

1. Introduction & Business Problem

Problem Background:

New York City is also known as the City of New York or simply New York which is the most populous city in the United States. There are estimated 8,398,748 population over a land area of about 302.6 square miles as of 2018. New York is also the most densely populated major city in the United States. Located at the southern tip of the state of New York, the city is the center of the New York metropolitan area, the largest metropolitan area in the world by urban landmass and one of the world's most populous megacities, with an estimated 19,979,477 people in its 2018 Metropolitan Statistical Area and 22,679,948 residents in its Combined Statistical Area. A global power city, New York City has been described as the cultural, financial, and media capital of the world, and exerts a significant impact upon commerce, entertainment, research, technology, education, politics, tourism, art, fashion, and sports. The city's fast pace has inspired the term New York minute. Home to the headquarters of the United Nations, New York is an important center for international diplomacy.

Situated on one of the world's largest natural harbors, New York City consists of five boroughs, each of which is a separate county of the State of New York. The five boroughs – Brooklyn, Queens, Manhattan, The Bronx, and Staten Island – were consolidated into a single city in 1898. The city and its metropolitan area constitute the premier gateway for legal immigration to the United States. As many as 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world as of 2016. As of 2019, the New York metropolitan area is estimated to produce a gross metropolitan product (GMP) of US\$1.9 trillion. If greater New York City were a sovereign state, it would have the 12th highest GDP in the world. New York is home to the highest number of billionaires of any city in the world.

New York City traces its origins to a trading post founded by colonists from the Dutch Republic in 1624 on Lower Manhattan; the post was named New Amsterdam in 1626. The city and its surroundings came under English control in 1664 and were renamed New York after King Charles II of England granted the lands to his brother, the Duke of York. New York was the capital of the United States from 1785 until 1790 and has been the largest US city since 1790. The Statue of Liberty greeted millions of immigrants as they came to the U.S. by ship in the late 19th and early 20th centuries and is an international symbol of the U.S. and its ideals of liberty and peace. In the 21st century, New York has emerged as a global node of creativity and entrepreneurship, social tolerance, and environmental sustainability, and as a symbol of freedom and cultural diversity. In 2019, New York was voted the greatest city in the world per a survey of over 30,000 people from 48 cities worldwide, citing its cultural diversity.

Many districts and landmarks in New York City are well known, including three of the world's ten most visited tourist attractions in 2013; a record 62.8 million tourists visited in 2017. Several sources have ranked New York the most photographed city in the world. Times Square, iconic as the world's "heart" and "crossroads", is the brightly illuminated hub of the Broadway Theater District, one of the world's busiest pedestrian intersections, and a major center of the world's entertainment industry. The names of many of the city's landmarks, skyscrapers, and parks are known internationally. Manhattan's real estate market is among the most expensive in the world. New York is home to the largest ethnic Chinese population outside of Asia, with multiple distinct Chinatowns across the city. Providing continuous 24/7 service, the New York City Subway is the largest single-operator rapid transit system worldwide, with 472 rail stations. The city has over 120 colleges and universities, including Columbia University, New York University, and Rockefeller University, ranked among the top universities in the world. Anchored by Wall Street in the Financial District of Lower Manhattan, New York has been called both the most economically powerful city and world's leading financial center and is home to the world's two largest stock exchanges by total market capitalization, the New York Stock Exchange and NASDAQ.

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 analyzed 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 ROI will be reasonable.

2. 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. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens

Italian immigrants - New York-style pizza and Italian cuisine

Jewish immigrants and Irish immigrants - pastrami and corned beef

Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city

mobile food vendors - Some 4,000 licensed by the city

Middle Eastern foods such as falafel and kebabs examples of modern New York street food

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 very 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

The list can go on...

Therefore, it is necessary to choose the correct location to start its first venture. If this is successful, they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

Target Audience:

To recommend the correct location, the objective is to locate and recommend to the management which neighborhood of Network 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.

Data:

Data 1 : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

This dataset exists for free on the web. 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

Data 2 : Second data 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>

Grow NYC'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.

