The Battle of the Neighborhoods – Report

1. Introduction and Business Problem

New York City is the most populous city in the USA. It is considered as the financial capital and is a trade and commerce hub. The city is a melting pot of cultures and attracts people from all over the world who call it their home.

This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

Problem Description

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of New York is famous for its excellent cuisine. Its food culture includes an array of international cuisines influenced by the city's immigrant history.

1. Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Italian immigrants - New York-style pizza and Italian cuisine
3. Jewish immigrants and Irish immigrants - pastrami and corned beef
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
5. mobile food vendors - Some 4,000 licensed by the city
6. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
7. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin.

So, it is evident that to survive in such competitive market it is particularly important to strategically plan. Various factors need to be studied in order to decide on the Location such as:

1. New York Population
2. New York City Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?

4. Are there any venues like gyms, entertainment zones, parks etc nearby where floating population is high etc
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough
8. Untapped markets
9. Saturated markets etc
   The list can go on...

Even if the company is well-funded, it will need to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. The first move is especially important, thereby choice of location is very important.

Target Audience

To recommend the correct location, XYZ Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighbourhood of New York City will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made.

This would interest anyone who wants to start a new restaurant in New York city.

Success Criteria

The success criteria of the project will be a good recommendation of borough/neighbourhood choice to XYZ Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

2. Data

We will be using the following datasets for analysing New York City.

1. New York City has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood.
Link to the dataset is https://geo.nyu.edu/catalog/nyu_2451_34572

	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

2. The second data set which will be used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
Website-https://www.grownyc.org/greenmarketco/foodbox

GrowNYC's Fresh Food Box Program is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.
A farmers' market is often defined as a public site used by two or more local or regional producers for the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

	Borough	Market Name	Street Address	Latitude	Longitude	Days of Operation	Hours of Operations	Season Dates	Accepts EBT	Open Year-Round	Stellar Cooking Demonstrations	Food Activities for Kids	Location Point
0	Brooklyn	Urban Oasis Farmers' Market	681 Clarkson Ave.	40.656255	-73.936608	Wednesday	2pm-5:30pm	06/24/20-11/04/20	No	No	NaN	NaN	(40.656255, -73.936608)
1	Staten Island	Goodhue Center Go!Healthy Farm Stand	301 Prospect Ave	40.638983	-74.097422	Tuesday	11am-2pm	07/06/20-11/20/20	Yes	No	NaN	NaN	(40.638983, -74.097422)
2	Manhattan	Morningside Park's Down to Earth Farmers' Market	W 110th St & Manhattan Ave	40.801155	-73.959647	Saturday	9am-4pm (winter closes at 3pm)	Year-Round	Yes	Yes	NaN	NaN	(40.801155, -73.959647)
3	Bronx	170 Farm Stand	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30pm-6:30pm	7/15/20-11/25/20	Yes	No	NaN	NaN	(40.839882, -73.916783)
4	Manhattan	Go!Healthy Farm Stand at Mirabal Sisters Campus	21 Jumel Pl.	40.839305	-73.936159	Friday	11am-2pm	07/07/2020-11/20/20	Yes	No	NaN	NaN	(40.839305, -73.936159)

3. For the below analysis we will get data from Wikipedia as given below :

1. New York Population
2. New York City Demographics
3. Cuisine of 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/Portal: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

	Racial Composition	2010	1990	1970	1940
0	White	44.00%	52.30%	76.60%	93.60%
1	Non-Hispanic	33.30%	43.20%	62.90%	92%
2	Black or African American	25.50%	28.70%	21.10%	6.10%
3	Hispanic or Latino	28.60%	24.40%	16.20%	1.60%
4	Asian	12.70%	7.00%	1.20%	--

4. New York City geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City. The below is image of the Foursquare API data.

