

The Battle of the Neighborhoods – Report

Introduction/Business Problem

New York City is one of the most populated cities in the United States. It is a global hub of business and commerce. The city is a major centre for banking, finance, theatre, fashion, and the arts in the United States. This fact means that the market is highly competitive. It means that New York is a vibrant city and also very expensive. For an entrepreneur, it involves significant risk, so in the beginning, any start up venture or expansion needs to be analysed carefully.

Problem Description:

New York City is famous in restaurants with different cuisines. In New York City, you can find international cuisine which is influenced by the city's immigrant history. It's a competitive business to hold a restaurant in NYC. The problem is to make a clear picture of competition in the restaurant market and find the best location with lots of potential customers and lesser competitors in the same cuisine. To solve that problem you have to make investigation about such factors as:

1. Population
2. Demographics
3. Suppliers of ingredients
4. Infrastructure
5. Pressure of competitors
6. Popular cuisine among local customers
7. Segmentation of the Borough
8. Untapped markets

Target Audience:

The objective is to locate and support the management which neighbourhood of New York City will be the best choice to start a restaurant. It also includes understanding the rationale of the recommendations made. This investigation can be helpful for every company who needs to choose the correct location to start its first venture. The first move is significant; thereby, the choice of location is crucial.

Success Criteria:

The success criteria of the project will be a good recommendation of borough choice based on nearest suppliers of ingredients, amount potential customers and pressure of competitors.

2. Data

For the analyses of capstone project «The Battle of Neighborhoods» for the New York City, we will be using the below datasets:

Data #1

The neighbourhood has in total around five boroughs and 306 districts. To segment the regions, we will need a dataset with five boroughs. This dataset exists for free on the web-page [here](#)

2014 New York City Neighborhood Names

Description: This New York City Neighborhood Names point file was created as a guide to New York City's neighborhoods that appear on the web resource, "New York: A City of Neighborhoods." Best estimates of label centroids were established at a 1:1,000 scale, but are ideally viewed at a 1:50,000 scale.

Publisher: [New York \(City\). Department of City Planning](#)

Collection: [Bytes of the Big Apple](#)

Place(s): [New York, New York, United States](#)

Subject(s): [Neighborhoods](#), [Neighborhood planning](#), and [Communities](#)

Format(s): [Shapefile](#)

Year(s): 2014

Held by: [NYU](#)

Preservation record: <http://hdl.handle.net/2451/34572>



Data #2

Other data will help us to analyse the location of the nearest suppliers of fresh food. For that investigation, we will use two sources:

- DOHMH Farmers Markets (data set is [here](#))



Data Dictionary - Dataset Information

General

Dataset Name	Farmers Markets and Food Boxes
Agency Name	DOHMH
Update Frequency	Biweekly
Dataset Description	This is a list of farmers markets and food box sites across the five boroughs. Farmers market vendors, products, and EBT acceptance varies by venue.
Dataset Keywords	Farmers Markets and Food Boxes
Dataset Category	

Detailed Description

List any additional information in order to provide context to the data for someone not familiar with your agency's operations.	This data is collected to help inform New Yorkers of the various farmers market locations around the city. These markets are physical sites where shoppers can purchase fresh-locally grown produce within the five boroughs. The Health Bucks program at the New York City Department collects this data from market to compile a list and a map of all New York City farmers markets and food box locations (map available on nyc.gov by searching "farmers markets"). The data is collected from market operators each spring and is updated upon hearing from market operators and community members about market closings, openings, updated hours of operation, and location changes. This data changes often and is only as accurate as the feedback we receive from farmers markets.
Why is this data collected?	
How is this data collected?	
What does each record represent?	
How can this data be used?	
What are the ideosyncracies or limitations of the data to be aware of?	This data changes frequently. Farmers markets and food box sites are indicated on this list only if information can be collected from the site. It may not be the most exhaustive list of farmers markets and food box sites in the city.

- Food Boxes dataset (data set is [here](#))



Data #3

Next important source is Wikipedia. From there we will get the data set about New York City:

1. Population (data set is [here](#))
2. Economic (data set is [here](#))
3. Cuisine (data set is [here](#))

Data #4

We will use the Foursquare API to explore areas in NYC. That analysis will be leveraging provision venues information for each neighbourhood

3.Methodology

Business Understanding:

Our main goal is to get optimum location for a new restaurant business in New York City.

