IBM DATA SCIENCE PROFESSIONAL CERTIFICATION CAPSTONE COURSERA PROJECT

MARCO SAENZ

Week 4 - Battle of the Neighborhoods

PROBLEM

- A new customer is wanting to invest in a Mexican cuisine restaurant in New York, a Mexican restaurant may prove to be a great investment opportunity if the location is right.
- We will help the customer to find an optimal location area for his restaurant to be able address the opportunity of covering a need in a certain area.
- We need to understand where the Mexican Restaurants and all other restaurants are in New York, then use a clustering algorithm to find similar areas in New York considering demographic data of each borough.
    - Investors with real state in New York can use the data analytics results for marketing.

APPROACH

- Data description and use:
  - We will use 2 datasets to merge into one dataset for this exercise and analysis.
    - New York Neighborhoods and Boroughs
    - All related locations for Mexican restaurants will be obtained via de foursquare API

METHODOLOGY

1. Import the dataset with New York boroughs and use Data Wrangling to prepare data into a data frame.
2. Create Map for Manhattan to visualize the neighborhoods
3. Use the Foursquare API to request venues data for Manhattan and map the results on top of the created map.
4. Using One hot encoding to analyze each neighborhood. Group by Neighborhood and then use the mean of the frequency of occurrence of each category.
5. Use K-Means Clustering to cluster the neighborhoods into clusters with the most frequent (“Common”) type of venues.
6. Conclude the best location area for the restaurant

1.- IMPORT THE DATASET WITH NEW YORK BOROUGHS AND USE DATA WRANGLING TO PREPARE DATA INTO A DATA FRAME.

Download and Data Wrangling

In [2]:
```
!wget -q -O 'newyork_data.json' https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json
print('Data downloaded!')
```

```
Data downloaded!
```

In [3]:
```
with open('newyork_data.json') as json_data:
    newyork_data = json.load(json_data)
```

In [4]:
```
neighborhoods_data = newyork_data['features']
neighborhoods_data[0]
```

Out[4]:

```
{'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]}}
```


Transform the Data into pandas Dataframe

In [5]:
```
# define the dataframe columns

column_names = ['Borough', 'Neighborhood', 'Latitude', 'Longitude']

# instantiate the dataframe
neighborhoods = pd.DataFrame(columns=column_names)
neighborhoods
```

Out[5]:

Borough	Neighborhood	Latitude	Longitude

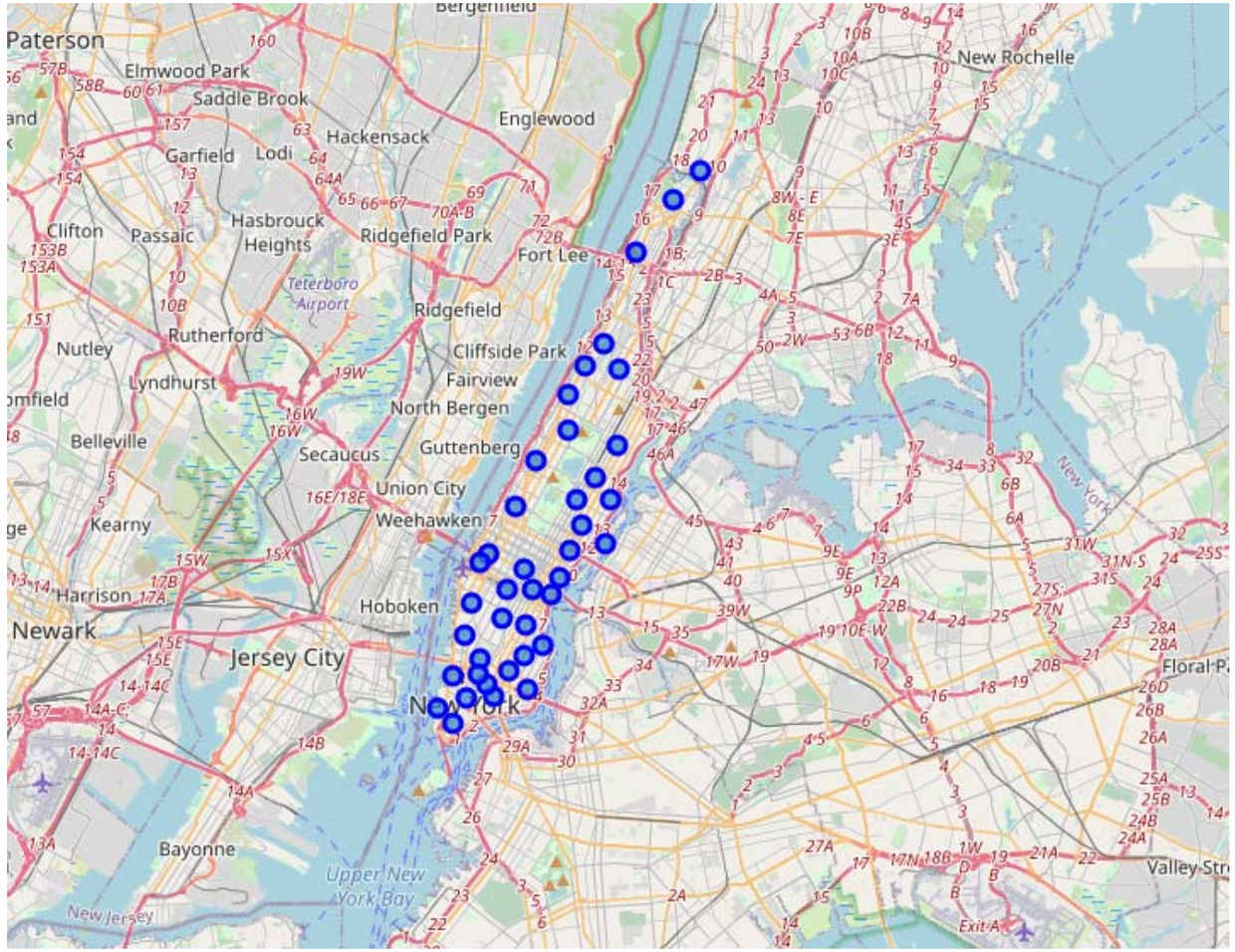
In [6]:
```
for data in neighborhoods_data:
    borough = neighborhood_name = data['properties']['borough']
    neighborhood_name = data['properties']['name']

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

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

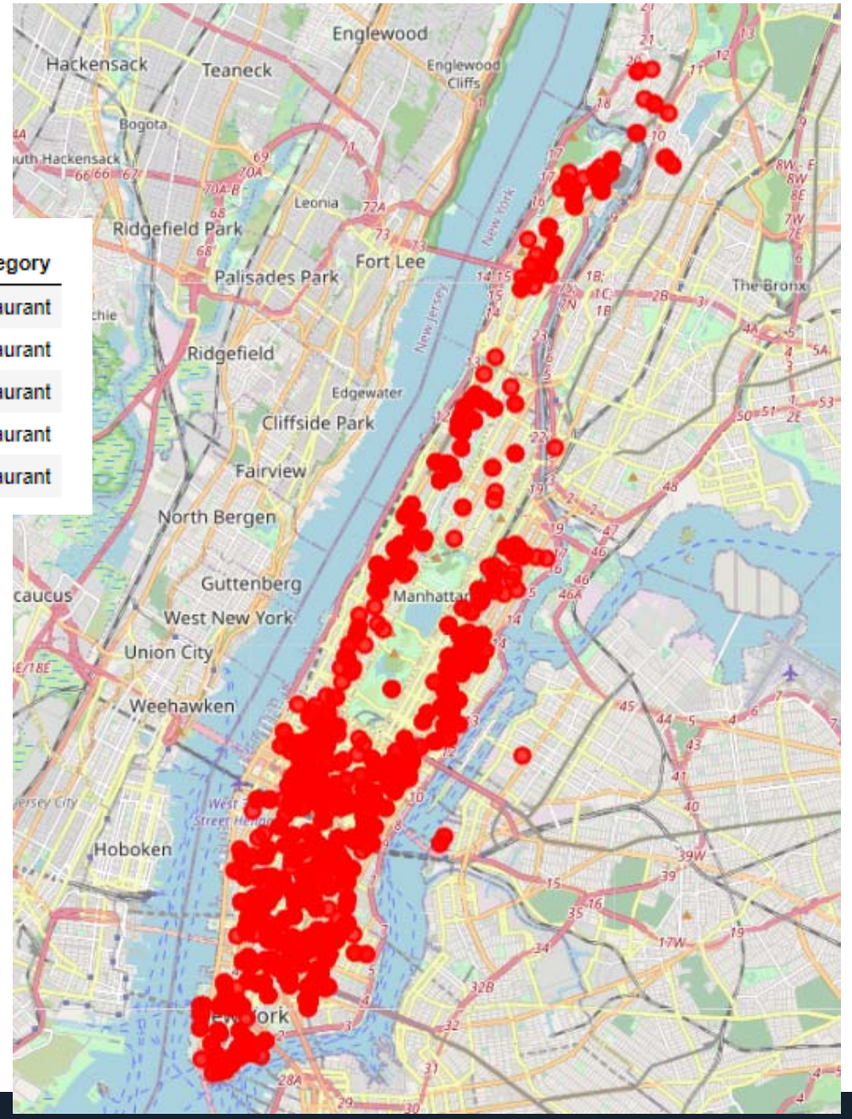
2.- CREATE MAP FOR MANHATTAN TO VISUALIZE THE NEIGHBORHOODS

	Borough	Neighborhood	Latitude	Longitude
0	Manhattan	Marble Hill	40.876551	-73.910660
1	Manhattan	Chinatown	40.715618	-73.994279
2	Manhattan	Washington Heights	40.851903	-73.936900
3	Manhattan	Inwood	40.867684	-73.921210
4	Manhattan	Hamilton Heights	40.823604	-73.949688



3.- USE THE FOURSQUARE API TO REQUEST VENUES DATA FOR MANHATTAN AND MAP THE RESULTS ON TOP OF THE CREATED MAP.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Guacamole	40.874511	-73.910708	Mexican Restaurant
1	Marble Hill	40.876551	-73.91066	Taqueria Sinaloense	40.874574	-73.910687	Mexican Restaurant
2	Marble Hill	40.876551	-73.91066	Mi Lindo San Miguelito	40.880023	-73.906488	Mexican Restaurant
3	Marble Hill	40.876551	-73.91066	Picante Picante Mexican Restaurant	40.878252	-73.902936	Mexican Restaurant
4	Marble Hill	40.876551	-73.91066	Estrellita Poblana V	40.879687	-73.906257	Mexican Restaurant



4.- USING ONE HOT ENCODING TO ANALYZE EACH NEIGHBORHOOD. GROUP BY NEIGHBORHOOD AND THEN USE THE MEAN OF THE FREQUENCY OF OCCURRENCE OF EACH CATEGORY.

	Neighborhood	American Restaurant	Bar	Breakfast Spot	Burger Joint	Burrito Place	Caribbean Restaurant	Chinese Restaurant	Deli / Bodega	Empanada Restaurant	Fast Food Restaurant	Food Stand	Food Truck	Latin American Restaurant	Mexican Restaurant	Seafood Restaurant	Taco Place	Tex-Mex Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant
0	Battery Park City	0.00	0.020833	0.000000	0.000000	0.104167	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.020833	0.000000	0.666667	0.000000	0.187500	0.000000	0.000000	0.000000
1	Carnegie Hill	0.00	0.000000	0.000000	0.023810	0.023810	0.00	0.023810	0.000000	0.00	0.023810	0.00	0.023810	0.000000	0.714286	0.000000	0.166667	0.000000	0.000000	0.000000
2	Central Harlem	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.037037	0.888889	0.000000	0.037037	0.037037	0.000000	0.000000
3	Chelsea	0.00	0.000000	0.020833	0.000000	0.041667	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.041667	0.020833	0.562500	0.020833	0.250000	0.041667	0.000000	0.000000