	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

Business Understanding:

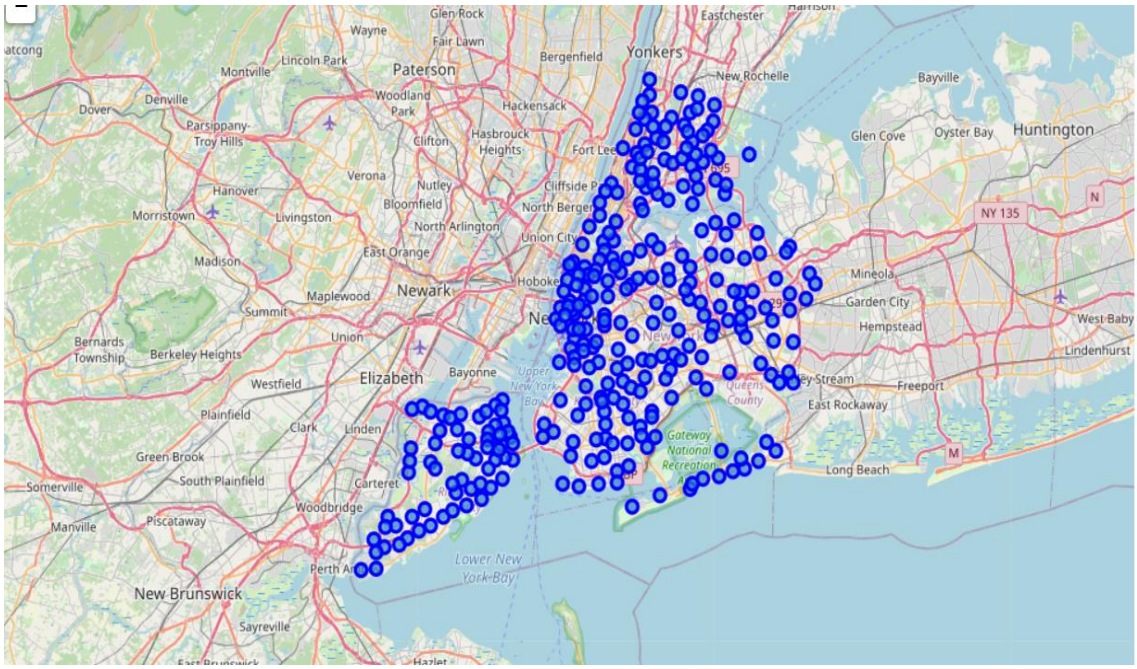
The main goal is to find an optimum location for a new restaurant business in New York City.

Analytic Approach:

New York City has a total of 5 boroughs and 306 neighborhoods. In the first part, Manhattan and Brooklyn are clustered. In the second part, The Bronx, Queens and Staten Island are clustered.

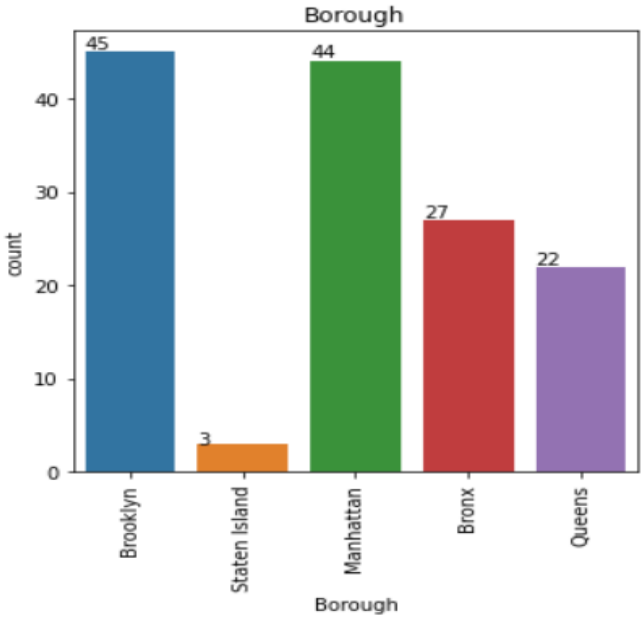
Exploratory Data Analysis:

1. NYC Geographical data
    a. The data from newyork_data.json is loaded and explored. It contains the coordinates of NYC neighborhoods
    b. It is transformed to a pandas dataframe
    c. This data is further used to get venues data from Foursquare
    d. Geopy and folium libraries are used to create maps with neighborhoods superimposed on top

