2 branches in Manhattan (Lower East Side and Noho). Stuyvesant Town has Mama Fina's Restaurant which second behind Pig and Khao.

- **Hypothesis 2**

Manhattan could be the best place to start a Filipino Restaurant given that;

1. It has the best neigborhoods for Filipino cusine.

2. It has the top rating Filipino Restaurants.

3. It is the best place to stay if you prefer Filipino Cuisine.

VI. Conclusion and Recommendation

1. Lower East Side or Noho in Manhattan would be the best choice to start a restaurant given that;

- It is the third most populous urban area in New York City (NYC).

- It has a population density of 27,826 people per square km, and confirmed highest of any borough in the United States.

- It has a good economy (gdp_per_capita).

- It has some of the top rated Filipino restaurants located in that area

- It has the second highest Asian ethnic minority population in NYC.

- Tough competitors are there but people looks for other choices.

2. There are very few Filipino Restaurants around Manhattan and the competition is very weak therefore, this the best time to start a Filipino Restaurant in the neighborhood of Manhattan.