4	Chinatown	0.00	0.020408	0.000000	0.020408	0.061224	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.673469	0.000000	0.204082	0.000000	0.000000	0.020408
5	Civic Center	0.00	0.020833	0.000000	0.020833	0.104167	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.625000	0.000000	0.208333	0.000000	0.000000	0.020833
6	Clinton	0.00	0.000000	0.000000	0.000000	0.061224	0.00	0.000000	0.000000	0.00	0.020408	0.00	0.040816	0.020408	0.653061	0.000000	0.183673	0.020408	0.000000	0.000000
7	East Harlem	0.00	0.000000	0.000000	0.000000	0.021277	0.00	0.021277	0.000000	0.00	0.021277	0.00	0.042553	0.000000	0.723404	0.000000	0.170213	0.000000	0.000000	0.000000
8	East Village	0.00	0.020000	0.000000	0.000000	0.040000	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.740000	0.000000	0.180000	0.020000	0.000000	0.000000
9	Financial District	0.00	0.020833	0.000000	0.000000	0.083333	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.020833	0.000000	0.687500	0.000000	0.187500	0.000000	0.000000	0.000000
10	Flatiron	0.00	0.000000	0.000000	0.000000	0.061224	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.020408	0.000000	0.693878	0.000000	0.183673	0.040816	0.000000	0.000000
11	Gramercy	0.00	0.000000	0.000000	0.000000	0.061224	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.020408	0.000000	0.714286	0.000000	0.183673	0.020408	0.000000	0.000000
12	Greenwich Village	0.00	0.000000	0.000000	0.000000	0.060000	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.020000	0.020000	0.680000	0.000000	0.160000	0.020000	0.020000	0.020000
13	Hamilton Heights	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.076923	0.038462	0.807692	0.000000	0.038462	0.038462	0.000000	0.000000
14	Hudson Yards	0.00	0.021277	0.000000	0.000000	0.042553	0.00	0.000000	0.000000	0.00	0.021277	0.00	0.085106	0.021277	0.574468	0.000000	0.212766	0.021277	0.000000	0.000000
15	Inwood	0.00	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.00	0.038462	0.00	0.000000	0.000000	0.807692	0.000000	0.153846	0.000000	0.000000	0.000000
16	Lenox Hill	0.00	0.000000	0.000000	0.000000	0.057143	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.028571	0.000000	0.685714	0.000000	0.200000	0.028571	0.000000	0.000000
17	Lincoln Square	0.00	0.000000	0.000000	0.000000	0.083333	0.00	0.000000	0.020833	0.00	0.000000	0.00	0.041667	0.000000	0.708333	0.000000	0.145833	0.000000	0.000000	0.000000