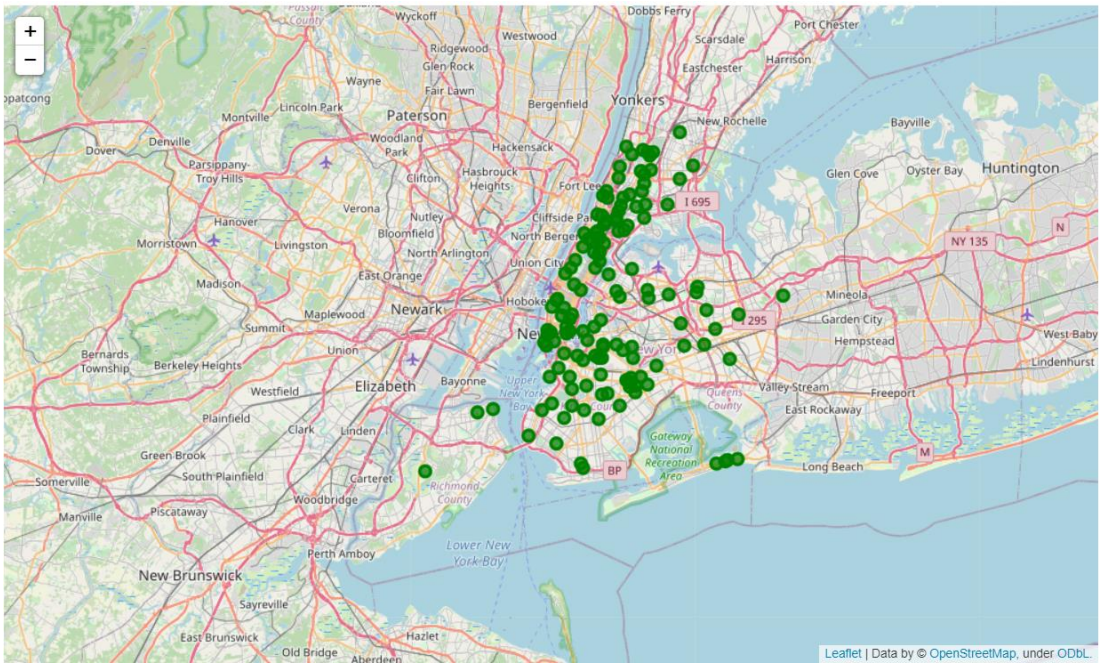


2. This data set is the DOHMH Farmers Market data.

There are a total of 141 farmer markets in New York City, mostly in Brooklyn and Manhattan.



The farmers markets are visualised as follows:



3. To analyse New York City's population, demographics and cuisine. Data was scrapped from Wikipedia pages. The Python library BeautifulSoup was used. BeautifulSoup is a python package for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

Population:

- Manhattan is the smallest and most densely populated borough
- Brooklyn has the highest population
- Queens is the largest borough in terms of area

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,418,207	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	368,500	22.83	59.13
3	Queens	Queens	2,253,858	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	30,500	58.37	151.18
5		City of New York	8,336,817	842.343	302.64	783.83
6		State of New York	19,453,561	1,731.910	47,126.40	122,056.82

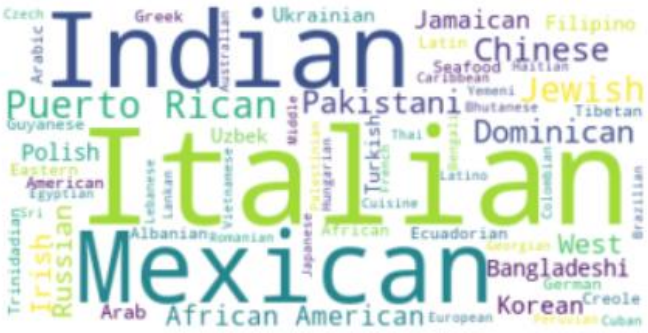
Demographics:

- NYC is the most populated city of the US, with an estimated record of high of 8,622,298 as of 2017. There is more immigration into the city that emigration
- The reason NYC has restaurants serving various cuisines from many countries is due to its racial composition. This also increases the scope for restaurant businesses in NYC

	Racial Composition	2010	1990	1970	1940
0	White	44.00%	52.30%	76.60%	93.60%
1	Non-Hispanic	33.30%	43.20%	62.90%	92%
2	Black or African American	25.50%	28.70%	21.10%	6.10%
3	Hispanic or Latino	28.60%	24.40%	16.20%	1.60%
4	Asian	12.70%	7.00%	1.20%	--

Cuisine:

- Data from the Wikipedia page for NYC cuisines was scraped and a word cloud was created
- Most preferred cuisine in NYC- Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani and Dominican



Most preferred cuisines in Brooklyn are Italian, Puerto Rican, Mexican.

Most preferred cuisines in Manhattan are Italian, Puerto Rican, Chinese, Indian.

