

# **Analysis of Neighborhoods of New York City - Report**

## **1. Introduction and business Problem**

As a part of the final IBM Capstone Project, I get to work on a final assignment where we have to define a business problem, look for data in the web and then use Foursquare location data to compare different neighborhoods of New York City (my chosen city) to figure out which neighborhood is suitable for starting a restaurant business. To complete the task, I begin with designing the problem, data preparation and then the final analysis section step by step. Detailed codes and images are given in Github and link can be found at the end of the post.

## **Background**

The City of New York, is the most populous city in the United States. With an 2018 population of 8,398,748 distributed over about 302.6 square miles (784 km<sup>2</sup>); it is the most densely populated major city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business opportunities and business friendly environment. It has attracted many different players into the market. The city is a major center for banking and finance, communication, real estate, retailing, world trade, transportation, tourism, media & advertising... NYC is distinctive for its high concentrations of advanced service sector firms in the law, accountancy, management consultancy fields, high technology..

And that means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. That being said, any new business venture or expansion needs to be analyzed carefully. The insights derived from analysis will give us good understanding of the business environment which help in strategically targeting the market.

## **Description**

Having a successful restaurant in New York City is no easy feat, especially for those who are looking to transition from staff to owner.

When it comes to real estate, the key is “location, location, location.” The same holds true for restaurants. Where you operate plays a significant role in the success or failure of a business. When looking for the right space in New York City, consider how much you can afford. The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

So it is evident that to survive in such competitive market it is very important to strategically plan, because despite the stressful demands of running a restaurant, if done successfully, there are plenty of rewards. Various factors need to be studied in order to decide on the Location such as :

1. NYC Population
2. Are there any Farmers Markets nearby to purchase raw ingredients.
3. Are there any venues like Parks nearby where floating population is high
4. Who are the competitors in that location?

5. Menu of the competitors
6. Segmentation of the Borough
7. Saturated markets
8. Untapped markets

An X Company needs to choose the correct location to start its first business. The choice of location is very important as it has to be in an area zoned for businesses.

## **Audience**

This would interest anyone who wants to start a new restaurant in NYC.

## **Success criteria**

The success criteria of this project will be a good recommendation of neighborhood choice to X Company.

The neighborhoods in New York City are located within the five boroughs of the City of New York. Their names and borders are not officially defined, and they change from time to time.

The five actual boroughs are : Brooklyn, the Bronx, Manhattan, Queens, and Staten Island. Each borough has the same boundaries as a county of the state.

## **2. Data**

We need data from reliable sources for analysis. To understand our problem and quantify results we will use following data:

1 : we will need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and longitude coordinates of each.[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

	<b>Borough</b>	<b>Neighborhood</b>	<b>Latitude</b>	<b>Longitude</b>
<b>0</b>	Bronx	Wakefield	40.894705	-73.847201
<b>1</b>	Bronx	Co-op City	40.874294	-73.829939
<b>2</b>	Bronx	Eastchester	40.887556	-73.827806
<b>3</b>	Bronx	Fieldston	40.895437	-73.905643
<b>4</b>	Bronx	Riverdale	40.890834	-73.912585

2 : the DOHMH Farmers Markets dataset :

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2> <https://www.grownyc.org/greenmarketco/foodbox>

	FacilityName	Service Category	Service Type	Address	Address 2	Borough	ZipCode	Latitude	Longitude	AdditionalInfo	StartDa
0	Inwood Park Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	Isham St bet Seaman & Cooper	NaN	Manhattan	10034	40.869009	-73.920320	Open year-round	NaN
1	82nd Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	82nd St bet 1st & York Aves	NaN	Manhattan	10028	40.773448	-73.948954	Open year-round	NaN
2	1 Centre Street	Farmers Markets and Food Boxes	Food Boxes	1 Centre Street	South Building, 9th Floor	Manhattan	11101	40.713028	-74.003753	Open year-round	NaN
3	125th Street Farmers Market	Farmers Markets and Food Boxes	Farmers Markets	125th St & Adam Clayton Powell Jr Blvd	NaN	Manhattan	10027	40.808981	-73.948327	Market open dates: 6/13/2017 to 11/21/2017	06/13/21
4	170 Farm Stand	Farmers Markets and Food Boxes	Farmers Markets	170th St & Townsend Ave	NaN	Bronx	10452	40.840095	-73.916827	Market open dates: 7/5/2017 to 11/22/2017	07/05/21

3 : we will get data from wikipedia as well :