18	Little Italy	0.00	0.020000	0.000000	0.020000	0.060000	0.00	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.720000	0.000000	0.160000	0.000000	0.000000	0.020000
19	Lower East	0.00	0.020000	0.000000	0.000000	0.020000	0.00	0.000000	0.000000	0.02	0.000000	0.00	0.000000	0.000000	0.720000	0.000000	0.220000	0.000000	0.000000	0.000000

5.- USE K-MEANS CLUSTERING TO CLUSTER THE NEIGHBORHOODS INTO CLUSTERS WITH THE MOST FREQUENT ("COMMON") TYPE OF VENUES.

	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	Mexican Restaurant	Taco Place	Burrito Place	Food Truck	Bar	American Restaurant	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant	Latin American Restaurant
1	Carnegie Hill	Mexican Restaurant	Taco Place	Fast Food Restaurant	Food Truck	Burger Joint	Burrito Place	Chinese Restaurant	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant
2	Central Harlem	Mexican Restaurant	Tex-Mex Restaurant	Taco Place	Latin American Restaurant	American Restaurant	Food Stand	Vegetarian / Vegan Restaurant	Seafood Restaurant	Food Truck	Fast Food Restaurant
3	Chelsea	Mexican Restaurant	Taco Place	Food Truck	Burrito Place	Tex-Mex Restaurant	Breakfast Spot	Seafood Restaurant	Latin American Restaurant	American Restaurant	Vegetarian / Vegan Restaurant
4	Chinatown	Mexican Restaurant	Taco Place	Burrito Place	Vietnamese Restaurant	Burger Joint	Bar	Food Truck	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	1	Mexican Restaurant	Taco Place	American Restaurant	Food Stand	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant	Latin American Restaurant	Food Truck	Fast Food Restaurant
1	Manhattan	Chinatown	40.715618	-73.994279	0	Mexican Restaurant	Taco Place	Burrito Place	Vietnamese Restaurant	Burger Joint	Bar	Food Truck	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	3	Mexican Restaurant	Taco Place	Latin American Restaurant	American Restaurant	Caribbean Restaurant	Food Truck	Food Stand	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	3	Mexican Restaurant	Taco Place	Fast Food Restaurant	Food Stand	Vegetarian / Vegan Restaurant	Tex-Mex Restaurant	Seafood Restaurant	Latin American Restaurant	Food Truck	American Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	1	Mexican Restaurant	Food Truck	Tex-Mex Restaurant	Taco Place	Latin American Restaurant	American Restaurant	Food Stand	Vegetarian / Vegan Restaurant	Seafood Restaurant	Fast Food Restaurant

5.- USE K-MEANS CLUSTERING TO CLUSTER THE NEIGHBORHOODS INTO CLUSTERS WITH THE MOST FREQUENT ("COMMON") TYPE OF VENUES.

Cluster 1

