Battle of the Neighborhoods:
Project Viva New York
Recommender Simulation using Data Science

1. Business Problem

The United States of America is well known for its large Mexican population along the states that compound the American border. However, this analysis takes place far away from the border reaching Canadian lands. As we move up north, native Spanish speakers (Mexican) population decreases and alongside with it, their cultural background. Mexico is well known for its famous and delicious cuisine. The gastronomy is the ensemble of culinary dishes and techniques from Mexico that takes part in traditions and Mexicans daily life, enriched by its country regions, which in turn are a mix of pre-Hispanic Mexico and European cuisine.

Being New York City one of the largest cities in the United States and in the world, known as a great land of opportunities, makes it a perfect place for a venue idea such as a Mexican restaurant. Even due to its large, populated boroughs such as Manhattan that could create a positive impact on sales, costs can be extremely high especially around downtown and its surrounding areas. Bringing to topic multiculturalism, ethnicity plays a vital role into selecting which type of venue will be a good fit and New York City is not an exception.

New York City Boroughs:

1. **Manhattan**
2. **Bronx**
3. **Brooklyn**
4. **Queens**
5. **Staten Island**

2. Data Overview

The main purpose of this project is to analyze neighborhoods that compound all 5 boroughs from the city of New York and conclude through a series of analyses which areas will be considered, taking into account a series of analyses, the locations that will best fit the conditions for a Mexican restaurant.

The data used on this project is based alone on New York City. Through Foursquare, data such as venues, venues categories and coordinates will be used alongside neighborhoods latitudes and longitudes coordinates to create a merged dataset that will allow the audience to comprehend throughout geospatial data the geography of New York City, population density, and number of local venues in each neighborhood and borough. Machine learning clustering analyses allows to create data visualization of the city neighborhoods and venues dataset and to conclude which areas could be an excellent fit for a Mexican restaurant based on venues frequency and population.

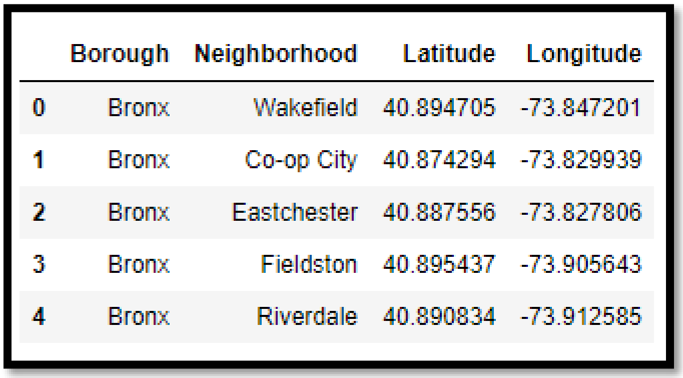
3. Target Audience

The project aims Entrepreneurs or Business owners who want to expand their Mexican restaurant franchises or grow a new business outside their hometown. This analysis will provide helpful geospatial information that could be used by the targeted audience.

4. Data Acquisition

4.1 New York Boroughs and Neighborhoods Dataset

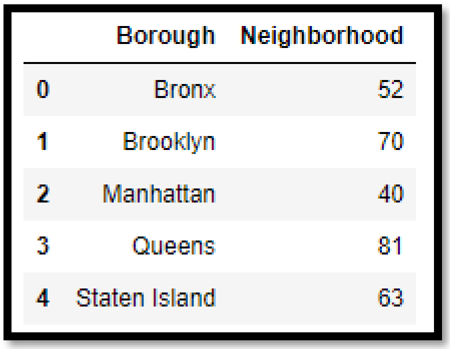
First, the dataset related to Boroughs and Neighborhoods from New York city that displays Latitude and Longitude coordinates was extracted from New York JSON file acquired from IBM Skills Network Labs repository.



	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

Fig 4.1 New York Boroughs and Neighborhoods dataset Pandas Dataframe 'newyork_n'

The dataset compounds all different Boroughs from New York City and their respective neighborhoods.



	Borough	Neighborhood
0	Bronx	52
1	Brooklyn	70
2	Manhattan	40
3	Queens	81
4	Staten Island	63

Fig 4.2 New York Count of Neighborhoods group by Boroughs

4.1 New York City Population by Neighborhoods (2010)

Source: https://data.cityofnewyork.us/City-Government/New-York-City-Population-By-Neighborhood-Tabulatio/swpk-hqdp

NYC OpenData

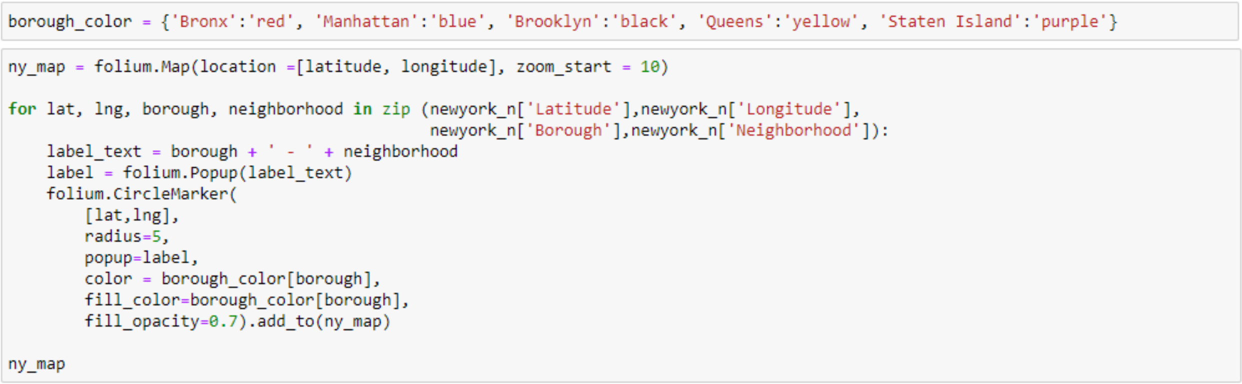
The data was collected from Census Bureaus' Decennial data dissemination (SF1). Neighborhood Tabulation Areas (NTAs) are aggregations of census tracts that are subsets of New York City's 55 Public Use Microdata Areas (PUMAs). Primarily due to these constraints, NTA boundaries and their associated names may not definitively represent neighborhoods.

5. Methodology

From geopy.geocoders library, Nominatim module was imported to acquire specific latitude and longitude coordinates for New York City location in order to center map as shown below.

Geographical Coordinates of New York City: 40.7127281, -74.0060152

After acquiring all related data from above mentioned sources, the next step is to analyze these sources. This script displays a map with coded colors according to dataset **“newyork_n”** Boroughs column that codes with a specified color all neighborhoods clustered in each borough.



```
borough_color = {'Bronx':'red', 'Manhattan':'blue', 'Brooklyn':'black', 'Queens':'yellow', 'Staten Island':'purple'}
```

```
ny_map = folium.Map(location =[latitude, longitude], zoom_start = 10)

for lat, lng, borough, neighborhood in zip (newyork_n['Latitude'],newyork_n['Longitude'],
                                           newyork_n['Borough'],newyork_n['Neighborhood']):
    label_text = borough + ' - ' + neighborhood
    label = folium.Popup(label_text)
    folium.CircleMarker(
        [lat,lng],
        radius=5,
        popup=label,
        color = borough_color[borough],
        fill_color=borough_color[borough],
        fill_opacity=0.7).add_to(ny_map)

ny_map
```

For the map shown below, using Folium module library a code is assigned to each neighborhood depending on which Borough it is located.