	FacilityName	Service Category	Service_Type	Address	Address_2	Borough	ZipCode	Latitude	Longitude	AdditionalInfo	StartDate	EndDate	Monday	Tuesday	Wednesday	Thursday
0	Inwood Park Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	Islem St bet Seaman & Cooper	NaN	Manhattan	10034	40.869009	-73.920320	Open year-round	NaN	NaN	NaN	NaN	NaN	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	NaN	NaN	NaN	NaN	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	06/13/2017	11/21/2017	NaN	10am-7pm	NaN	NaN
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	07/05/2017	11/22/2017	NaN	NaN	2:30pm-6:30pm	NaN
5	175th Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	175th St bet Wadsworth Ave & Broadway	NaN	Manhattan	10033	40.845956	-73.937813	Market open dates: 6/29/2017 to 11/30/2017	06/29/2017	11/30/2017	NaN	NaN	NaN	8am-5pm

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

New York Population

New York City Demographics

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

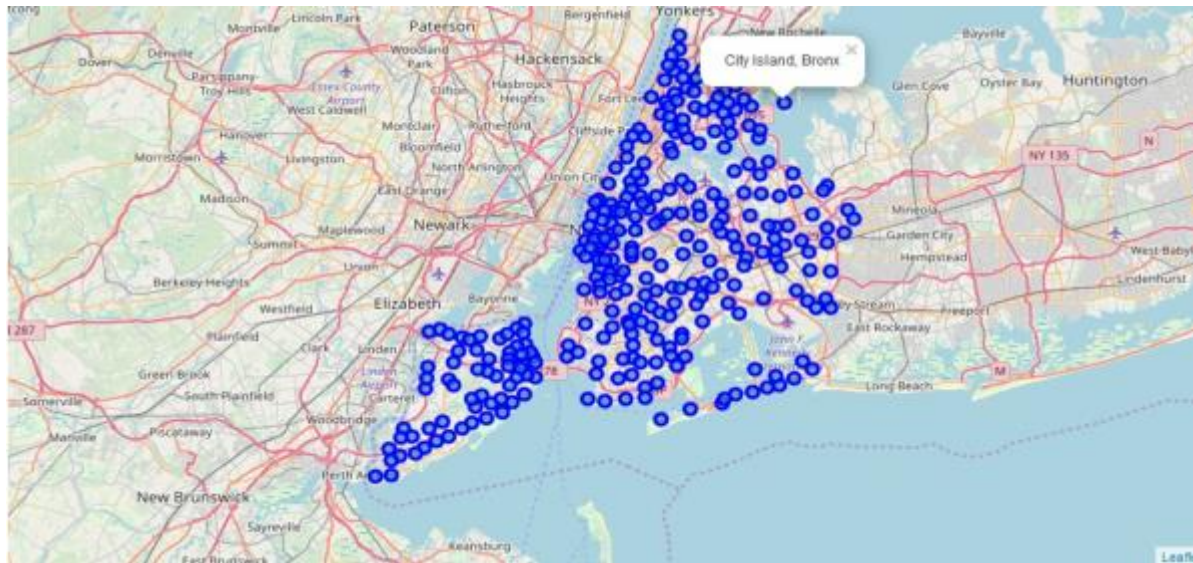
Data 4 : New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighborhood. We will use the Foursquare API to explore neighborhoods in New York City.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
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	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

Exploratory Data Analysis:

Data 1- New york city Geographical Coordinates Data.

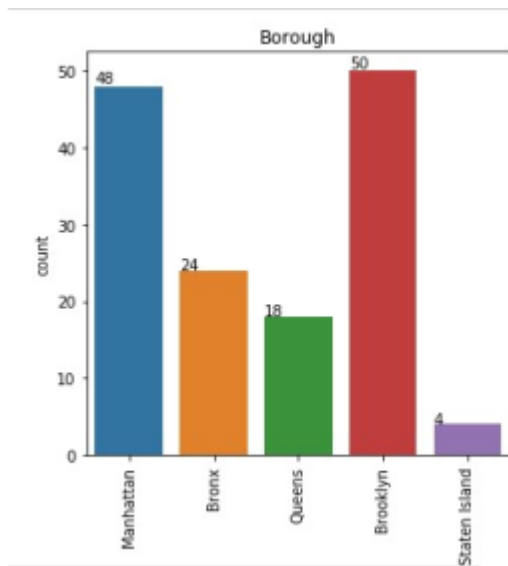
1. In this we load the data and explore data from newyork_data.json file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Foursquare.
5.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