```
In [29]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 0, manhatta
```

Out[29]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
1	Chinatown	Mexican Restaurant	Taco Place	Burrito Place
10	Lenox Hill	Mexican Restaurant	Taco Place	Burrito Place
13	Lincoln Square	Mexican Restaurant	Taco Place	Burrito Place
14	Clinton	Mexican Restaurant	Taco Place	Burrito Place
16	Murray Hill	Mexican Restaurant	Taco Place	Burrito Place
18	Greenwich Village	Mexican Restaurant	Taco Place	Burrito Place
21	Tribeca	Mexican Restaurant	Taco Place	Burrito Place
23	Soho	Mexican Restaurant	Taco Place	Burrito Place
28	Battery Park City	Mexican Restaurant	Taco Place	Burrito Place
29	Financial District	Mexican Restaurant	Taco Place	Burrito Place
32	Civic Center	Mexican Restaurant	Taco Place	Burrito Place
33	Midtown South	Mexican Restaurant	Taco Place	Burrito Place
36	Tudor City	Mexican Restaurant	Taco Place	Burrito Place
38	Flatiron	Mexican Restaurant	Taco Place	Burrito Place

South Area

Cluster 0 in Map = Cluster 1 in Table

5.- USE K-MEANS CLUSTERING TO CLUSTER THE NEIGHBORHOODS INTO CLUSTERS WITH THE MOST FREQUENT ("COMMON") TYPE OF VENUES.

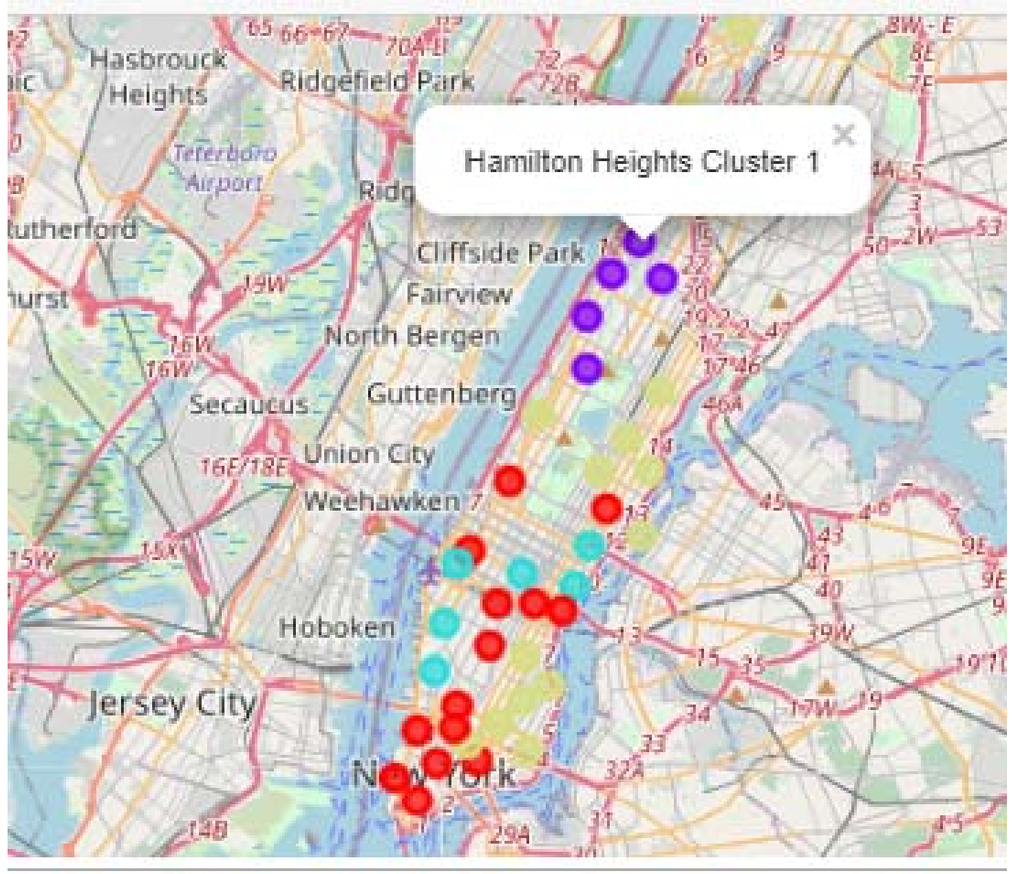
Cluster 2