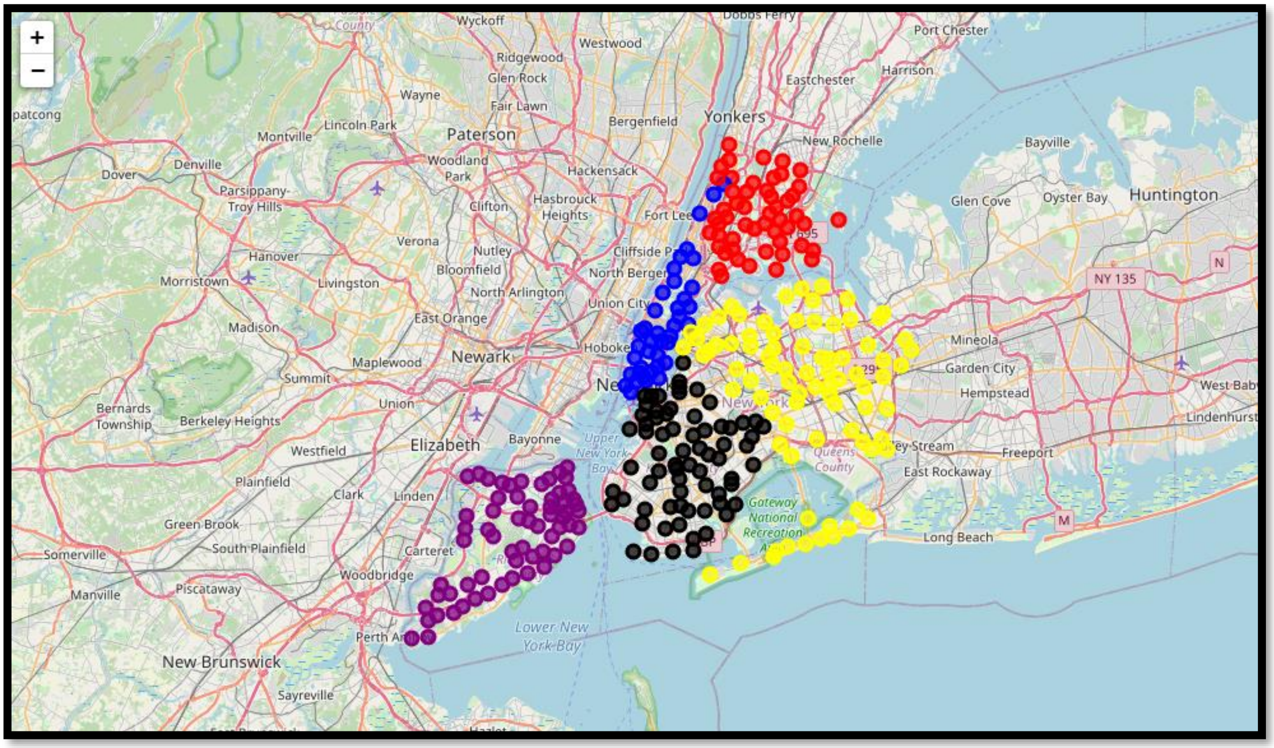
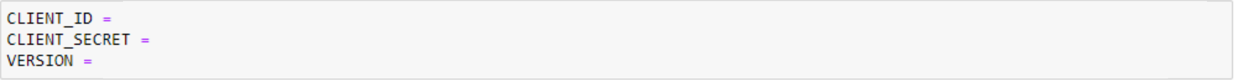


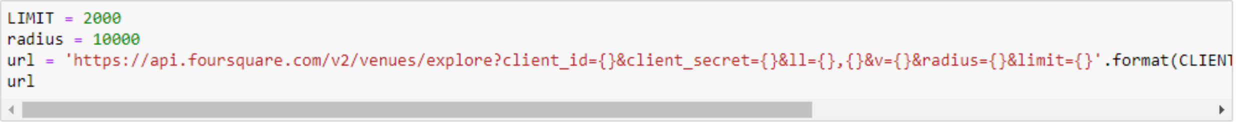
Fig 5.1 Boroughs (districts) of New York City (“ny_map”)

Staten Island	**Brooklyn**	**Manhattan**	**Bronx**	**Queens**

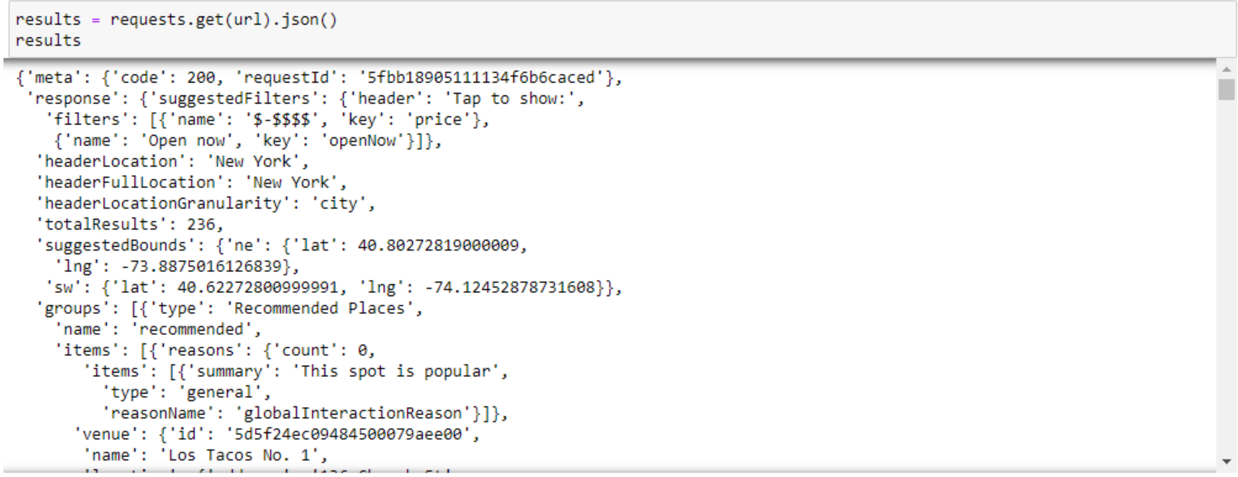
After creating New York borough clustering map, we proceed using Foursquare API calls to retrieve limited number of venues within a specified radius.



```
CLIENT_ID =
CLIENT_SECRET =
VERSION =
```



```
LIMIT = 2000
radius = 10000
url = 'https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&ll={},{}&v={}&radius={}&limit={}'.format(CLIENT
url
```

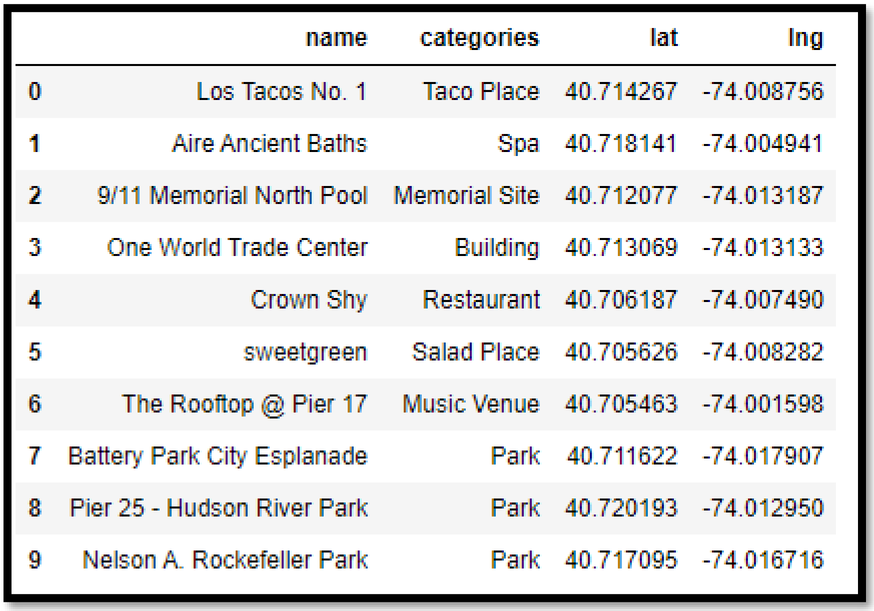


```
results = requests.get(url).json()
results
```

```
{'meta': {'code': 200, 'requestId': '5fbb18905111134f6b6caced'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': '$-$$$$', 'key': 'price'},
    {'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'New York',
  'headerFullLocation': 'New York',
  'headerLocationGranularity': 'city',
  'totalResults': 236,
  'suggestedBounds': {'ne': {'lat': 40.80272819000009,
    'lng': -73.8875016126839},
   'sw': {'lat': 40.62272800999991, 'lng': -74.12452878731608}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '5d5f24ec09484500079aee00',
       'name': 'Los Tacos No. 1',
```

A function get_category_type is defined and will retrieve name, categories, latitude and longitudes coordinates from the data extracted from api.foursquare.com url.

