The Battle Of Neighborhoods

Introduction

Manhattan is the most densely populated of New York City's 5 boroughs. It's mostly made up of Manhattan Island, bounded by the Hudson, East and Harlem rivers. Among the world's major commercial, financial and cultural centers, it's the heart of "the Big Apple". New York City has been called both the most economically powerful city and the leading financial center of the world. City with high population and fast developing resources requires more food outlets. Due to its dense population there is a huge increase in demand for restaurant business and Manhattan is an ideal place to open a restaurant.

Business Problem

A entrepreneur wants to open an Mexican Restaurant in Manhattan,New York. The success of any restaurant depends on multiple factors like location, menu, interior, consistency in taste and many other. Entrepreneur approached us to help him to find a perfect location which is economic and profitable to them.

Target Audience

A entrepreneur wants to open an Mexican Restaurant in Manhattan,New York.

Data Overview

The data prepared for the purposes of the analysis from multiple sources which will provide the list of neighborhoods in New York (https://cocl.us/new_york_dataset), the Geographical location of the neighborhoods (via Geo coder package) and Venue data pertaining to Mexican restaurants (via Foursquare). The venues data will help find which neighborhood is best suitable to open an Mexican restaurant.

Methodology

- New York City data from https://cocl.us/new_york_dataset

Download the NewYork City data from https://cocl.us/new_york_dataset

Save it as 'newyorkgeo.json'

In [2]:
```python
with open('newyorkgeo.json') as json_data:
    newyorkCity_data = json.load(json_data)
newyorkCity_data
```

Out[2]:

```
{'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]}},
  {'type': 'Feature',
```

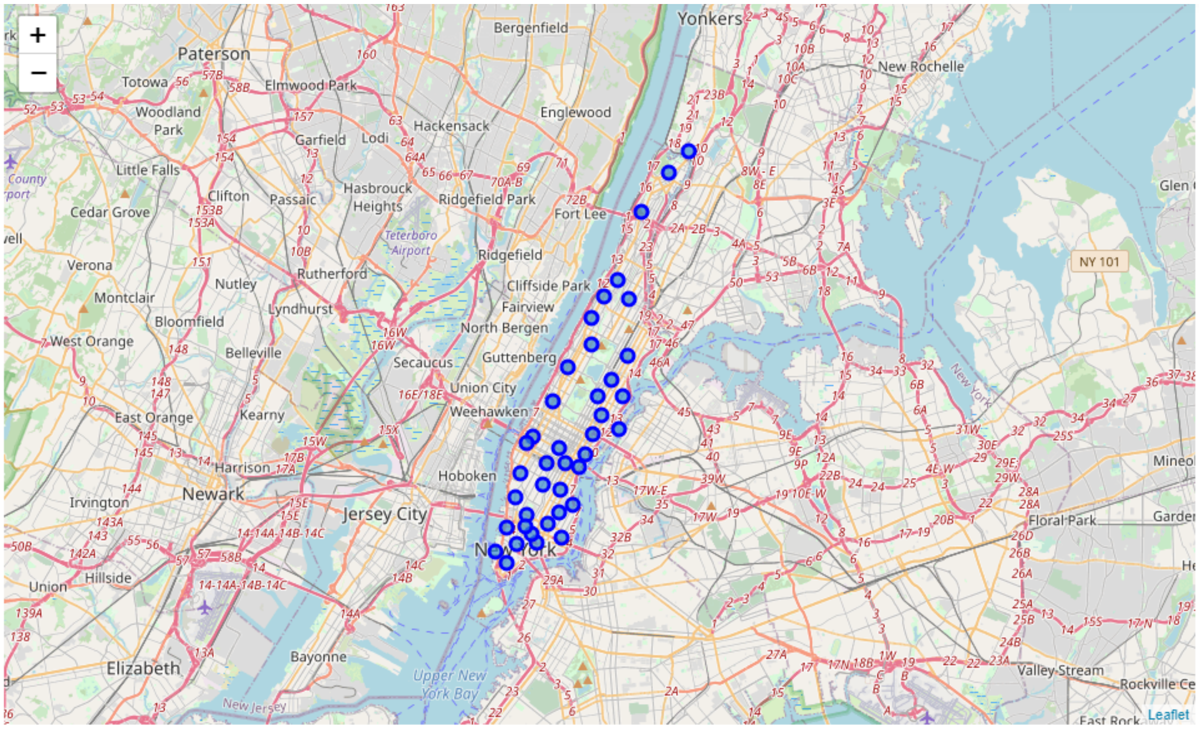

- Using geopy package we could get,Geo location like latitude and longitude of each location