Most preferred cuisines in Queens are Pakistani, Indian, Irish, Chinese, Korean.

Most preferred cuisine in the Bronx are Italian, Puerto Rican, Dominican, Albanian.

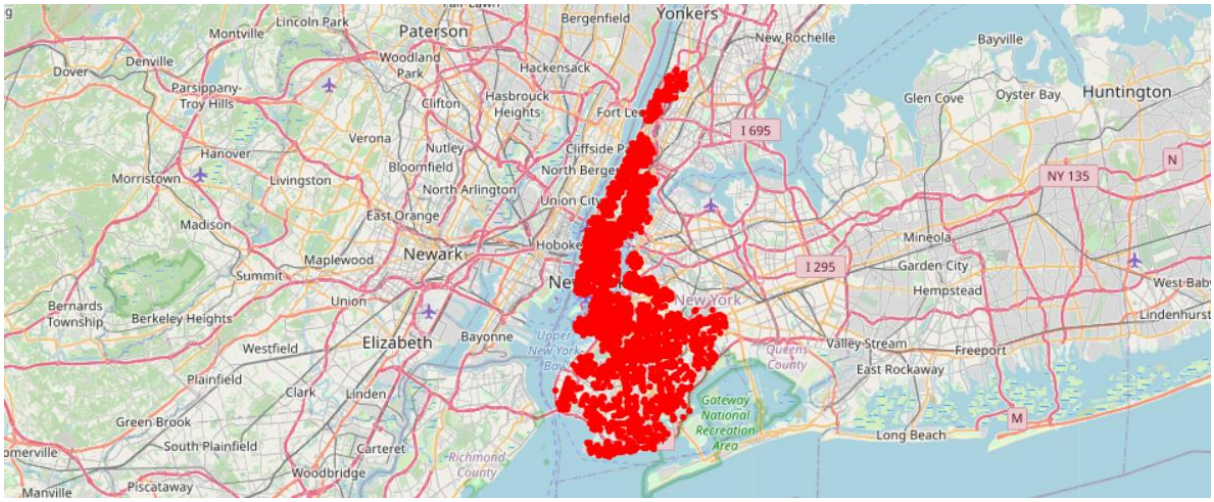
Unfortunately, the data about the cuisines in Staten Island was limited. Hence, a word cloud could not be generated.

4. The geographical coordinates of NYC must be utilized as input for the Foursquare API which is then leveraged to provide venue information for each neighborhood. We used Foursquare API data to explore the neighborhoods of NYC. Using the geographical coordinates of each neighborhood Foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

Part 1- Brooklyn and Manhattan

The generated data frame has 9627 venues and 403 unique venue types across 110 neighborhoods.

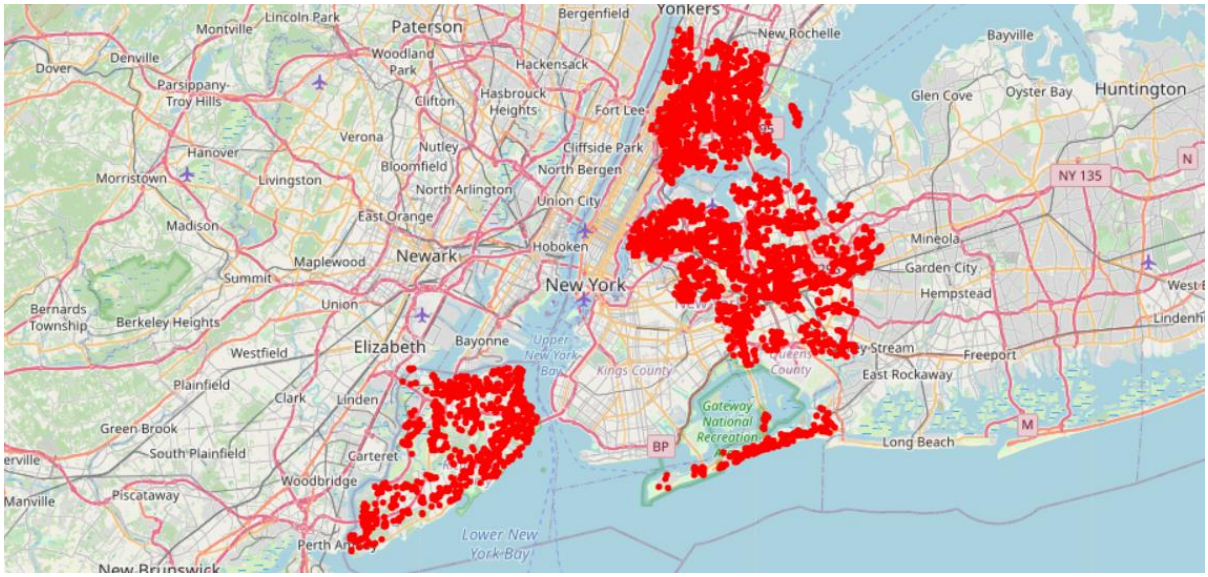
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett 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