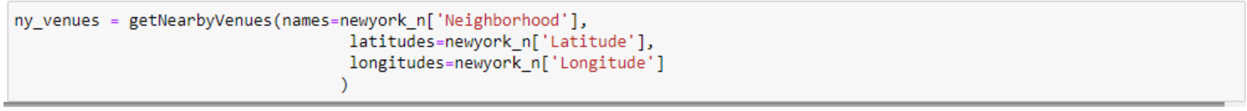
From the metadata extracted, all information was retrieved and structured into a data frame. This data frame named "newyork_nearby_venues" shows name, unique categories, and geographical coordinates from all venues within a specified geographical range.



	name	categories	lat	lng
0	Los Tacos No. 1	Taco Place	40.714267	-74.008756
1	Aire Ancient Baths	Spa	40.718141	-74.004941
2	9/11 Memorial North Pool	Memorial Site	40.712077	-74.013187
3	One World Trade Center	Building	40.713069	-74.013133
4	Crown Shy	Restaurant	40.706187	-74.007490
5	sweetgreen	Salad Place	40.705626	-74.008282
6	The Rooftop @ Pier 17	Music Venue	40.705463	-74.001598
7	Battery Park City Esplanade	Park	40.711622	-74.017907
8	Pier 25 - Hudson River Park	Park	40.720193	-74.012950
9	Nelson A. Rockefeller Park	Park	40.717095	-74.016716

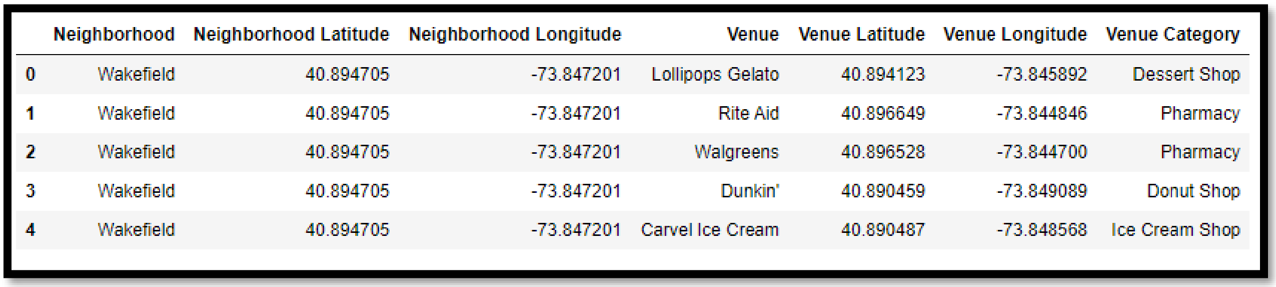
Fig 5.2 Dataframe "newyork_nearby_venues"

Once again, we defined another function called getNearbyVenues that will retrieve a specified number of venues from the same specified geographical range as mentioned above and will convert it into a data frame. This defined function utilizes a Foursquare API call.



```
ny_venues = getNearbyVenues(names=newyork_n['Neighborhood'],
                                   latitudes=newyork_n['Latitude'],
                                   longitudes=newyork_n['Longitude']
                                  )
```

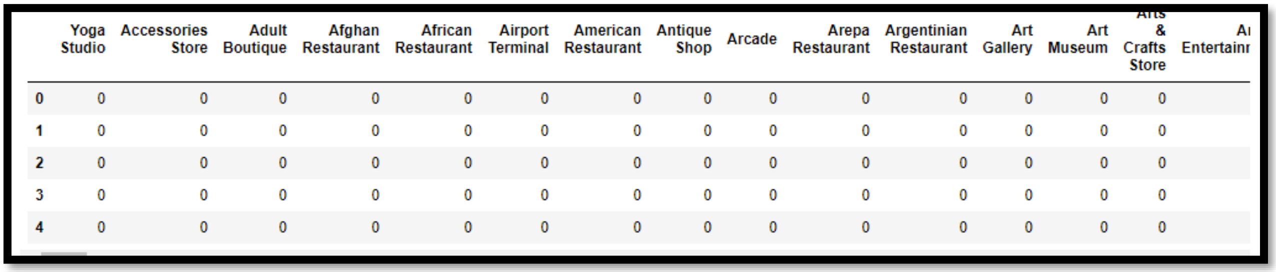
The data frame will be called “ny_venues” and displays neighborhood names, respective latitudes and longitudes and all venues that fall into the same area within each specified neighborhood as shown in the following data frame:



	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy
2	Wakefield	40.894705	-73.847201	Walgreens	40.896528	-73.844700	Pharmacy
3	Wakefield	40.894705	-73.847201	Dunkin'	40.890459	-73.849089	Donut Shop
4	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop

Fig 5.3 Dataframe “ny_venues”

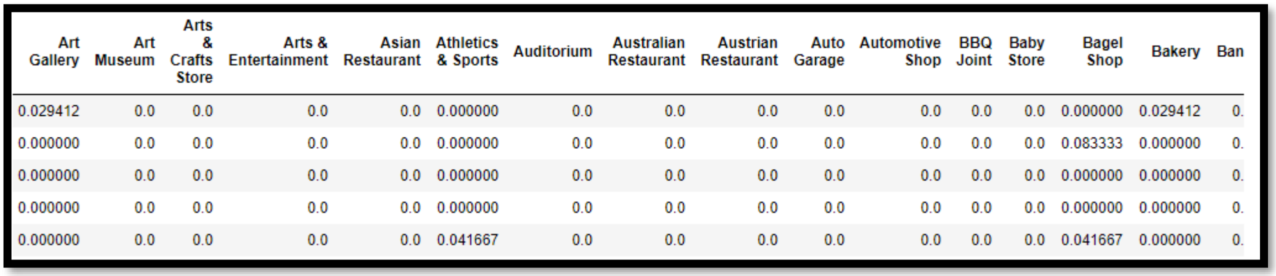
For categorical values, the integer encoding will allow the model to assume a natural order between categories and may result in poor performance or unexpected results. We use a technique called one-hot encoding in which the integer encoded variable is removed and a new binary variable is added for each unique integer. The coding 1 and 0 are assigned to each variable, if a variable exists in each statement a 1 is assigned, if not, the number 0 will appear.



	Yoga Studio	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Terminal	American Restaurant	Antique Shop	Arcade	Arepa Restaurant	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Ai Entertainr
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Fig 5.4 Dataframe “ny_onehot”
One-hot encoding technique

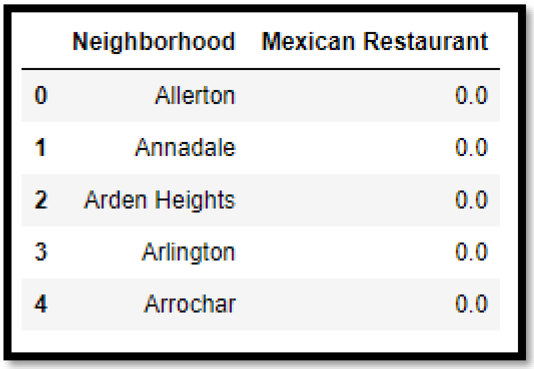
Subsequently, the mean of the frequency of occurrence of each Venue Category is calculated. All rows are grouped by Neighborhoods. Then, Neighborhood and Mexican Restaurant columns are extracted into a new data frame that will be used later for its respective analysis.