```
In [30]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 1, manhatta
```

Out[30]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Marble Hill	Mexican Restaurant	Taco Place	American Restaurant
4	Hamilton Heights	Mexican Restaurant	Food Truck	Tex-Mex Restaurant
5	Manhattanville	Mexican Restaurant	Taco Place	American Restaurant
6	Central Harlem	Mexican Restaurant	Tex-Mex Restaurant	Taco Place
25	Manhattan Valley	Mexican Restaurant	Food Truck	Bar
26	Morningside Heights	Mexican Restaurant	Taco Place	Bar

North Area



Cluster 1 in Map = Cluster 2 in Table

5.- USE K-MEANS CLUSTERING TO CLUSTER THE NEIGHBORHOODS INTO CLUSTERS WITH THE MOST FREQUENT ("COMMON") TYPE OF VENUES.

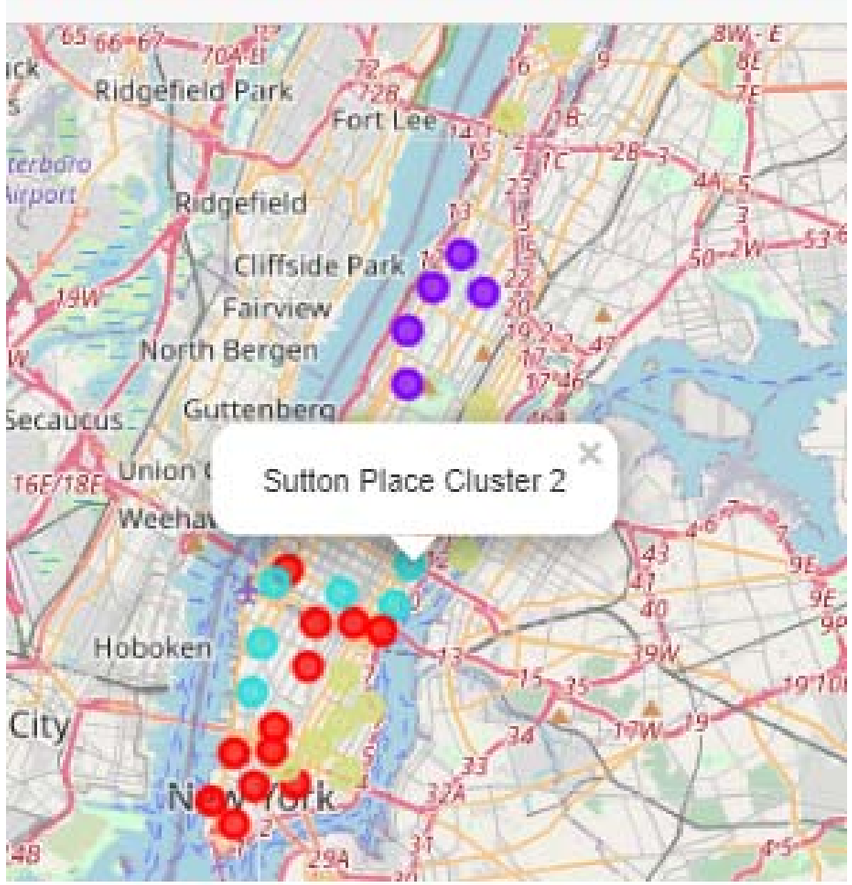
Cluster 3