Analytic Approach:

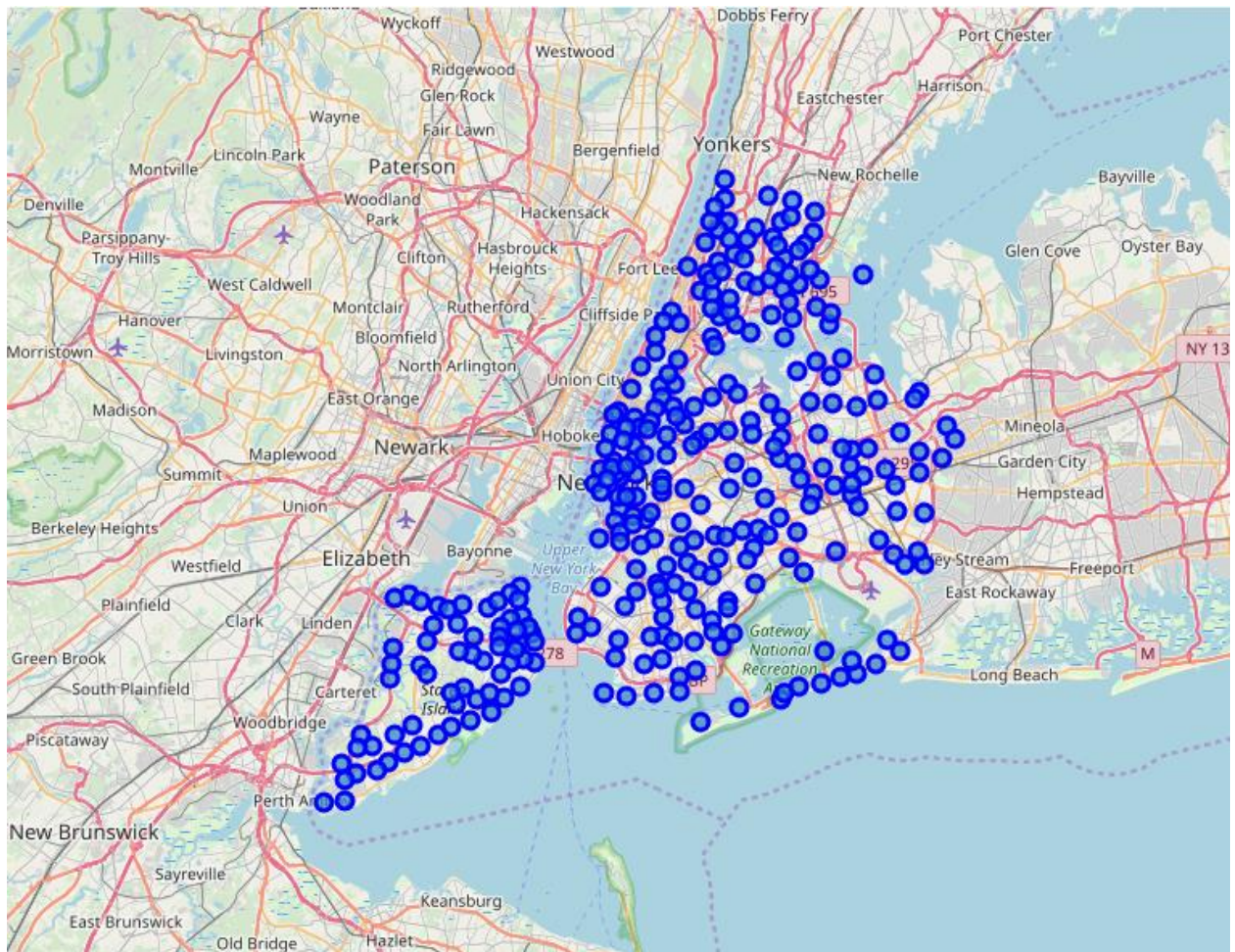
New York City neighbourhood has a total of 5 boroughs and 306 districts. In this project first part is clustering of Manhattan and Brooklyn. And second part is clustering of Bronx, Queens and Staten Island.

Exploratory Data Analysis

Step #1

New York City Geographical Coordinates Data.

1. In this, we load the data and explore data from newyork_data.json file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York City neighbourhoods.
4. This data will be used to get Venues data from Foursquare.
5. We used geopy and folium libraries to create a map of New York city with neighbourhoods superimposed on top.