Art Gallery	Art Museum	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	Auditorium	Australian Restaurant	Austrian Restaurant	Auto Garage	Automotive Shop	BBQ Joint	Baby Store	Bagel Shop	Bakery	Ban
0.029412	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.029412	0.
0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.083333	0.000000	0.
0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.
0.000000	0.0	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	0.000000	0.
0.000000	0.0	0.0	0.0	0.0	0.041667	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.041667	0.000000	0.

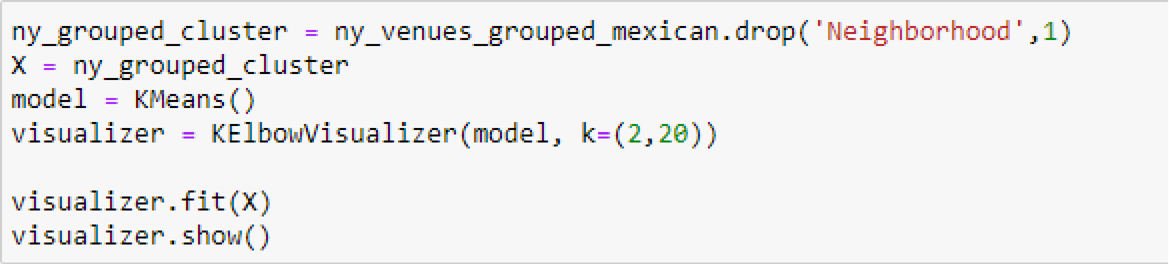
Fig 5.5 Dataframe "ny_venues_grouped"
Calculated mean on One-hot encoding technique

The Mexican Restaurant column on this data frame is removed to apply Machine Learning K-Means clustering, for it cannot function with categorical variables, in order to cluster all neighborhoods based on the same neighborhoods that had similar averages of Mexican Restaurants in that neighborhood. Inside K-Means clustering function an optimum k value must be assigned in order to classify neighborhoods by cluster labels. We imported "KElbowVisualizer" to fit our K-Means model using the optimum k value.



	Neighborhood	Mexican Restaurant
0	Allerton	0.0
1	Annadale	0.0
2	Arden Heights	0.0
3	Arlington	0.0
4	Arrochar	0.0

Fig 5.6 Dataframe "ny_venues_grouped_mexican"



```
ny_grouped_cluster = ny_venues_grouped_mexican.drop('Neighborhood',1)
X = ny_grouped_cluster
model = KMeans()
visualizer = KElbowVisualizer(model, k=(2,20))

visualizer.fit(X)
visualizer.show()
```

We assigned data frame "ny_venues_grouped_mexican" as our *X* value and model as K-Means. The visualizer runs the model given these two variables and finally X is fitted into the visualizer model.

This Elbow Point visualizer display the graph below:

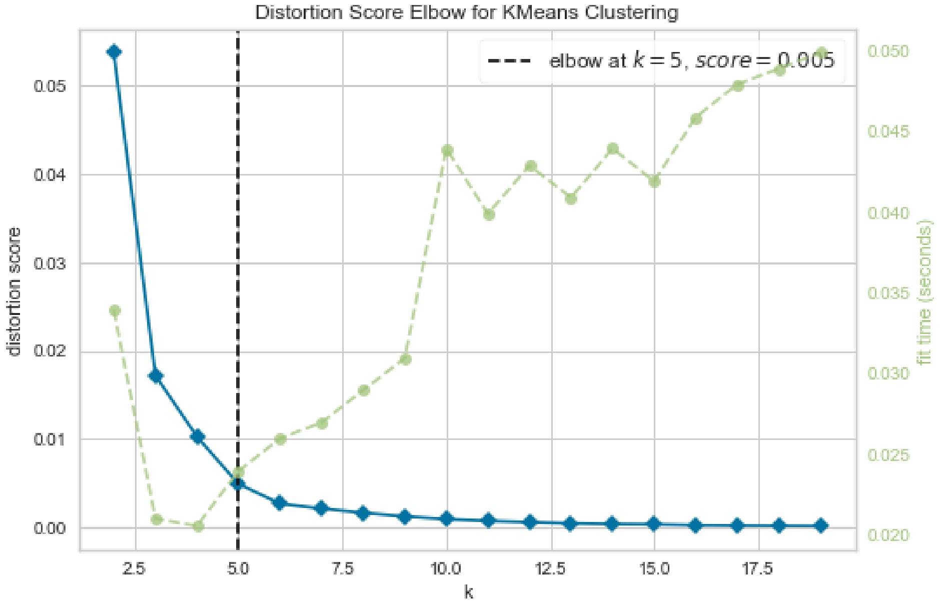
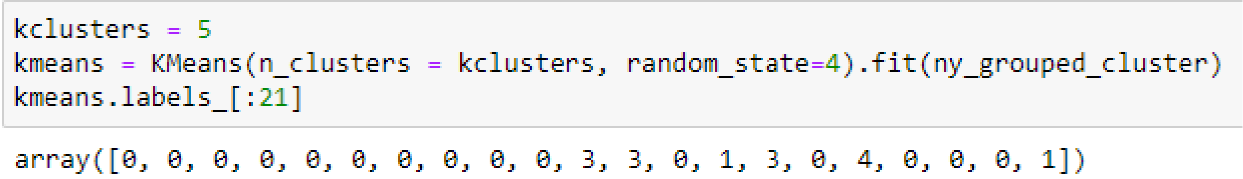


Fig 5.7 KElbowVisualizer graph

The graph above shows that the optimum K-value for K-Means clustering technique is when k = 5. As you can see, the KElbowVisualizer function runs several tests for different k-values and measures the accuracy of each given k-value. The optimum K-value is chosen at the point where the line has a sharpest turn k = 5.

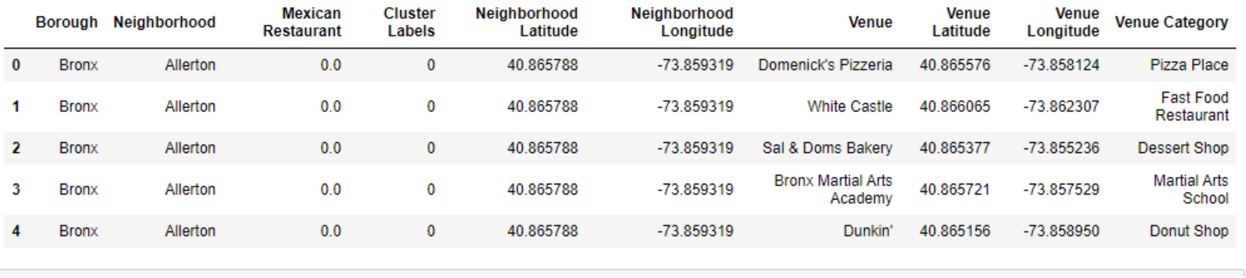
Now that optimum k-value is known, we proceed to cluster all neighborhoods into 5 different clusters using KMeans technique.



```
kclusters = 5
kmeans = KMeans(n_clusters = kclusters, random_state=4).fit(ny_grouped_cluster)
kmeans.labels_[:21]
```

```
array([0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 3, 0, 1, 3, 0, 4, 0, 0, 0, 1])
```