- a. New York City Demographics & Population
- b. Cuisine of New York city

[https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City) [https://en.wikipedia.org/wiki/Economy\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Economy_of_New_York_City) [https://en.wikipedia.org/wiki/Portal:New\\_York\\_City](https://en.wikipedia.org/wiki/Portal:New_York_City) [https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City) [https://en.wikipedia.org/wiki/List\\_of\\_Michelin\\_starred\\_restaurants\\_in\\_New\\_York\\_City](https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City)

	Racial composition	2010[237]	1990[239]	1970[239]	1940[239]
0	White	44.0%	52.3%	76.6%	93.6%\n
1	—Non-Hispanic	33.3%	43.2%	62.9%[240]	92.0%\n
2	Black or African American	25.5%	28.7%	21.1%	6.1%\n
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%[240]	1.6%\n
4	Asian	12.7%	7.0%	1.2%	—\n

4 : Foursquare Developers Access to venue data: <https://foursquare.com/>

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

### **3. Methodology :**

Our main goal is to get the best location for the X company to launch the restaurant business.

#### **Analytic approach**

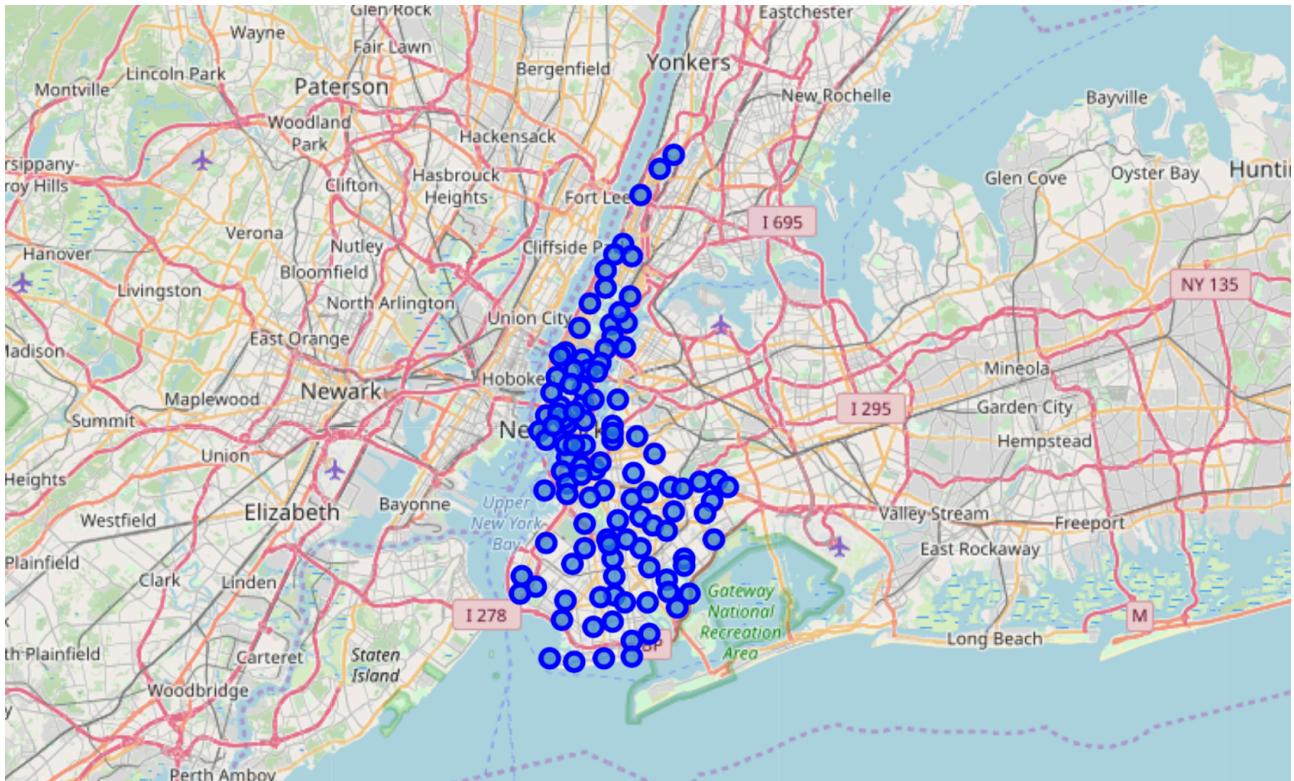
In NYC, we have 5 boroughs; we will be using the 'explore' function to get the most common venue categories in each neighborhood, and then using this feature to group the neighborhoods into clusters. To do so, we will use the k-means clustering algorithm. Finally, we will use the Folium library to visualize the neighborhoods in Brooklyn and Manhattan and their emerging clusters. We will proceed as follows : we'll divide this notebook in two sections; the first one to explore neighborhoods in Brooklyn and Manhattan, and then the second one to explore neighborhoods in Bronx, Queens and Staten Island.