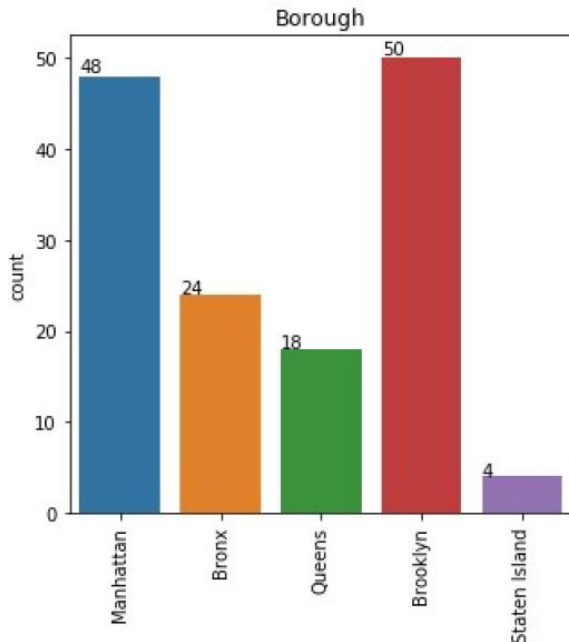
Step #2

New York City farmers market data.

We used DOHMH Farmers Markets and Food Boxes dataset data set. There are 144 Farmers Markets in NYC in total. The highest number is in Manhattan and Brooklyn.

And lowest in Queens, Bronx and Staten Island.

The visualisation of data is given below.



Step #3

To analyse New York city Population, Demographics and Cuisine, scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library. BeautifulSoup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping

1.New York Population:

- Manhattan (New York County) is the geographically smallest and most densely populated borough.
- Manhattan's (New York County's) population density of 72,033 people per square mile (27,826/km²) and 629,682 billion in GBP, what is the TOP of rate in USA.
- Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough.
- Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough.

	Borough	County	Estimate_2017	GBP_billions	GBP_per_capita	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,471,160	28.787	19,570	42.10	109.04	34,653	13,231
1	Brooklyn	Kings	2,648,771	63.303	23,900	70.82	183.42	37,137	14,649
2	Manhattan	New York	1,664,727	629.682	378,250	22.83	59.13	72,033	27,826
3	Queens	Queens	2,358,582	73.842	31,310	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	479,458	11.249	23,460	58.37	151.18	8,112	3,132
5		City of New York	8,622,698	93.574	302.64	806.863	783.83	28,188	10,947
6		State of New York	19,849,399	78.354	47,214	1,547.116	122,284	416.4	159

2.New York City Demographics:

New York City is the most populous city in the United States, with an estimated record high of 8,622,698 residents as of 2017.

The racial composition is as given below. It is the reason why New York city has restaurants serving cuisine from many countries such as Indian, African, Japan etc. It also increases the scope of restaurants business in New York City.

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

Step #4

3.The cuisine of New York City:

This data has been manually prepared. Data is taken from Wikipedia page [here](#). Using this data we did word cloud.

#1 NEW YORK CITY CUISINE : Most Preferred Food in New York City –Italian, Purto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.



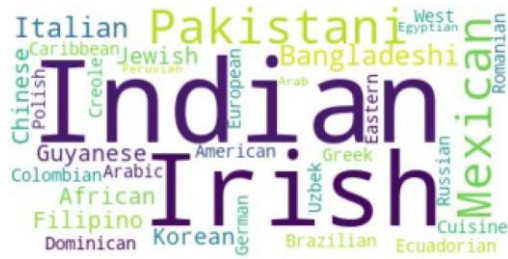
#2 BROOKLYN CUISINE -Most Preferred Food in Brooklyn is –Italian, Purto Rican & Mexican



#3 MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.



#4 QUEENS CUISINE - Most Preferred Food in Queens is – Indian, Irish, Pakistani and Mexican.



#5 THE BRONX CUISINE - Most Preferred Food in The Bronx is – Italian, Puerto Rican, Albanian and Dominican.