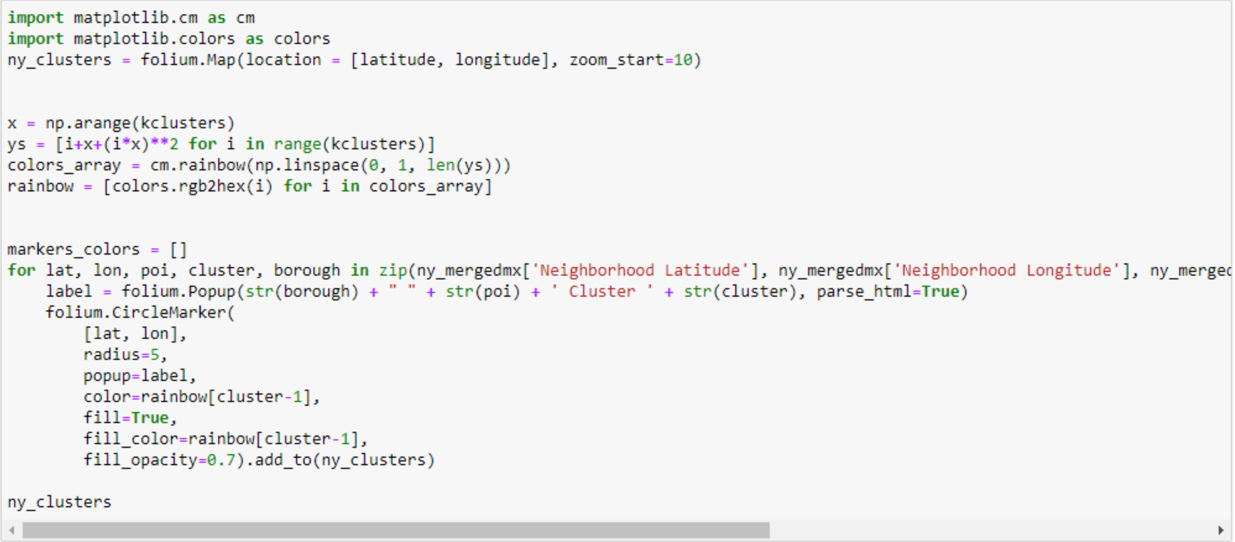
As seen on the previous image, clusters are assigned starting from 0, not 1. Hence, having a k value of 5, neighborhoods will be ranked from 0 up to 4.



	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	Allerton	0.0	0	40.865788	-73.859319	Domenick's Pizzeria	40.865576	-73.858124	Pizza Place
1	Bronx	Allerton	0.0	0	40.865788	-73.859319	White Castle	40.866065	-73.862307	Fast Food Restaurant
2	Bronx	Allerton	0.0	0	40.865788	-73.859319	Sal & Doms Bakery	40.865377	-73.855236	Dessert Shop
3	Bronx	Allerton	0.0	0	40.865788	-73.859319	Bronx Martial Arts Academy	40.865721	-73.857529	Martial Arts School
4	Bronx	Allerton	0.0	0	40.865788	-73.859319	Dunkin'	40.865156	-73.858950	Donut Shop

Fig 5.8 Data Frame "ny_mergedmx"

After clustering all neighborhoods, data frames are merged to compile into a single data frame boroughs, neighborhoods, venues, clusters and latitudes and longitudes coordinates for each respective neighborhood and venue.



```
import matplotlib.cm as cm
import matplotlib.colors as colors
ny_clusters = folium.Map(location = [latitude, longitude], zoom_start=10)


x = np.arange(kclusters)
ys = [i+x+(i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]


markers_colors = []
for lat, lon, poi, cluster, borough in zip(ny_mergedmx['Neighborhood Latitude'], ny_mergedmx['Neighborhood Longitude'], ny_merged
    label = folium.Popup(str(borough) + " " + str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(ny_clusters)

ny_clusters
```

For the map shown below, we can say there is intraclustering between clusters. The red dots are Cluster #0 and compounds the majority of New York City area owning about 6,242 venues and 220 neighborhoods.

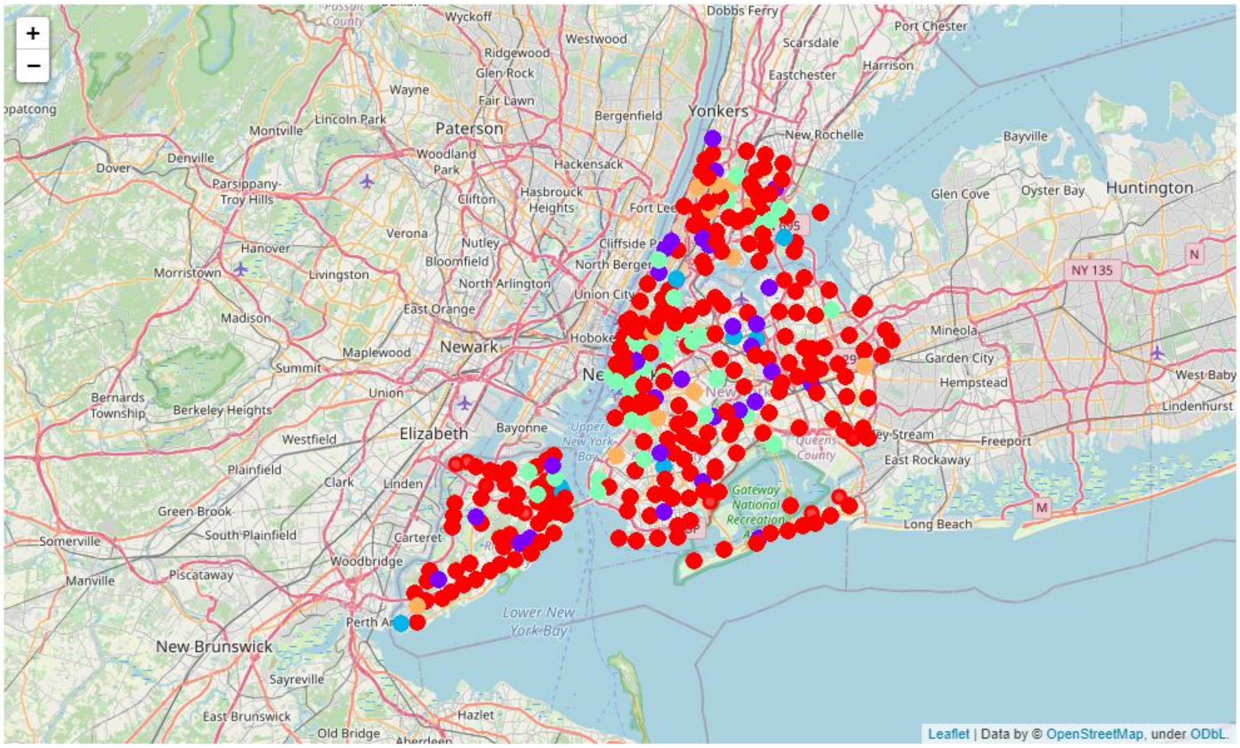
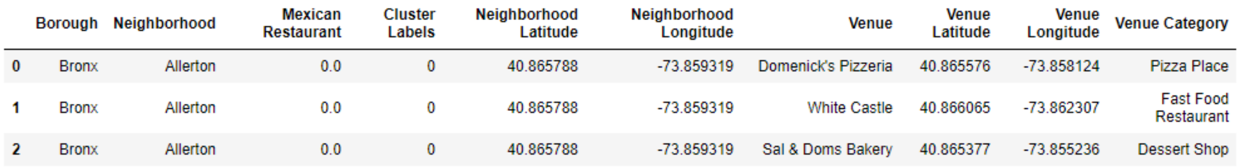


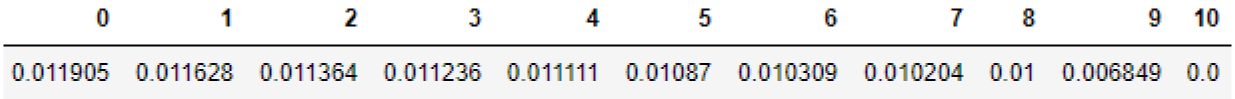
Fig 5.9 New York City Map
Neighborhoods grouped into 5 different clusters

Following up, we will be discussing number of venues and neighborhoods per cluster shown in map. The following analyses was made using one-hot encoding technique to analyze the means of the frequency of venues, in this case, Mexican Restaurants given by foursquare API call used previously.