```
In [31]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 2, manhatta
```

Out[31]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
15	Midtown	Mexican Restaurant	Taco Place	Burrito Place
17	Chelsea	Mexican Restaurant	Taco Place	Food Truck
24	West Village	Mexican Restaurant	Taco Place	Food Truck
34	Sutton Place	Mexican Restaurant	Taco Place	Burrito Place
35	Turtle Bay	Mexican Restaurant	Taco Place	Burrito Place
39	Hudson Yards	Mexican Restaurant	Taco Place	Food Truck

Midtown Area



Cluster 2 in Map = Cluster 3 in Table

5.- USE K-MEANS CLUSTERING TO CLUSTER THE NEIGHBORHOODS INTO CLUSTERS WITH THE MOST FREQUENT ("COMMON") TYPE OF VENUES.

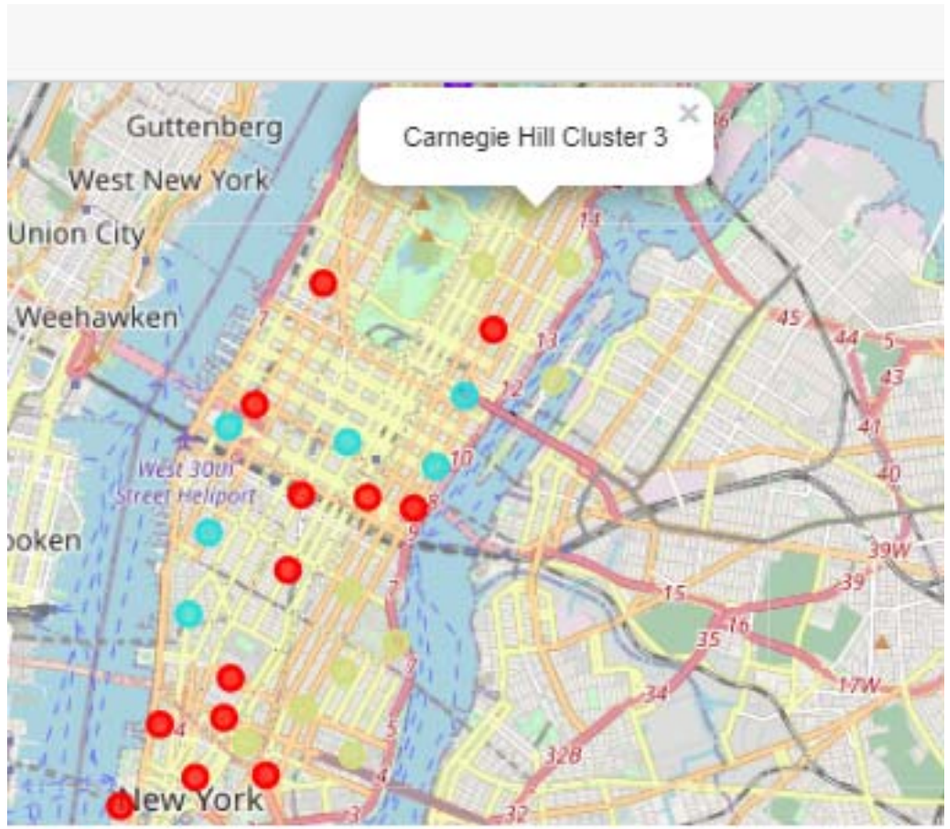
Cluster 4