#### **Methodology:**

New York City geographical coordinates data has been used as input for the Foursquare API, that has been leveraged to provision for venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

#### **Brooklyn and Manhattan :**

#### **Visualization :**

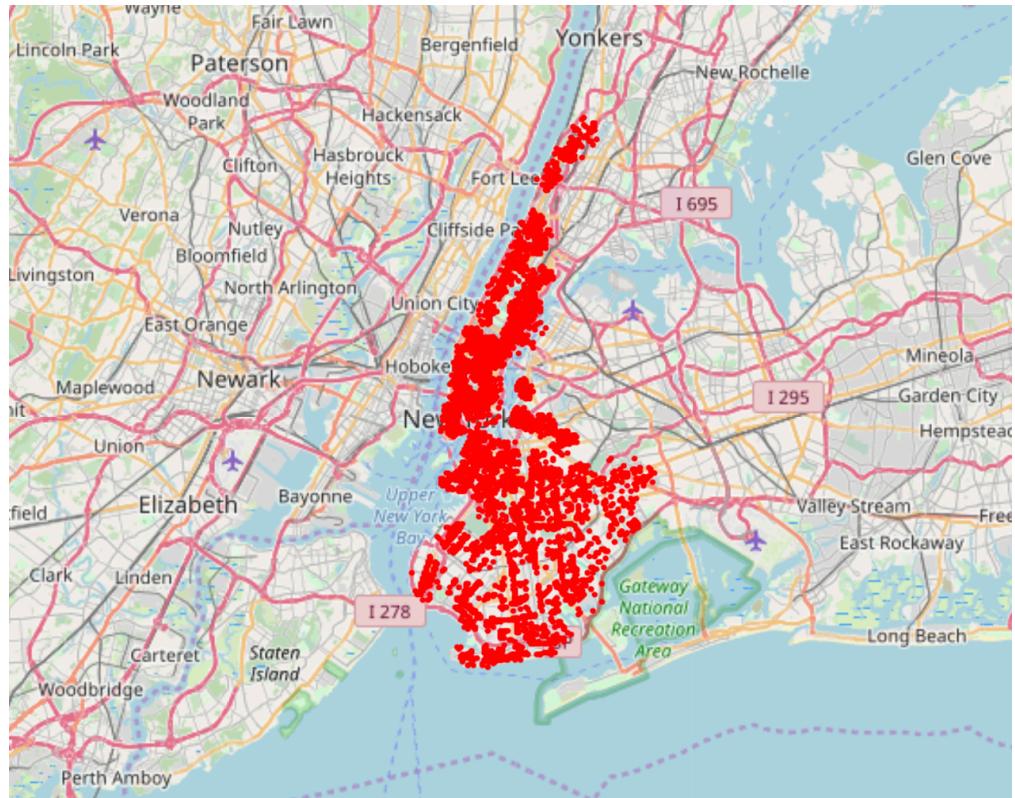


The venues data we get after using the geographical coordinates are the top 200 venues in a radius of 1000 meters.

	<b>Neighborhood</b>	<b>Neighborhood Latitude</b>	<b>Neighborhood Longitude</b>	<b>Venue</b>	<b>Venue Latitude</b>	<b>Venue Longitude</b>	<b>Venue Category</b>
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop

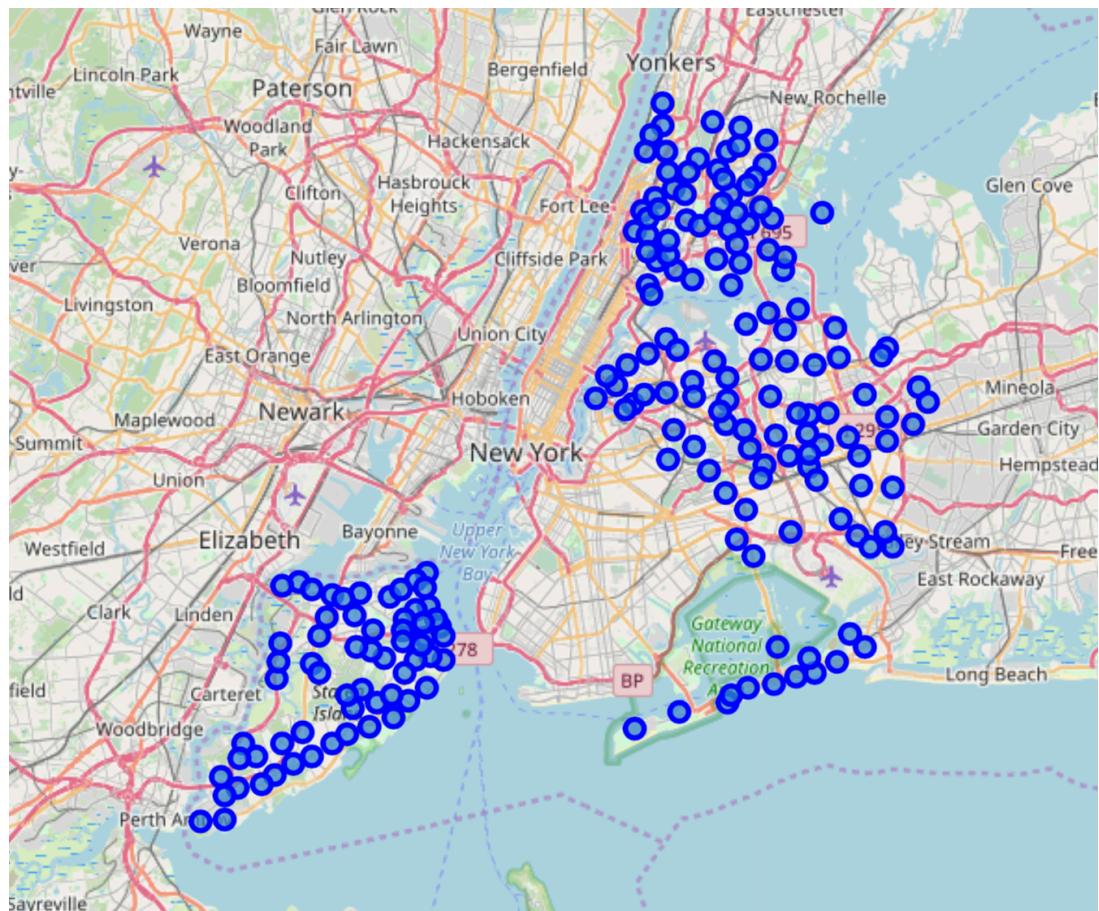
## **Venues visualization :**

The dataframe has 9708 venues and 397 unique venue types.



## Bronx, Queens and Staten Island :

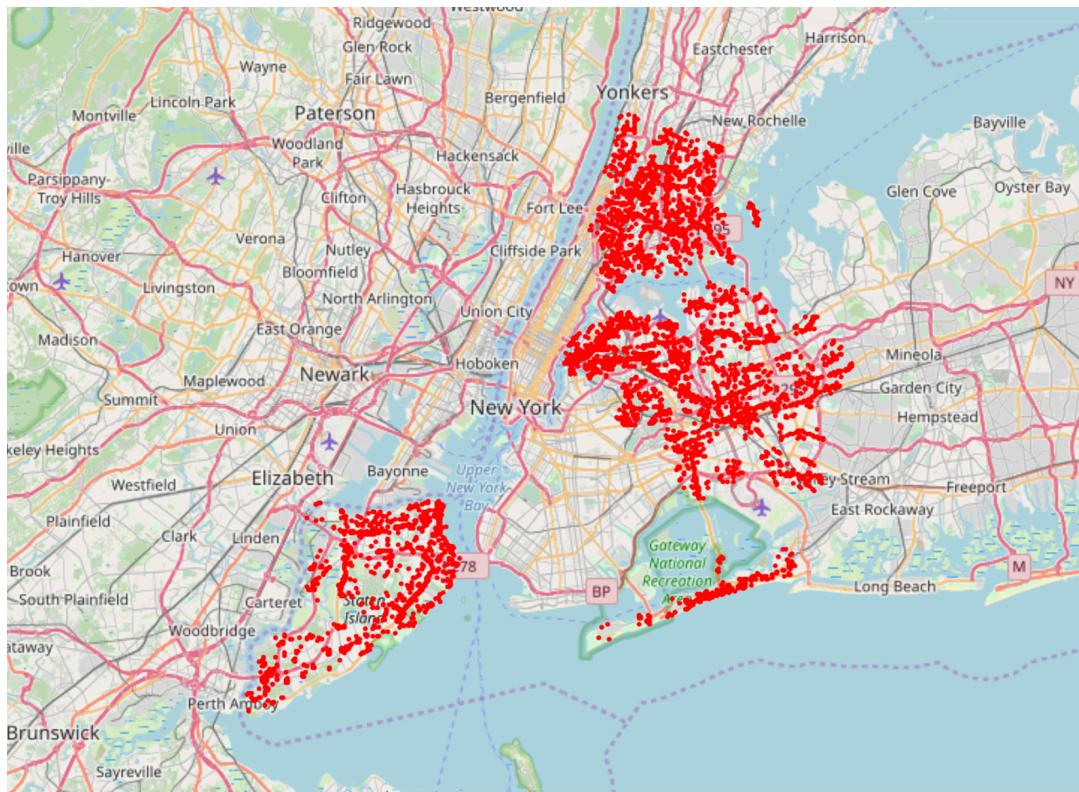
**Bronx, Queens and Staten Island neighborhoods visualization:**