Cluster 0



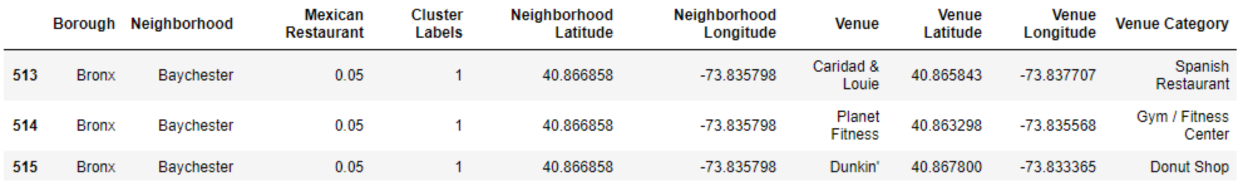
	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Bronx	Allerton	0.0	0	40.865788	-73.859319	Domenick's Pizzeria	40.865576	-73.858124	Pizza Place
1	Bronx	Allerton	0.0	0	40.865788	-73.859319	White Castle	40.866065	-73.862307	Fast Food Restaurant
2	Bronx	Allerton	0.0	0	40.865788	-73.859319	Sal & Doms Bakery	40.865377	-73.855236	Dessert Shop



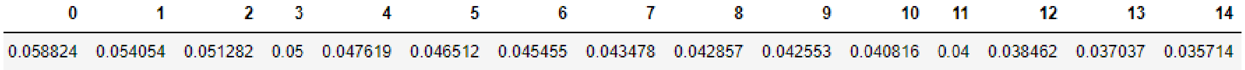
0	1	2	3	4	5	6	7	8	9	10
0.011905	0.011628	0.011364	0.011236	0.011111	0.01087	0.010309	0.010204	0.01	0.006849	0.0

Mexican Restaurant means on Cluster 0 range from 0 to 0.011905 with 220 Neighborhoods and 6,242 Venues, 71.9% and 59.9% of total Neighborhoods and Venues, respectively. We can conclude that cluster 0 has on average 28 Venues per Neighborhood with the smallest range of means according to One-hot Encoding technique.

Cluster 1



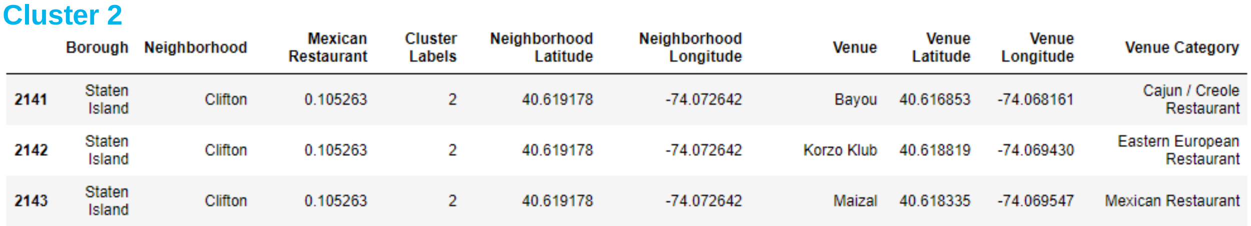
	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
513	Bronx	Baychester	0.05	1	40.866858	-73.835798	Caridad & Louie	40.865843	-73.837707	Spanish Restaurant
514	Bronx	Baychester	0.05	1	40.866858	-73.835798	Planet Fitness	40.863298	-73.835568	Gym / Fitness Center
515	Bronx	Baychester	0.05	1	40.866858	-73.835798	Dunkin'	40.867800	-73.833365	Donut Shop



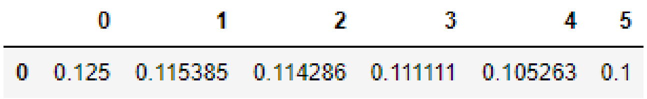
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0.058824	0.054054	0.051282	0.05	0.047619	0.046512	0.045455	0.043478	0.042857	0.042553	0.040816	0.04	0.038462	0.037037	0.035714

Mexican Restaurants means on Cluster 1 range from 0.035714 to 0.058824 with 30 Neighborhoods and 1,240 Venues, 9.8% and 11.9% of total Neighborhoods and Venues, respectively. Cluster 1 has on average 41 Venues per Neighborhood ranking 3rd on means according to One-hot Encoding technique.

Cluster 2



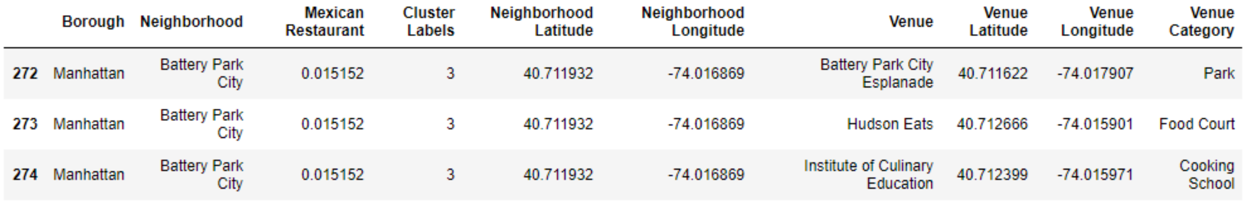
	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2141	Staten Island	Clifton	0.105263	2	40.619178	-74.072642	Bayou	40.616853	-74.068161	Cajun / Creole Restaurant
2142	Staten Island	Clifton	0.105263	2	40.619178	-74.072642	Korzo Klub	40.618819	-74.069430	Eastern European Restaurant
2143	Staten Island	Clifton	0.105263	2	40.619178	-74.072642	Maizal	40.618335	-74.069547	Mexican Restaurant



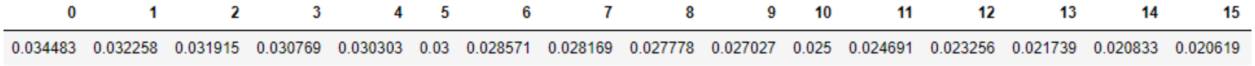
	0	1	2	3	4	5
0	0.125	0.115385	0.114286	0.111111	0.105263	0.1

Mexican Restaurants means on Cluster 2 range from 0.1 to 0.125 with 8 Neighborhoods and 184 Venues, 2.6% and 1.8% of total Neighborhoods and Venues, respectively. Cluster 2 has on average 23 Venues per Neighborhood and it has the highest means according to One-hot Encoding technique.

Cluster 3



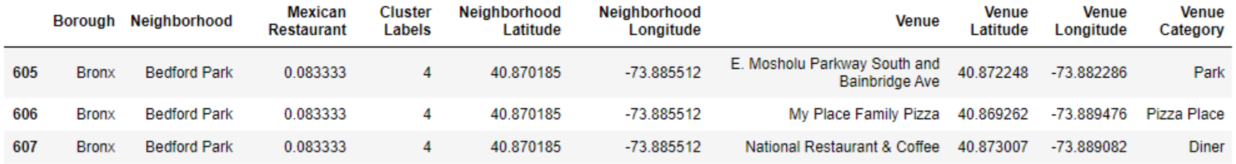
	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
272	Manhattan	Battery Park City	0.015152	3	40.711932	-74.016869	Battery Park City Esplanade	40.711622	-74.017907	Park
273	Manhattan	Battery Park City	0.015152	3	40.711932	-74.016869	Hudson Eats	40.712666	-74.015901	Food Court
274	Manhattan	Battery Park City	0.015152	3	40.711932	-74.016869	Institute of Culinary Education	40.712399	-74.015971	Cooking School



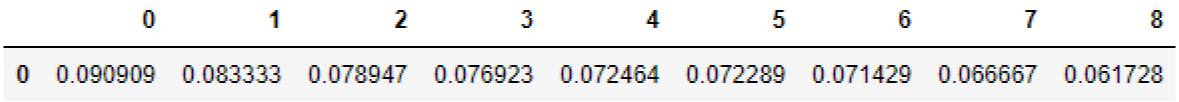
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.034483	0.032258	0.031915	0.030769	0.030303	0.03	0.028571	0.028169	0.027778	0.027027	0.025	0.024691	0.023256	0.021739	0.020833	0.020619

Mexican Restaurants means on Cluster 3 range from 0.014286 to 0.034483 with 36 Neighborhoods and 2,292 Venues, 11.8% and 21.9% of total Neighborhoods and Venues, respectively. Cluster 3 has the highest average of Venues per Neighborhood with 64 Venues and ranking the second to lowest on means according to One-hot Encoding technique.