creating a map of Manhattan using latitude and longitude values

In [7]:
```python
map_manhattan = folium.Map(location=[latitude, longitude], zoom_start=11)

# add markers to map
for lat, lng, label in zip(manhattan_data['Latitude'], manhattan_data['Longitude'], manhattan_data['Neighborhood']):
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_manhattan)

map_manhattan
```



- From foursquare API we get
    1. Restaurant names
    2. Id
    3. Location

The retrieval of the location, name and category about the various venues in Manhattan was collected through the Foursquare explore API. To obtain the data, it was required to make an account where it would provide a 'Secret Key' as well as a 'Client ID' which would allow me to pull any data.

	Neighborhood	Id	Name	Latitude	Longitude	Category
0	Marble Hill	5bd90c33a35dce002c1e26fc	Taqueria Sinaloense	40.874574	-73.910687	Mexican Restaurant
1	Marble Hill	590e5d2ce96d0c61de2dcf1d	Cocina Chente	40.886235	-73.907108	Mexican Restaurant
2	Marble Hill	5217dd2811d2d06ccafb77d3	Estrellita Poblana V	40.879687	-73.906257	Mexican Restaurant
3	Marble Hill	4ce81d330f196dcb5d2b43ae	Picante Picante Mexican Restaurant	40.878252	-73.902936	Mexican Restaurant
4	Marble Hill	5407a176498ee25dcaf36cf0	Mi Lindo San Miguelito	40.880023	-73.906488	Mexican Restaurant
5	Marble Hill	59fbb1cd1bc7043d43ffbf64	Guacamole	40.874511	-73.910708	Mexican Restaurant
6	Marble Hill	4babf8b4f964a5207dda3ae3	New Fresco Tortillas	40.885753	-73.910390	Mexican Restaurant
7	Marble Hill	5830ff3044587f7beb2271cc	Guacamole	40.869659	-73.916736	Mexican Restaurant
8	Marble Hill	5696a649498ee06e2efbbbd3	sazon de lupita	40.870069	-73.903562	Taco Place
9	Marble Hill	5b06e46035811b00393375b1	Amor Eterno	40.880898	-73.908693	Mexican Restaurant

- Using New York neighborhood data and Foursquare API we found the neighborhood details.
- Used K Means clustering to group the same type of neighborhoods.
- Found the places of same neighborhood as of Mexican Restaurants,more economic and less in number.

Then to analyze the data we performed a technique in which Categorical Data is transformed into Numerical Data for Machine Learning algorithms. This technique is called **One hot encoding**. For each of the neighborhoods, individual venues were turned into the frequency at how many of those Venues were located in each neighborhood.

```
# one hot encoding
manhattan_onehot = pd.get_dummies(manhattan_venues[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
manhattan_onehot['Neighborhood'] = manhattan_venues['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [manhattan_onehot.columns[-1]] + list(manhattan_onehot.columns[:-1])
manhattan_onehot = manhattan_onehot[fixed_columns]

manhattan_onehot.head()
```

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	...	Video Store	Vietnamese Restaurant	Volleyball Court
0	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	...	0	0	0

5 rows × 330 columns

Then we grouped those rows by Neighborhood and by taking the Average of the frequency of occurrence of each Venue Category.

	Neighborhood	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	...	Video Store	Vietnamese Restaurant	Volleyball Court
0	Battery Park City	0.0	0.0	0.0	0.000000	0.015873	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.0
1	Carnegie Hill	0.0	0.0	0.0	0.000000	0.011494	0.0	0.0	0.0	0.011494	...	0.0	0.022989	0.0
2	Central Harlem	0.0	0.0	0.0	0.066667	0.044444	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.0
3	Chelsea	0.0	0.0	0.0	0.000000	0.040000	0.0	0.0	0.0	0.000000	...	0.0	0.000000	0.0
4	Chinatown	0.0	0.0	0.0	0.000000	0.030000	0.0	0.0	0.0	0.000000	...	0.0	0.030000	0.0