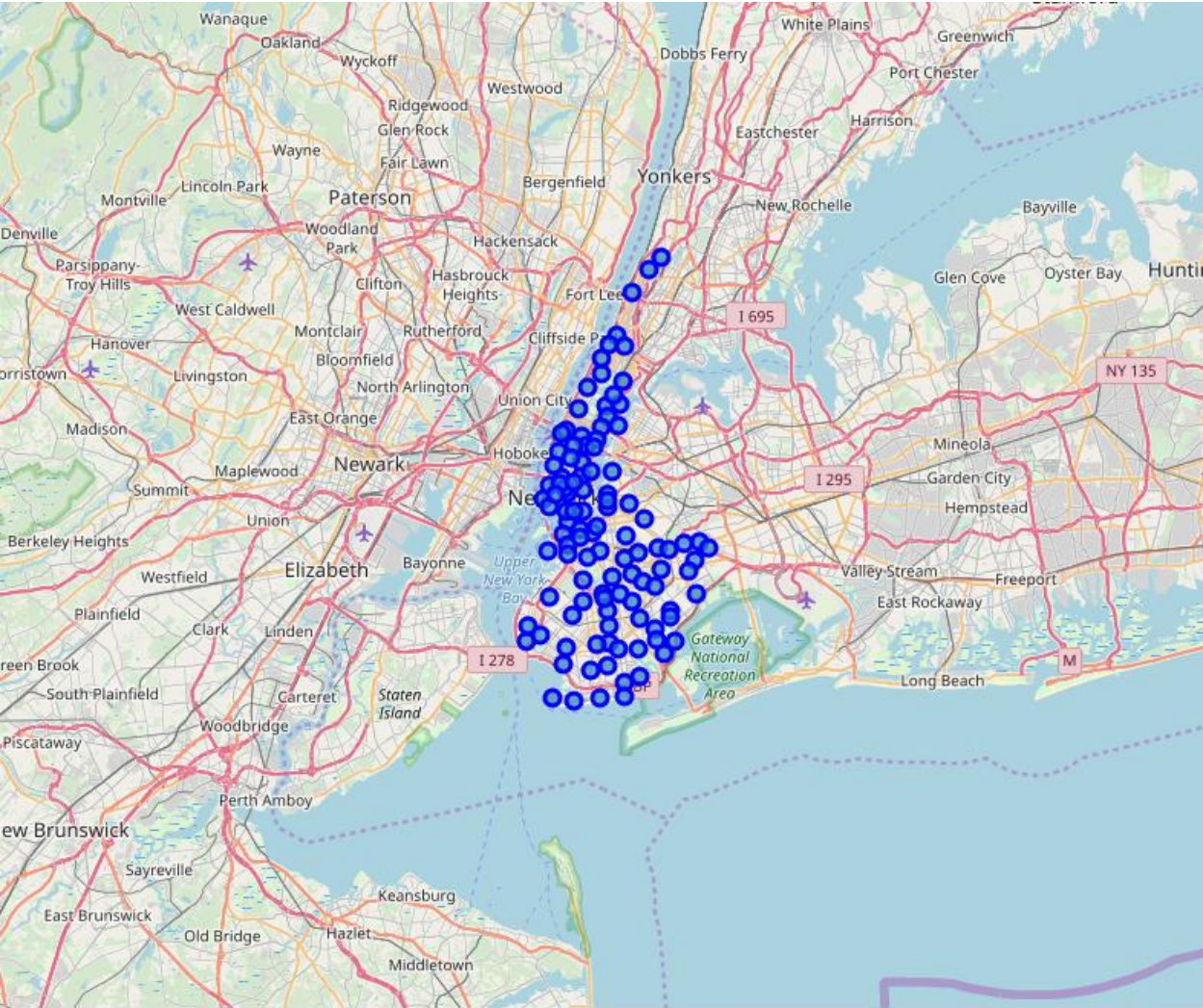
#6 THE STATEN ISLAND - Most Preferred Food – Italian, Puerto Rican, Albanian and Dominican as at Bronx.



Step #5

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

Brooklyn and Manhattan Visualization:

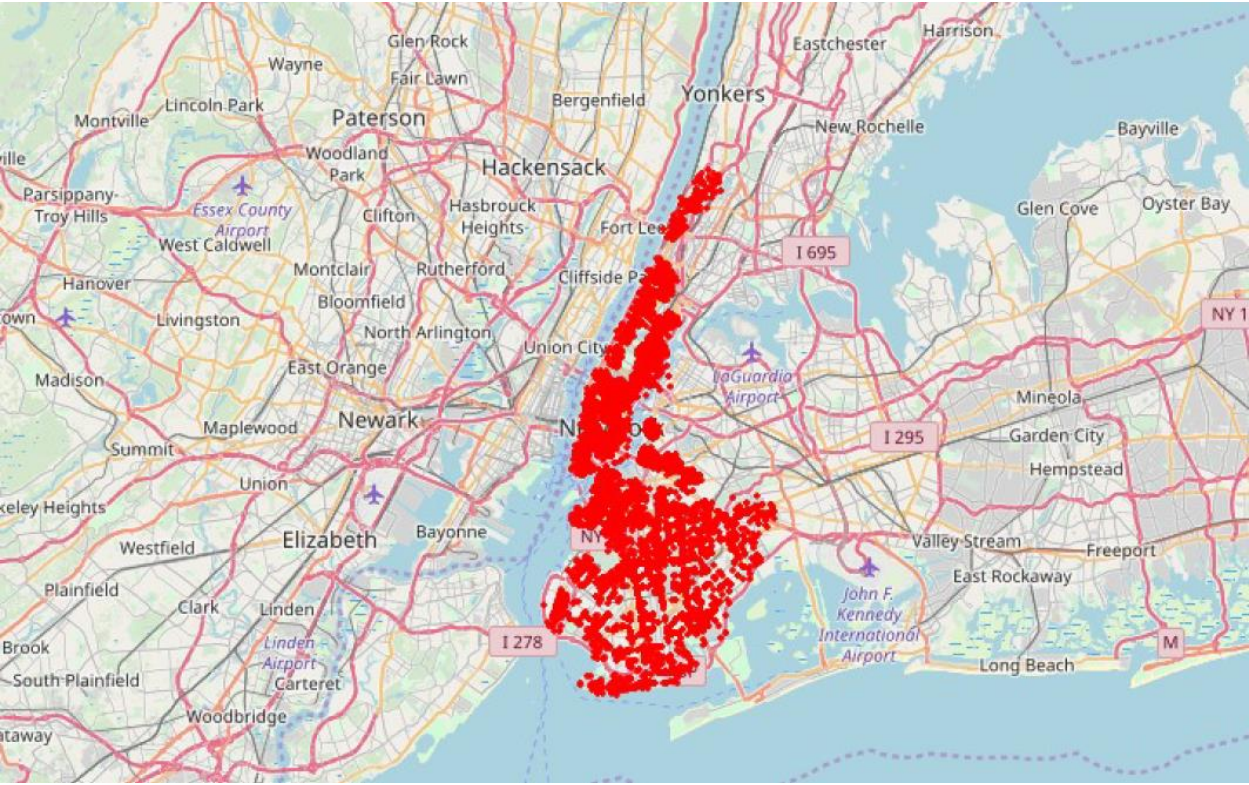


Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below.

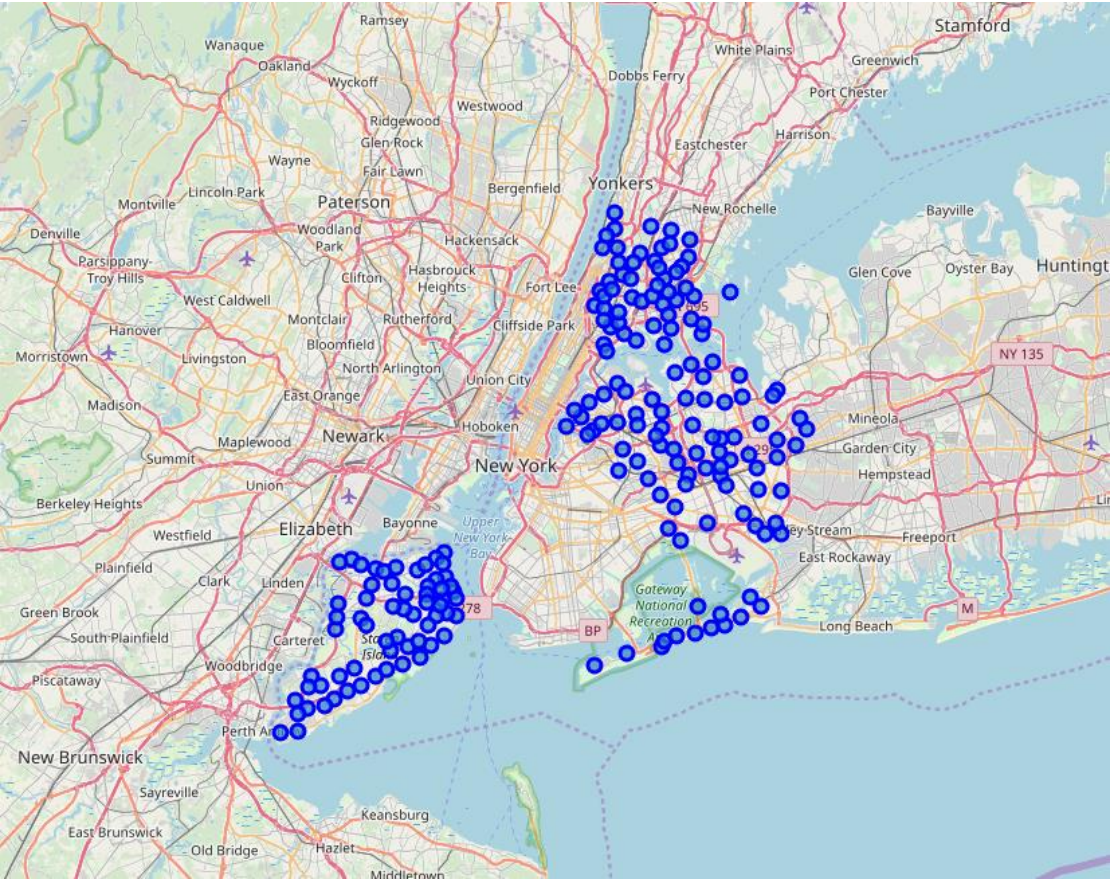
Brooklyn and Manhattan Venues:

	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	Rite Aid	40.896649	-73.844846	Pharmacy
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant

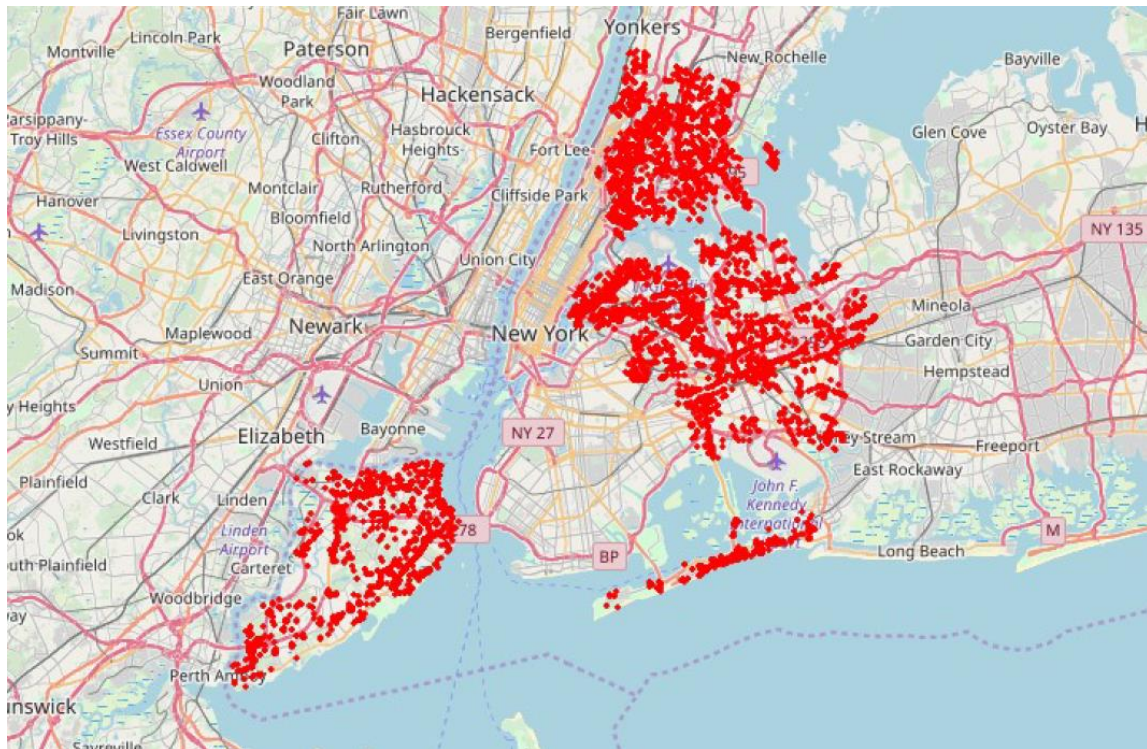
Brooklyn and Manhattan Venues Visualization: Generated below Brooklyn and Manhattan Venues Visualization around 9708 venues and 397 unique venue types.