Cluster 4

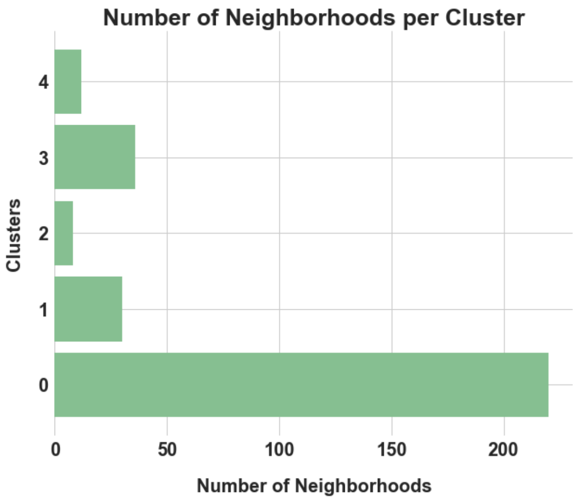


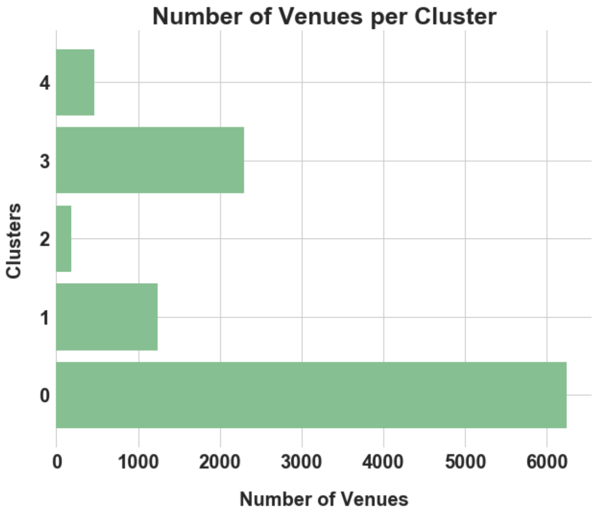
	Borough	Neighborhood	Mexican Restaurant	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
605	Bronx	Bedford Park	0.083333	4	40.870185	-73.885512	E. Mosholu Parkway South and Bainbridge Ave	40.872248	-73.882286	Park
606	Bronx	Bedford Park	0.083333	4	40.870185	-73.885512	My Place Family Pizza	40.869262	-73.889476	Pizza Place
607	Bronx	Bedford Park	0.083333	4	40.870185	-73.885512	National Restaurant & Coffee	40.873007	-73.889082	Diner



	0	1	2	3	4	5	6	7	8
0	0.090909	0.083333	0.078947	0.076923	0.072464	0.072289	0.071429	0.066667	0.061728

Mexican Restaurants means on Cluster 4 range from 0.061728 to 0.090909 with 12 Neighborhoods and 463 Venues, 3.9% and 4.4% of total Neighborhoods and Venues, respectively. Cluster 4 has on average 39 Venues per Neighborhood and ranking the second to highest means according to One-hot Encoding technique.





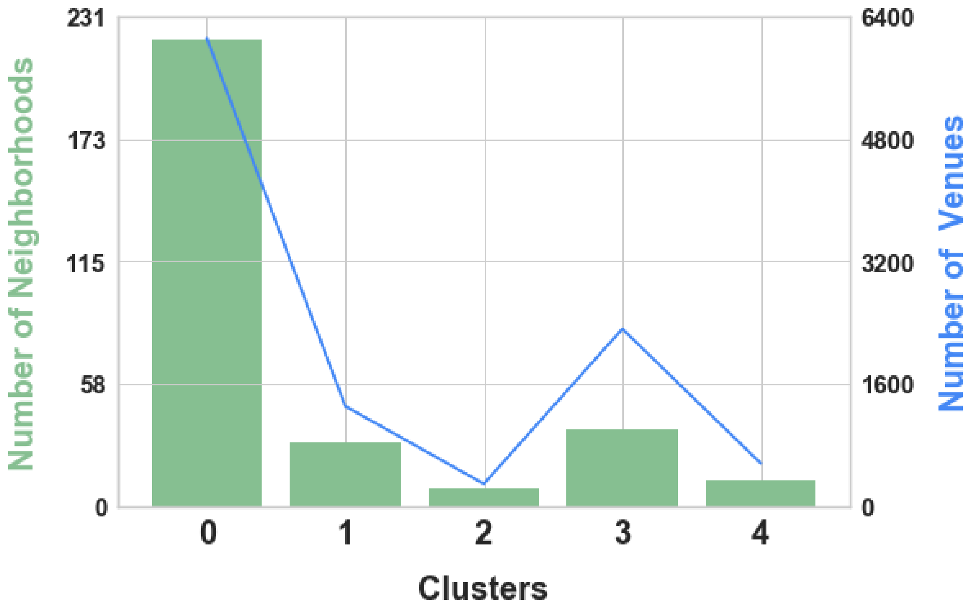


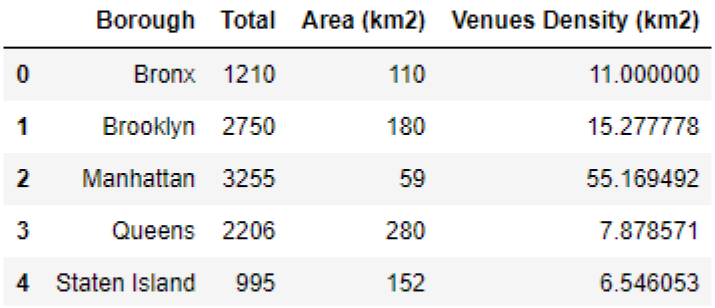
Fig 5.10 Number of Neighborhoods and Venues per Cluster

Cluster Analysis

Therefore, the ranking of Mexican restaurants by cluster will be ordered as follows from highest to lowest in descending order:

1. Cluster 2 (0.1 – 0.125)
2. Cluster 4 (0.061728 – 0.090909)
3. Cluster 1 (0.35714 – 0.058824)
4. Cluster 3 (0.014286 – 0.034483)
5. Cluster 0 (0 – 0.011905)

We cannot reach a conclusion yet by focusing only on clusters. The larger the observations are in a single cluster the smaller the average number of a given venue category, in this project, Mexican restaurants.



	Borough	Total	Area (km2)	Venues Density (km2)
0	Bronx	1210	110	11.000000
1	Brooklyn	2750	180	15.277778
2	Manhattan	3255	59	55.169492
3	Queens	2206	280	7.878571
4	Staten Island	995	152	6.546053

Nevertheless, more analyses can be made along with machine learning clustering especially when we are working with geospatial data. On the data frame above we can see that Manhattan has an overwhelming number of Venues despite the area it covers. We will follow up with choropleth maps regarding New York City Population Density, Venues per Neighborhood and Population and Venues Ratio across the area, all of them grouped by Neighborhoods.