## Queens and Staten Island Venues visualization:

The dataframe has 10807 venues and 387 unique venue types.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Jimbo's	40.891740	-73.858226	Burger Joint



### 4. Results :

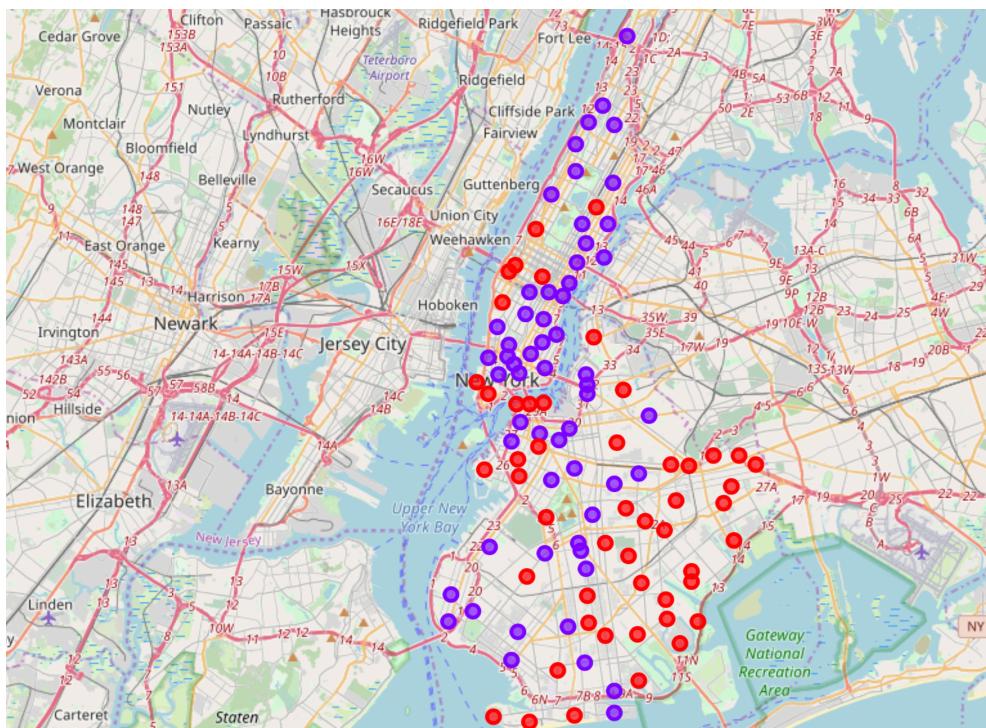
The venues data shown below, we filtered them, and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens ans Staten Island. Because we are focusing only on ‘restaurants’ business.

## Clustering :

To do so, to the neighborhoods, we used K-Means clustering algorithms. K-Means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

## Brooklyn & Manhattan :

In the following map, we can see different types of clusters created by using K-Means for Brooklyn and Manhattan.