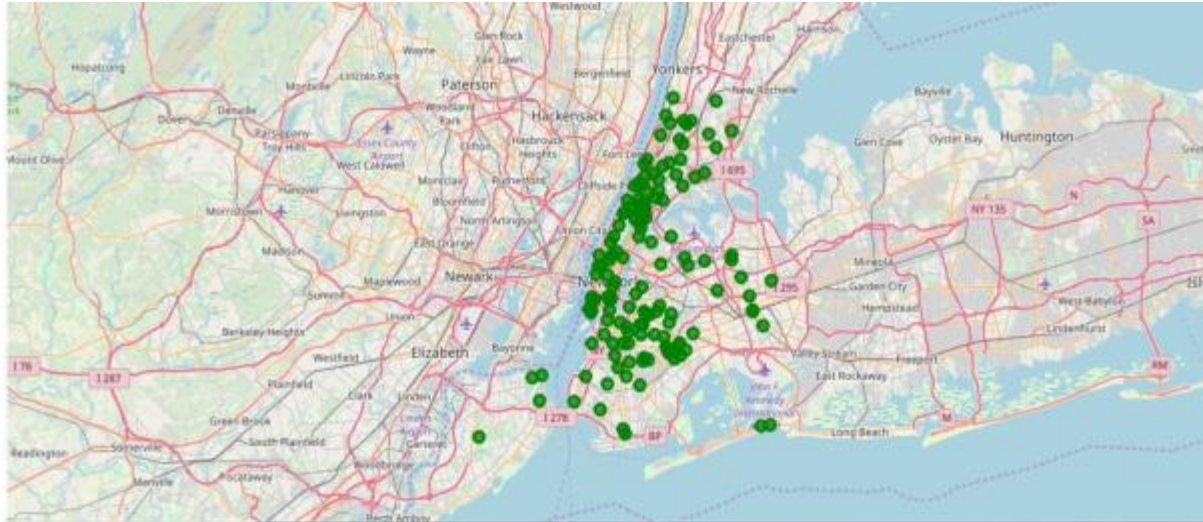


Data 2- Second data which is used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets data.

There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn.

And lowest in Queens, Bronx and Staten Island.





Data 3 : To analyze 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 : Insights from the data :

☐ 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 (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual American city.

☐ 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	square_miles	square_km	persons_sq_mi	persons_sq_km
0	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
1	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
3	Queens	Queens	2,358,582	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	479,458	58.37	151.18	8,112	3,132
5		City of New York	8,622,698	302.64	783.83	28,188	10,947
6		State of New York	19,849,399	47,214	122,284	416.4	159

2.New York City Demographics : New York City is the most populous city in the United States, [9] with an estimated record high of 8,622,698 residents as of 2017, [7] incorporating more immigration into the city than outmigration since the 2010 United States Census.

The racial composition is as given below. This is the reason New York city has restaurants serving cuisine from many countries such as Indian, African, Japan etc. This also increases the scope for 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%	–

3.Cuisine of New York city : This data has been manually prepared. Data is taken from Wikipedia page - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City . Using this data we did word cloud.

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

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

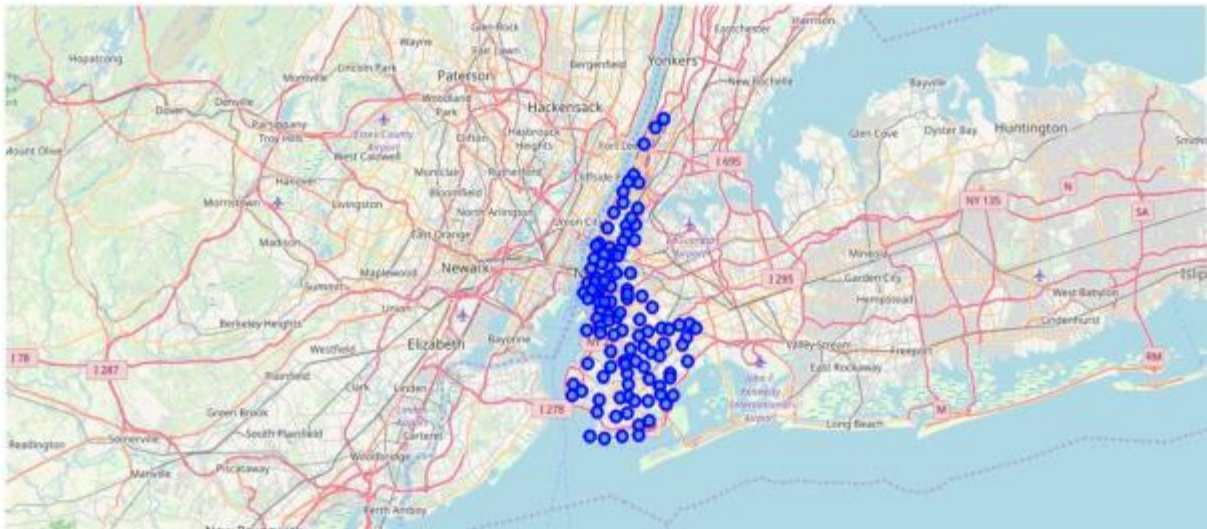
MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian .

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

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

There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

Data 4 : NewYork city geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.



Brooklyn and Manhattan Venues :

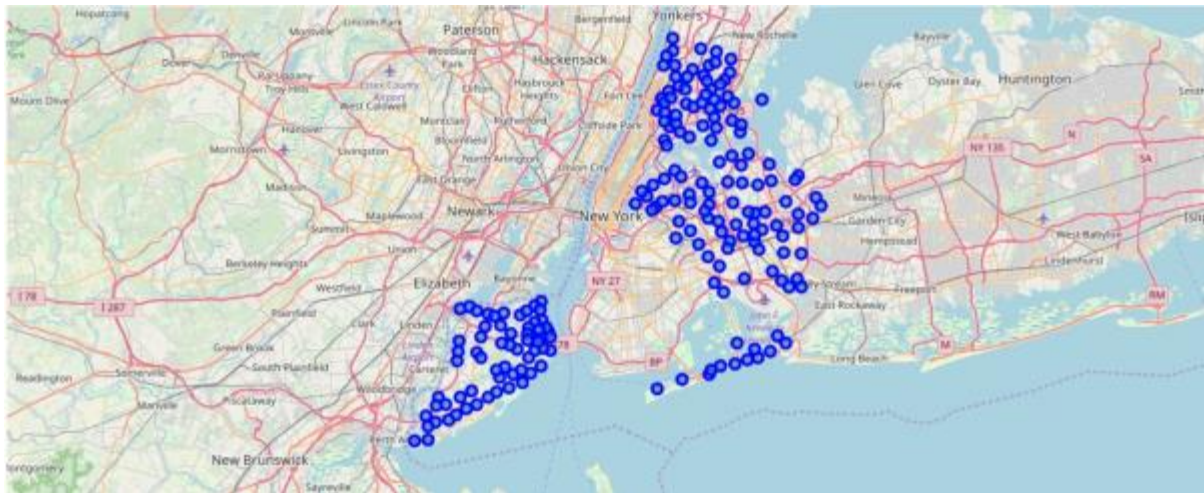
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett 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	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9708 venues and 397 unique venue types.



Bronx, Queens and Staten Island :

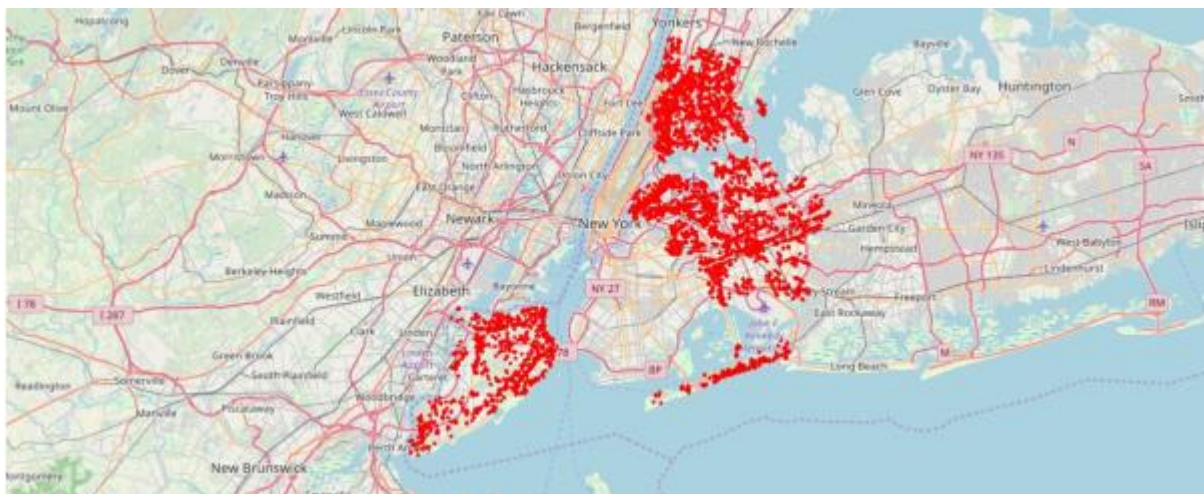
Bronx, Queens and Staten Island Neighborhoods Visualization :



Bronx, Queens and Staten Island Venues Visualization : The "BQS_venues" dataframe has 10805 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	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	40.896521	-73.844580	Pharmacy