Bronx, Queens and Staten Island Neighborhoods Visualization:



Bronx, Queens and Staten Island Venues Visualization has 10805 venues and 387 unique venue types.



Step #6

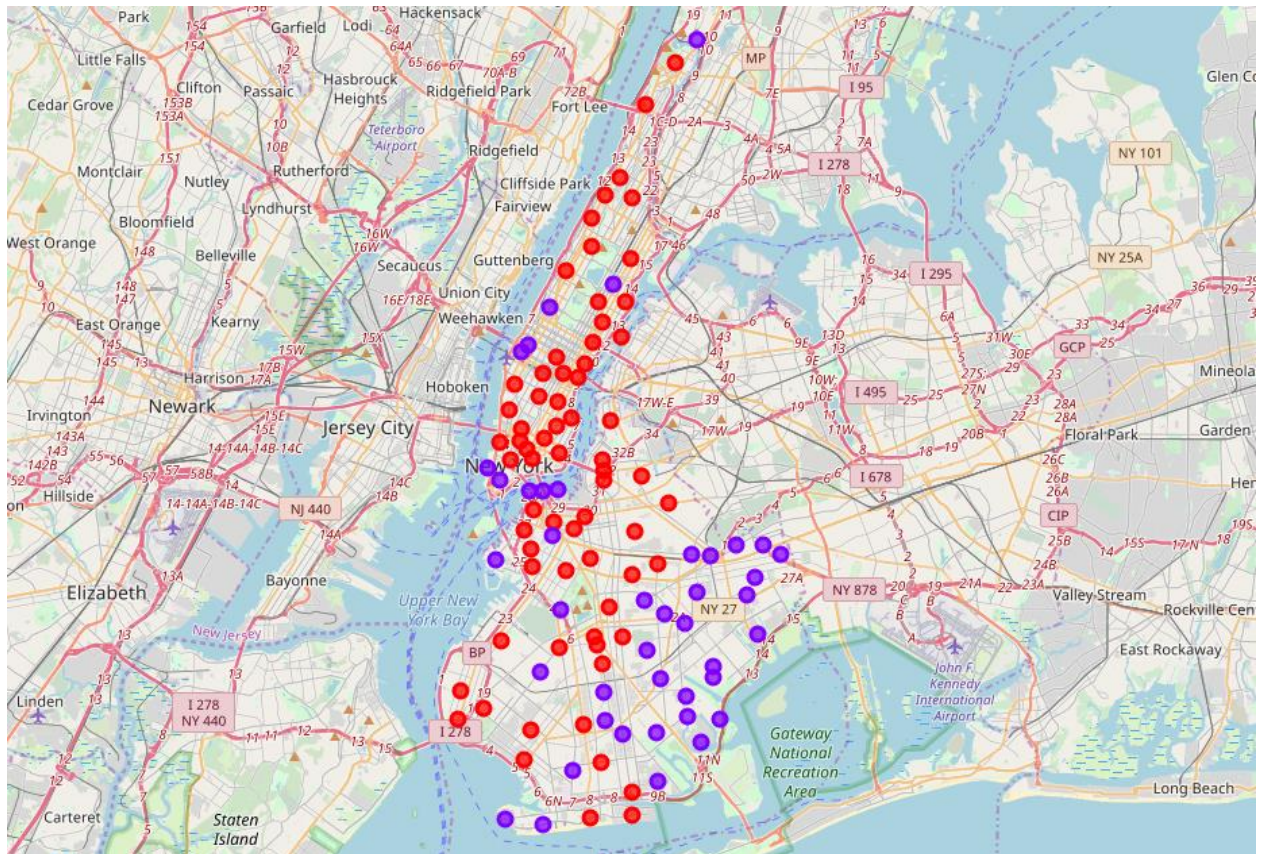
From these venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focus only on restaurants business.

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

To cluster the neighbourhoods into two clusters, we used the K-Means clustering Algorithm. k-means clustering aims to partition n observations into k clusters in which each view belongs to the group with the nearest mean. It uses iterative refinement approach.

Brooklyn & Manhattan:

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.