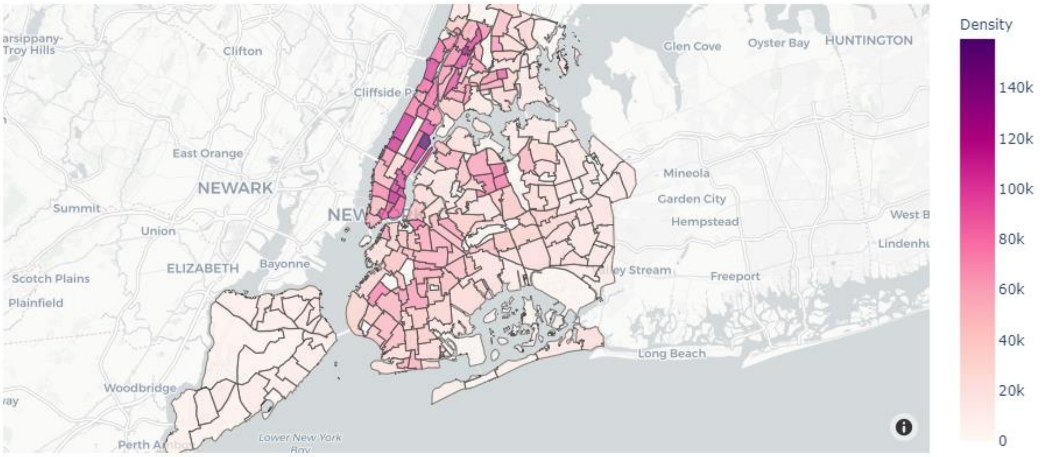


Fig 5.11 New York City population density

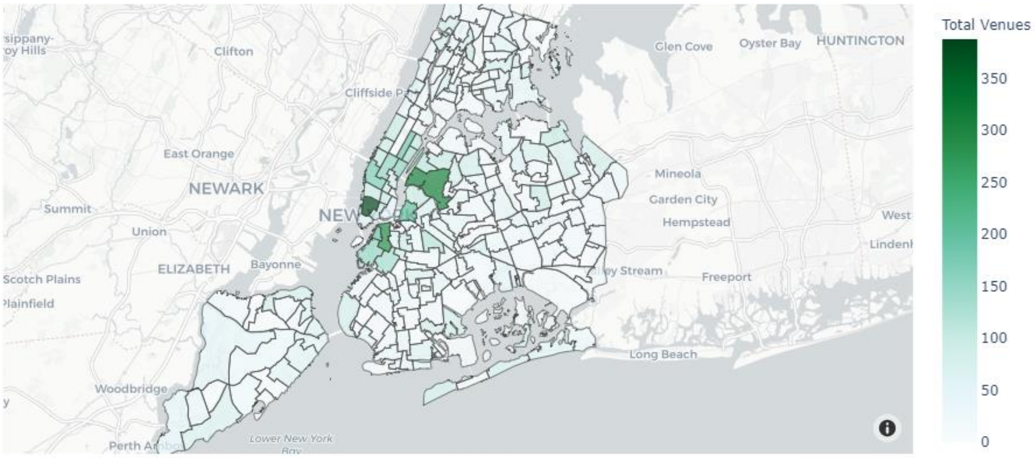


Fig 5.12 Total Venues per Neighborhood in New York Citv

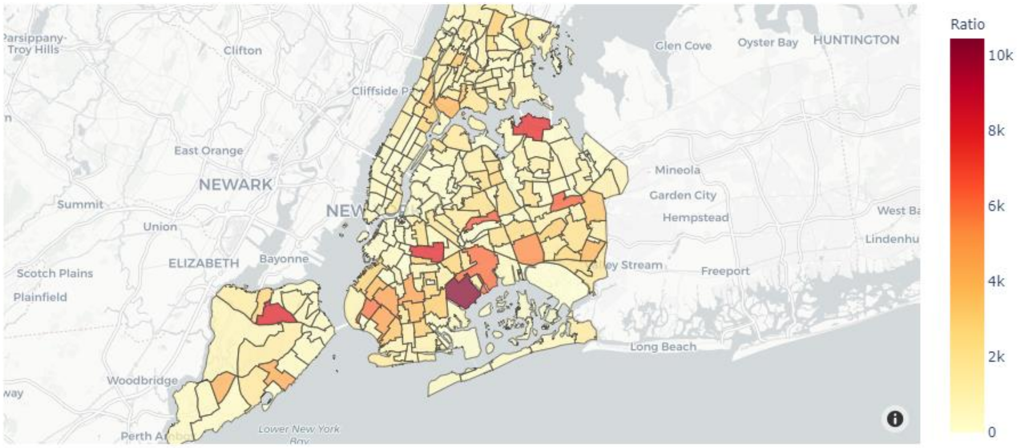


Fig 5.13 Population – Venues Ratio

6. Discussion

In Fig. 5.11, we clearly see that the northern side of New York City formed by Manhattan area has neighborhoods with the largest population density such as Yorkville, East Village, Upper West Side and Lincoln Square. The Bronx area has also considerably a large population density with neighborhoods such as Fordham South and Mount Hope. Staten Island and Queens have the least population density almost entirely in all their neighborhoods.

Now as for Total Venues per Neighborhood, it is noticeably clear that the area towards Downtown Manhattan has the highest number of Venues per neighborhood as stated in Fig 5.12. Neighborhoods such as Soho, Tribeca, Civic Center, Little Italy, Hudson Yards, Chelsea, Flat Iron, Union Square and much more can be found inside Manhattan.

Nonetheless, it is not only Manhattan area that ensues this principle. The northern side of Brooklyn with neighborhoods like Hunters Point, Sunnyside, and West Maspeth area, North Side and South Side area, and Dumbo, Vinegar Hill, Downtown Brooklyn and Boerum Hill area all have a significant number of venues.
We can conclude by observing the map that the highest concentration of venues without taking into account the population density, revolves around the area of downtown Manhattan and northern Brooklyn. Being Downtown Manhattan a place with a wide diversity of ethnicity and large population, it can be selected as a good fit for a Mexican restaurant, however, we need to take into account the number of venues found alongside the area.

At last but not at least, we will merge both Population Density and Total Number of Venues information to observe which Neighborhoods share the highest Density – Venues Ratio. Neighborhood areas such as Westerleigh in Staten Island, Beechhurst, Whitestone and Malba area and Jamaica Estates and Holliswood area in Queens, Crown Heights North area, East New York and Canarsie in Brooklyn share a considerable large demand over offer, which means that population surpasses in a substantial way the number of venues located within the mentioned areas.

This also means these areas could have a small number of Venues compared to population density.

Due to the analyses recently made, I will strongly suggest as best fits for a Mexican restaurant location Westerleigh area in Staten Island and Crown Heights North in Brooklyn. Westerleigh neighborhood has a large population density – venue ratio and belongs to Cluster 0 with the lowest frequency of occurrence for Mexican restaurants. The entire has the lowest number of venues compared with the rest and ranks 3rd in square kilometers with 995 venues and 152km2, with that said, Staten Island has the lowest number of venues per square kilometer.

As mentioned before, there is a strong magnet of many venues towards Downtown Manhattan area. Following this premise, Crown Heights North is in a location that could follow this pattern. Nevertheless, it shows a not considerable number of venues and this could lead to think that it is not consider a great location for building a restaurant or any venue of any category but, on the other hand, it can be considered a good fit because of having no competition whatsoever and having a significant population – venue ratio.

7. Conclusion

In conclusion, to summarize all done on this project, we had the opportunity to attack a business problem, acquire all relevant information from websites using python libraries and machine learning, create data frames from scratch and come up with complex choropleth maps for analyses throughout New York City area. Using Foursquare API calls, we were able to retrieve data relevant to the area we were analyzing. In the same way, we put into test machine learning techniques to overcome obstacles found in our way, as so to speak, we were given the challenge to work in a similar way a real data scientist would do.

It is imperative to state that this project is open for improvement as there are several different techniques and libraries that could easily fetch more relevant information or create better understanding data frames or charts, with that mentioned, I really hope this project could help tackle real life problem situations or craft new ideas as to think of what else could be done with this information and techniques used on this project.