```
In [32]: manhattan_merged.loc[manhattan_merged['Cluster Labels'] == 3, manhattan_
```

Out[32]:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	
2	Washington Heights	Mexican Restaurant	Taco Place	Latin American Restaurant	A
3	Inwood	Mexican Restaurant	Taco Place	Fast Food Restaurant	
7	East Harlem	Mexican Restaurant	Taco Place	Food Truck	Fa
8	Upper East Side	Mexican Restaurant	Taco Place	Fast Food Restaurant	
9	Yorkville	Mexican Restaurant	Taco Place	Fast Food Restaurant	
11	Roosevelt Island	Mexican Restaurant	Taco Place	Tex-Mex Restaurant	
12	Upper West Side	Mexican Restaurant	Taco Place	Burrito Place	
19	East Village	Mexican Restaurant	Taco Place	Burrito Place	T
20	Lower East Side	Mexican Restaurant	Taco Place	Burrito Place	
22	Little Italy	Mexican Restaurant	Taco Place	Burrito Place	
27	Gramercy	Mexican Restaurant	Taco Place	Burrito Place	T
30	Carnegie Hill	Mexican Restaurant	Taco Place	Fast Food Restaurant	
31	Noho	Mexican Restaurant	Taco Place	Vietnamese Restaurant	
37	Stuyvesant Town	Mexican Restaurant	Taco Place	American Restaurant	

West Area



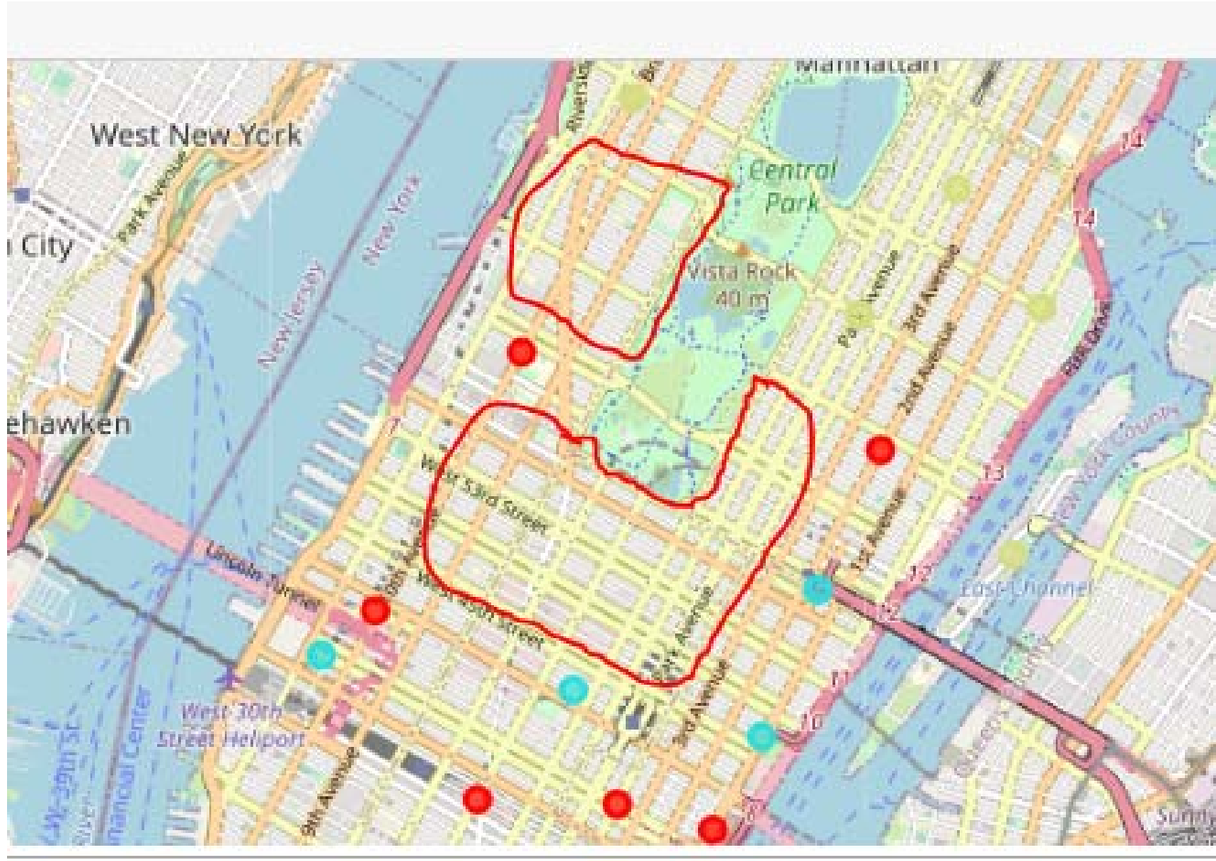
Cluster 2 in Map = Cluster 3 in Table

6.- CONCLUDE THE BEST LOCATION AREA FOR THE RESTAURANT

Based on the Data and locations of Mexican Restaurant Venues clusters we can conclude that the area marked in red are good locations for a Mexican Restaurant.

With the limited data we were able to identify that most Mexican restaurant places are identified as Taco Places, and after reviewing the clusters spread of type of restaurants, we can identify that midtown area is the most beneficial.

1.- The majority of their most common venues are Taco Places, Burrito places and Food Trucks.
2.- The cluster is smaller in proportion to the other neighborhood clusters.



DISCUSSION

Due to the limited data, we can only assume there are stronger factors that we are leaving out in this study.

One of the most important aspects for Data analysis will be the data availability, as it was mentioned during the course. My analysis is based on the available non premium API requests from Foursquare.

There is a High competition market in west Manhattan, and South Manhattan. Most of the places are identified as Taco places, so the full Mexican food restaurant experience would not be an incorrect option even with the competition. With more data regarding population, we may be able to improve this assessment.

It looks like Midtown has only taco places and there is some demand for Mexican food, therefore a Mexican food restaurant would be best located around south Central Park in Midtown.

THANK YOU