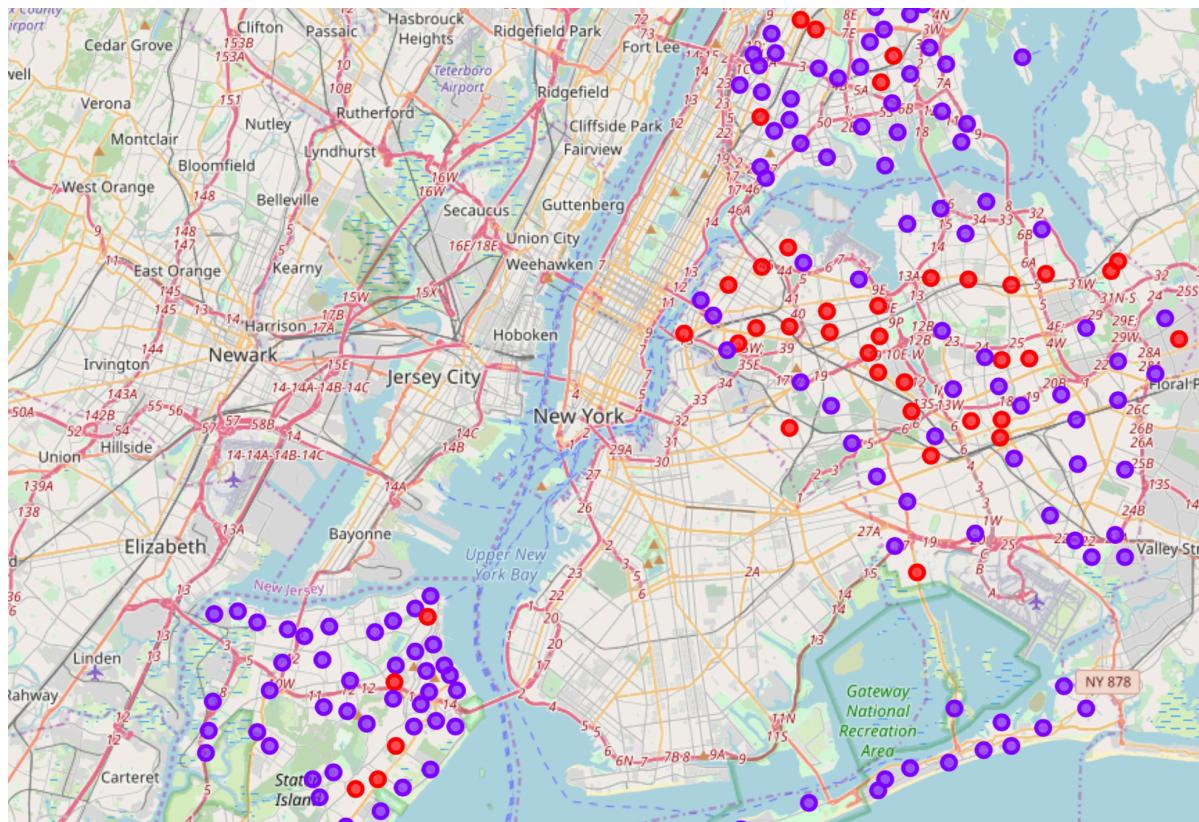
Cluster 0 :It has the smallest value. That means that the market is not saturated.

Cluster 1: It has the highest value. It shows that the market is saturated. Number of restaurants is very high.

—> There are no untapped neighborhoods in Brooklyn & Manhattan.

## Bronx, Queens and Staten Island

In the following map, we can see here too the different types of clusters created.



Cluster 0 : the total has the smallest value, so the market is not saturated. Thus, there are untapped neighborhoods.

Cluster 1 : the total is high. That means markets are saturated, and numbers of restaurants are very high.

### **5. Discussion :**

In the light of the results we had at the end, we can say that :

There is a possibility to increase Farmers markets in Bronx, Queens and Staten Island

In both Manhattan and Brooklyn, restaurants of cuisines of many countries are available. So the risk can be taken with great menu to propose.

Bronx, Queens and Staten Island also has good number of restaurants but not too many, thus it can be explored.

## **6. Conclusion :**

In a fast-moving world, there are many real-life problems or scenarios where data can be used to find solutions to those problems. Like seen in the example above, data was used to cluster neighborhoods in New York City based on the most common food venues (Restaurants) in its 5 boroughs. The results can help someone who wants to launch a restaurant business in the area.

I have made use of some frequently used python libraries to scrap web-data, use Foursquare API to explore neighborhoods of New York City and saw the results of segmentation of districts using Folium leaflet map.

Similarly, data can also be used to solve other problems, which most people face in metropolitan cities. Potential for this kind of analysis in a real-life problem is discussed in great detail.