Bronx, Queens and Staten Island Venues Map Visualization:



4.RESULTS :

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan

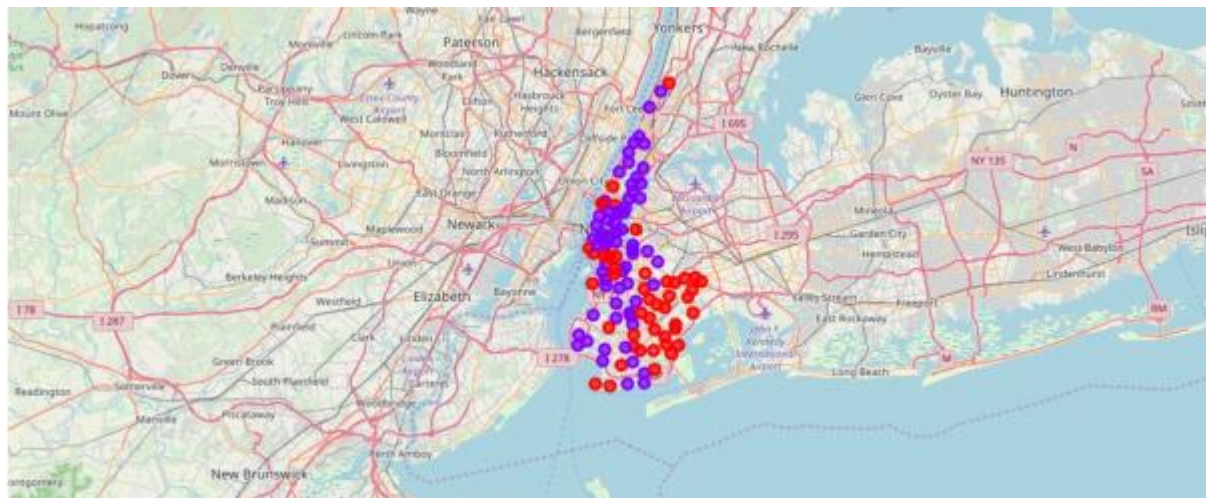
clustering and Bronx, Queens and Staten Island clustering. As we focussed only on restaurants business.

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

To cluster the neighborhoods into two clusters we used the K-Means clustering Algorithm. 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 below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.



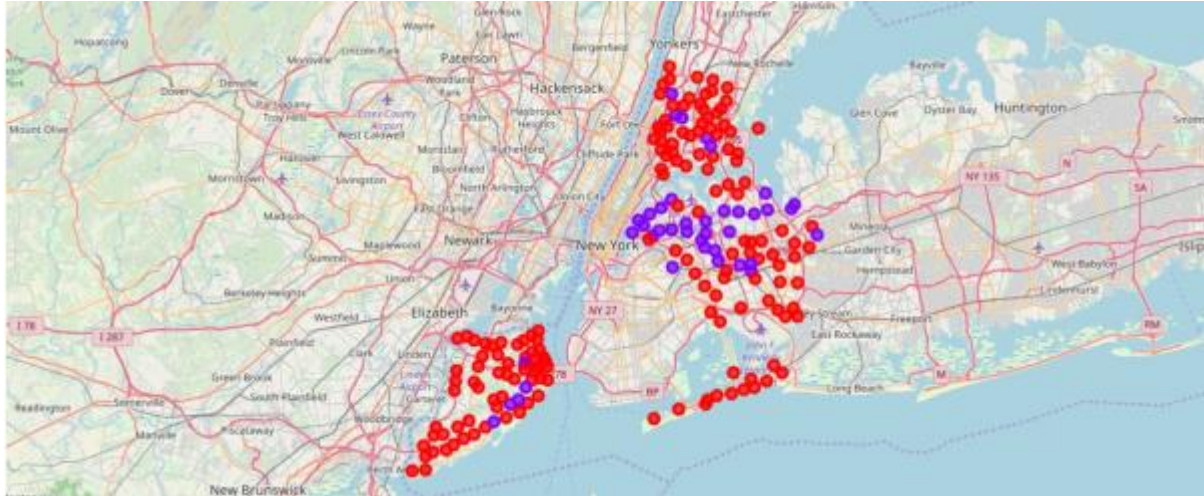
Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

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

There are no untapped neighborhoods 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 neighborhoods. 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	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. Number of restaurants are very high.

5.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 risk can be taken with great menu on board. It also shows people love eating cuisines of various countries.

6.CONCLUSION:

This analysis is performed on limited data. This may be right or may be wrong. But if good 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 has 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 neighborhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.