Part 2- The Bronx, Queens and Staten Island

The generated data frame has 10826 venues and 389 unique venue types across 110 neighborhoods.

	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	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy



4. Output

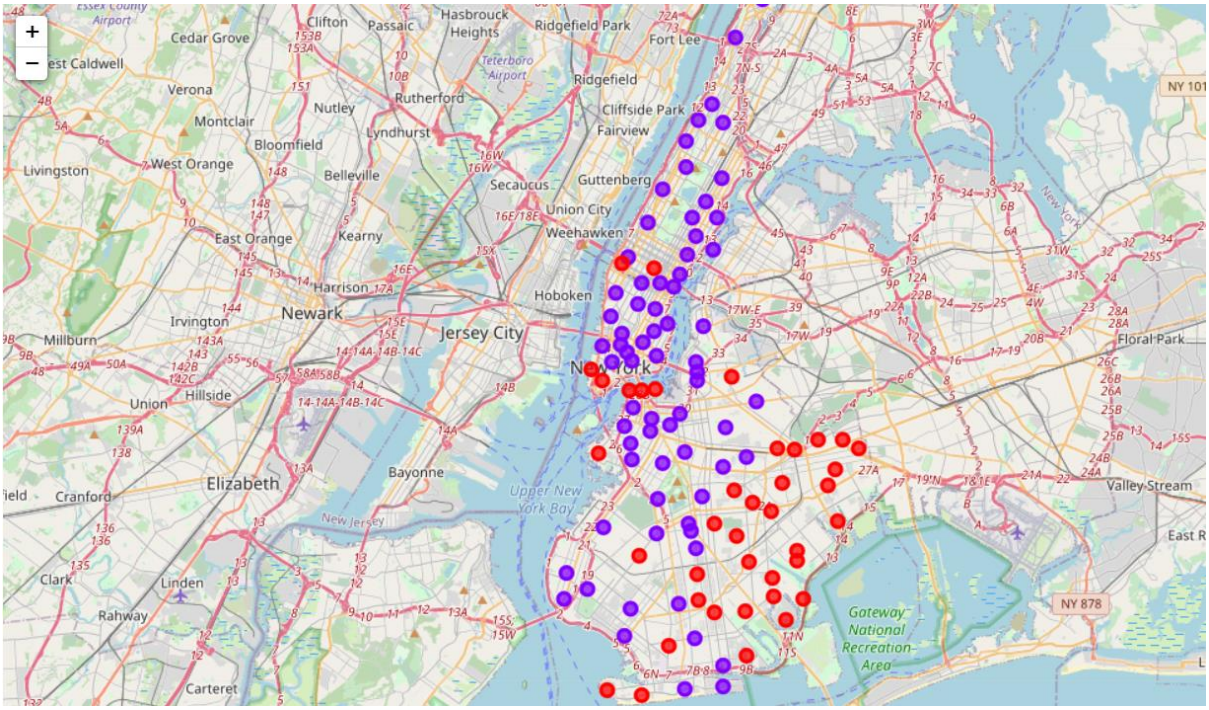
From the venues data, we filtered and used only restaurant data for Brooklyn and Manhattan clustering and for The Bronx, Queens and Staten Island clustering.

Neighborhoods K-Means clustering based on mean occurrence of venue category:

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

Brooklyn and Manhattan:

The map below shows the clustering of neighborhoods in Brooklyn and Manhattan.