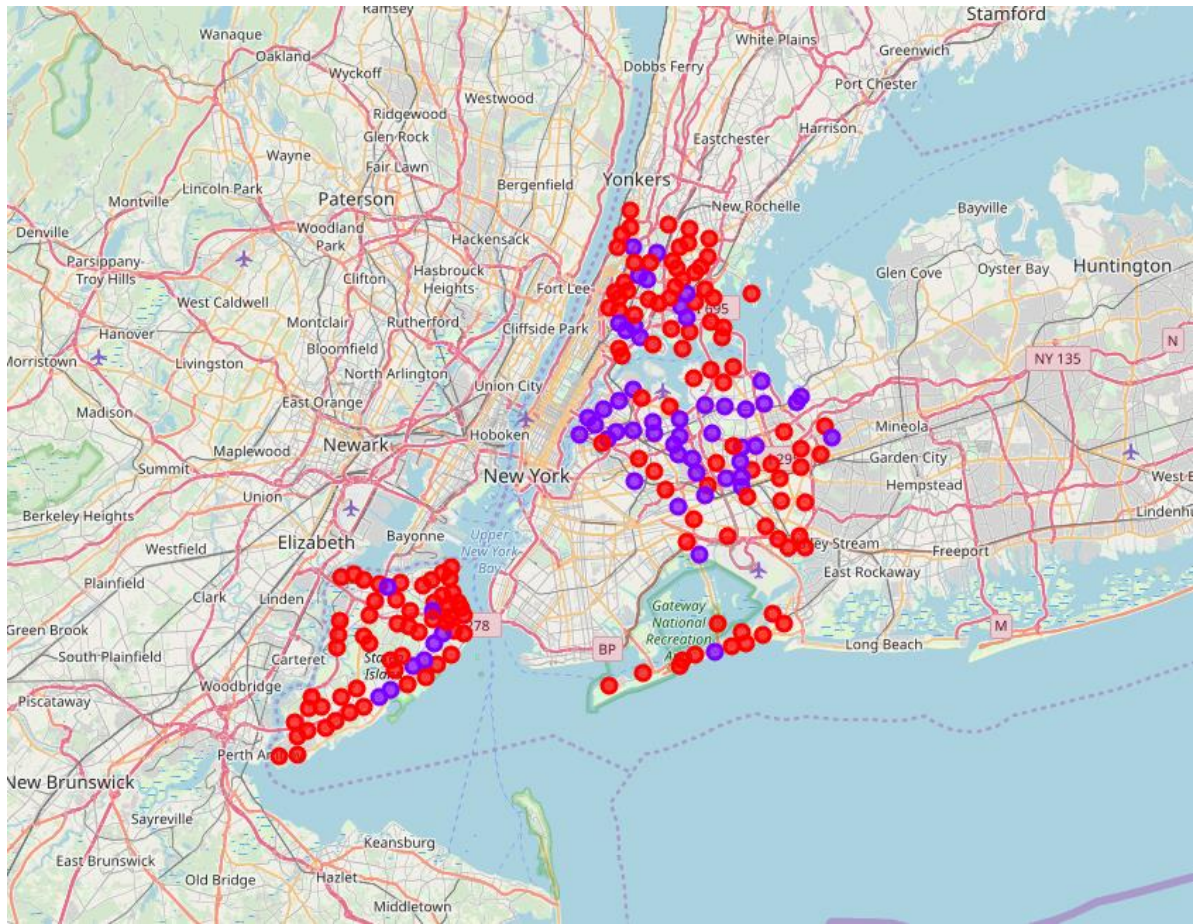


Cluster0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

Cluster1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. The number of restaurants are very high. There are no untapped neighbourhoods in Brooklyn and Manhattan.

Bronx, Queens and Staten Island:

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Bronx, Queens and Staten Island.



Cluster0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. There are untapped neighbourhoods. The list is as given below.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Staten Island	Todt Hill	40.597069	-74.111329	0	0
1	Staten Island	Port Ivory	40.639683	-74.174645	0	0
2	Staten Island	Butler Manor	40.506082	-74.229504	0	0
3	Staten Island	Bloomfield	40.605779	-74.187256	0	0

Cluster1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. The number of restaurants is very high.

Step #7

Discussion:

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.
2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So if the risk can be taken with the high menu on board. It also shows people love eating cuisines of various countries.

Step #8

Conclusion:

This analysis is performed on limited data. This may be right or maybe wrong. But if proper amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. Brooklyn and Manhattan have high concentration of restaurant business. Very competitive market. Bronx, Queens and Staten Island also has good number of restaurants but not as many as required. So this can be explored.

As per the neighbourhood or restaurant type mentioned like Indian Restaurant, analysis can be checked. A venue with lowest risk and competition can be identified.