To make the analysis more interesting, we wanted to cluster the neighborhoods based on the neighborhoods that had similar averages of Mexican Restaurants in that Neighborhood. To do this we used **K-Means** clustering.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

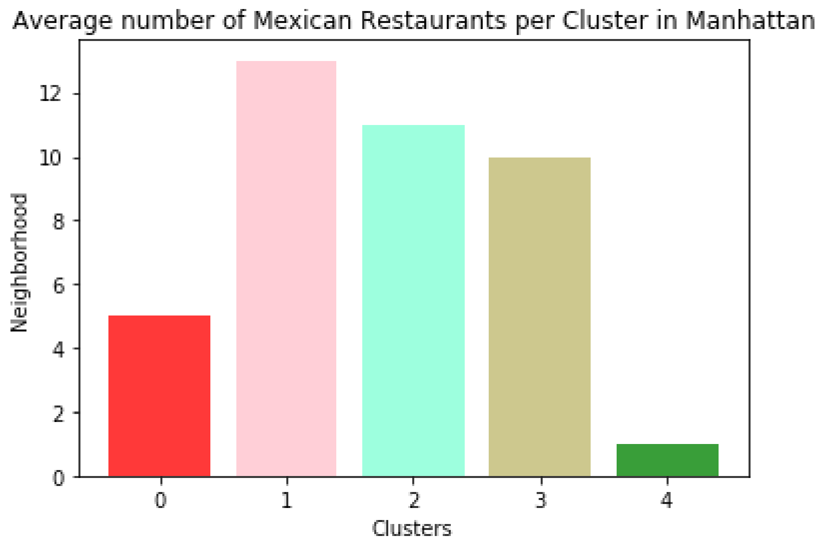
After, we created a new data frame which only stored the Neighborhood names as well as the common venues of Mexican Restaurants in that Neighborhood. This allowed the data to be summarized based on each individual Neighborhood and made the data much simpler to analyze.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Hotel	Coffee Shop	Gym	Memorial Site	Playground	Mexican Restaurant	Sandwich Place	Gourmet Shop	Shopping Mall
1	Carnegie Hill	Coffee Shop	Café	Italian Restaurant	Yoga Studio	Wine Shop	Bookstore	Gym / Fitness Center	Gym	Bar	Shipping Store
2	Central Harlem	African Restaurant	Bar	French Restaurant	American Restaurant	Gym / Fitness Center	Chinese Restaurant	Seafood Restaurant	Juice Bar	Public Art	Beer Bar
3	Chelsea	Coffee Shop	Art Gallery	Ice Cream Shop	American Restaurant	French Restaurant	Café	Italian Restaurant	Bar	Cupcake Shop	Cycle Studio
4	Chinatown	Chinese Restaurant	Bakery	Dessert Shop	American Restaurant	Ice Cream Shop	Spa	Bar	Hotpot Restaurant	Vietnamese Restaurant	Bubble Tea Shop

Clusters are labeled based on neighborhood shown in fig

	Neighborhood	Cluster Labels
0	Marble Hill	1
1	Chinatown	1
2	Washington Heights	0
3	Inwood	1
4	Hamilton Heights	0
5	Manhattanville	1
6	Central Harlem	1
7	East Harlem	2
8	Upper East Side	0
9	Yorkville	1

Then we calculated the average number of Mexican restaurants per cluster in Manhattan



Observations

Most of the Mexican Restaurants are in cluster 1 represented by the pink clusters. Looking at the nearby clusters, the optimum place to put a new Mexican Restaurant is Cluster 4 because in cluster 3 there is less Mexican restaurants compared to cluster2 and cluster1. There are many Neighborhoods in the area but less Mexican Restaurants therefore, eliminating any competition.

Results

- Using New York neighborhood data and Foursquare API we found the neighborhood details.
- Found neighborhood where there is a good Mexican restaurant and get the details of neighborhood using data analysis. Then we got neighborhood with less Mexican restaurants.
- Using K Mean Clustering, cluster those restaurants into groups.

Discussion

Utilizing more features from Foursquare API we could find its ratings and the menus they provide.

Conclusion

Finally, we found best neighborhoods to start a new Mexican restaurant in Manhattan, New York.