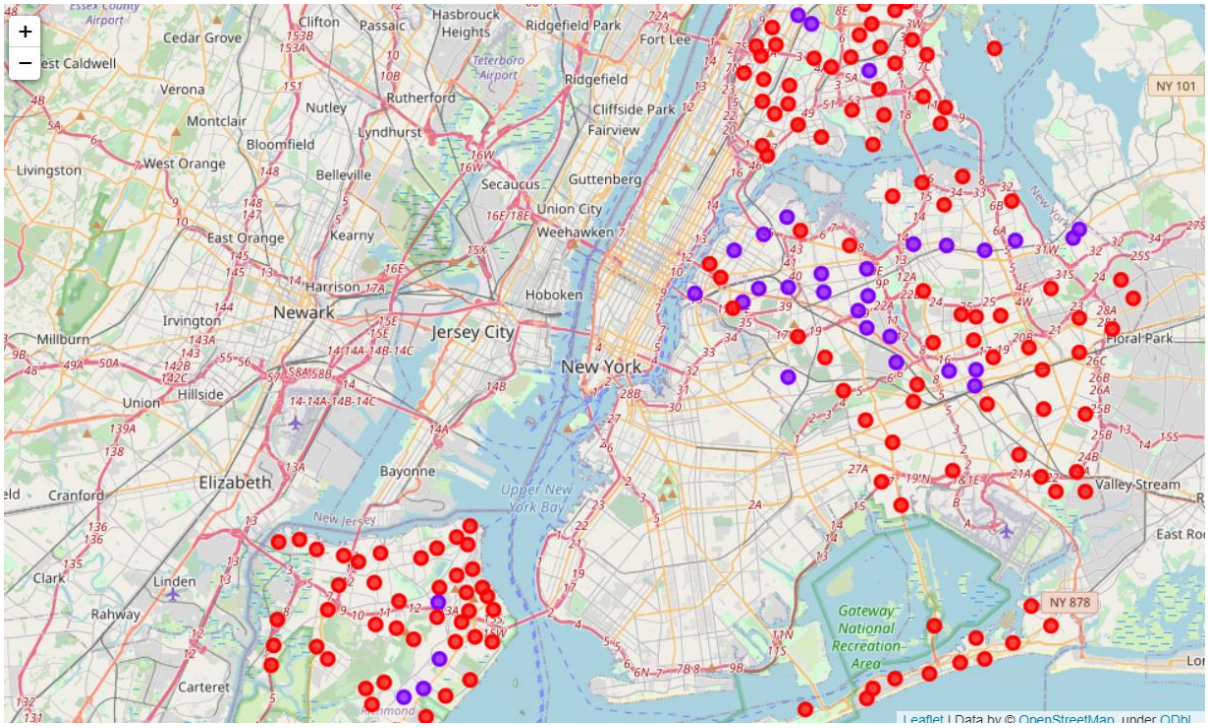


- Cluster 0: the total and total sum has the smallest value. It shows that the market is not saturated. Red on map
- Cluster 1: total and total sum has a large value. It shows that the market is saturated and that there is a lot of competition. Purple on map

In conclusion, there are no untapped neighborhoods in Brooklyn and Manhattan.

The Bronx, Queens and Staten Island:

The map below shows the clustering of neighborhoods in The Bronx, Queens and Staten Island.



- Cluster 0: the total and total sum has the smallest value. It shows that the market is not saturated. Red on map
- Cluster 1: total and total sum has a large value. It shows that the market is saturated and that there is a lot of competition. Purple on map

The untapped markets from Cluster 0 are as follows:

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Clason Point	40.806551	-73.854144	0	0
1	Staten Island	Todt Hill	40.597069	-74.111329	0	0
2	Staten Island	South Beach	40.580247	-74.079553	0	0
3	Staten Island	Port Ivory	40.639683	-74.174645	0	0
4	Staten Island	Butler Manor	40.506082	-74.229504	0	0
5	Staten Island	Rossville	40.549404	-74.215729	0	0
6	Staten Island	Bloomfield	40.605779	-74.187256	0	0

5. Points to be noted

- There is scope to explore cuisines in various neighborhoods in The Bronx, Queens and Staten Island
- There is also scope to increase the number of farmers markets in The Bronx, Queens and Staten Island
- In Manhattan and Brooklyn, a risk can be taken with a great menu on board. There are already a lot of cuisines from various countries available
- Neighborhoods- Staten Island: Todt Hill, South beach and Port Ivory

6. Conclusion

- Brooklyn and Manhattan have high concentration of restaurant businesses. It is an incredibly competitive market
- The Bronx, Staten Island and Queens also have a good number of restaurants but not as much as Brooklyn and Manhattan. Hence, these areas can be explored.
- Depending on the neighborhood or cuisine, a suitable venue with low risk and competition can be identified