The Battle of the Neighborhoods- Report

Introduction & Business Problem:

Problem Background:

The City of New York, is the most populous 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. It is a global hub of business and commerce. The city is a major center for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. 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 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 very important to strategically plan. Various factors need to be studied in order to decide on the Location such as:

- New York Population,
- New York City Demographics,
- Are there any Farmers Markets nearby so that the ingredients can be purchased fresh to maintain quality and cost?
- Are there any venues like Gyms, Entertainment zones, Parks nearby where floating population is high?
- Who are the competitors in that location?

- Cuisine served and Menus of the competitors,
- Segmentation of the Borough,
- Untapped markets,
- Saturated markets, etc.

Even though, well-funded MNO Company Ltd. need 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, MNO Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood 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/Neighborhood choice to XYZ Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

Data:

One city will be analyzed in this project: New York City. We will be using the below datasets for analyzing New York city

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

2. Second data which will be used is the DOHMH Farmers Markets dataset. In this we will be using the data of Farmers Markets.
   https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2

	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	Woodhull Hospital Youthmarket	Broadway & Flushing Ave	40.700726	-73.941932	Wednesday	9 a.m. - 2 p.m.	07/10/2019-11/27/2019	Yes	No	No	No	(40.700726, -73.941932)
1	Manhattan	Mount Sinai Hospital Greenmarket	E 99th St bet Madison & Park Aves	40.789169	-73.952743	Wednesday	8 a.m. - 5 p.m.	06/12/19-11/27/19	Yes	No	No	No	(40.789169, -73.952743)
2	Bronx	170 Farm Stand	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30 - 6:30 p.m.	07/10/2019-11/27/2019	Yes	No	No	Yes	(40.839882, -73.916783)
3	Manhattan	Greenmarket at Oculus Plaza	Church & Fulton Sts, on Oculus Plaza	40.711535	-74.010464	Tuesday	7 a.m. - 7 p.m.	07/09/2019-11/30/19	Yes	Yes	No	No	(40.711535, -74.010464)
4	Queens	Ditmars Park Youthmarket	Steinway St bet Ditmars Blvd & 23rd Ave, at Di...	40.772854	-73.906061	Saturday	8 a.m. - 3 p.m.	07/13/2019-11/23/2019	Yes	No	No	No	(40.772854, -73.906061)

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/Cuisine_of_New_York_City
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. The below is image of the Foursquare API data.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
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

Methodology:

Business Understanding:

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

Analytic Approach:

New York city's neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn. And second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

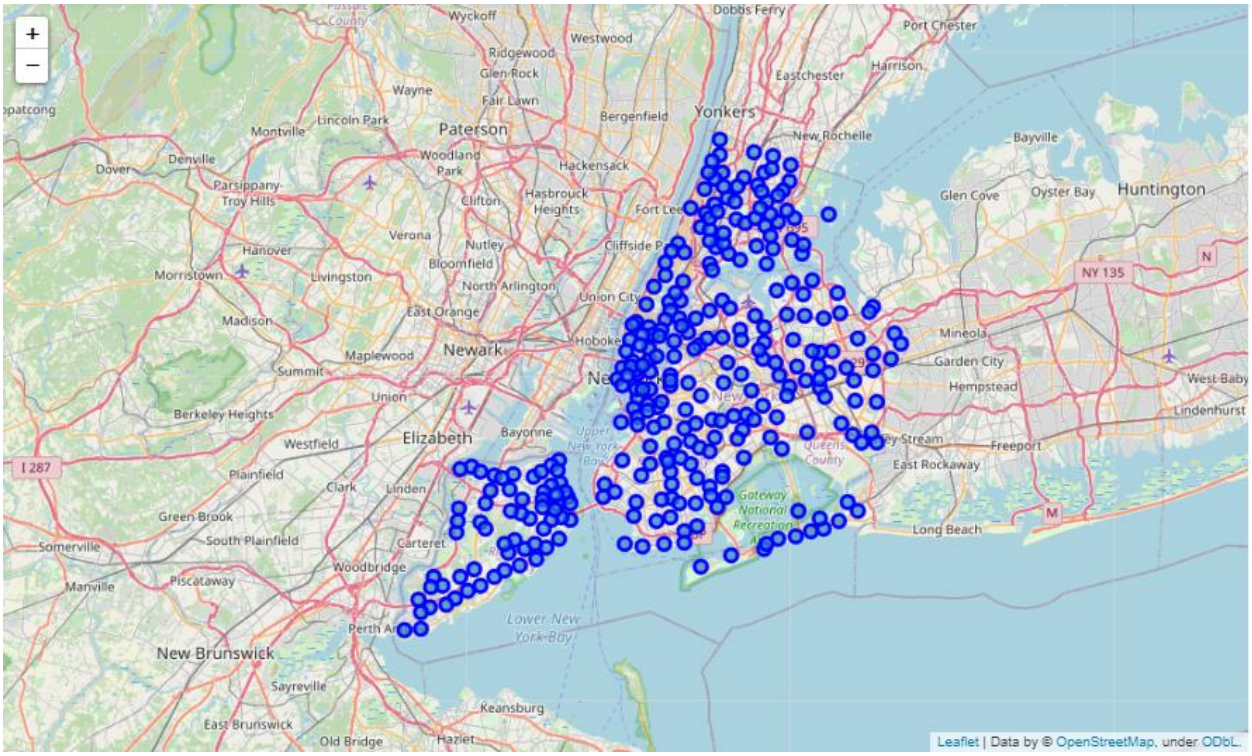
Exploratory Data Analysis:

Data 1:

New York city Geographical Coordinates Data.

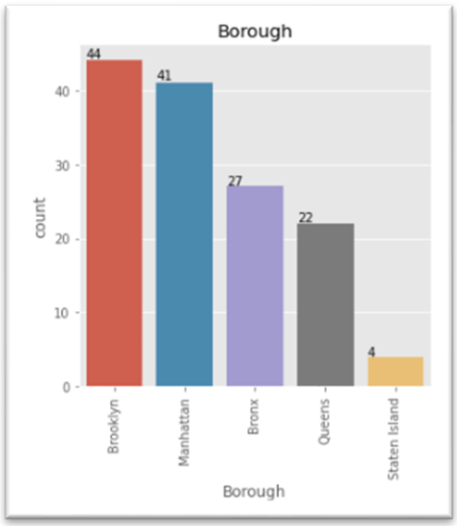
- In this we load the data and explore data from newyork_data.json file.
- Transform the data of nested python dictionaries into a pandas dataframe.
- This dataframe contains the geographical coordinates of New York city neighborhoods.
- This data will used to get Venues data from Fouresquare.
- We used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

New York neighborhood visualization



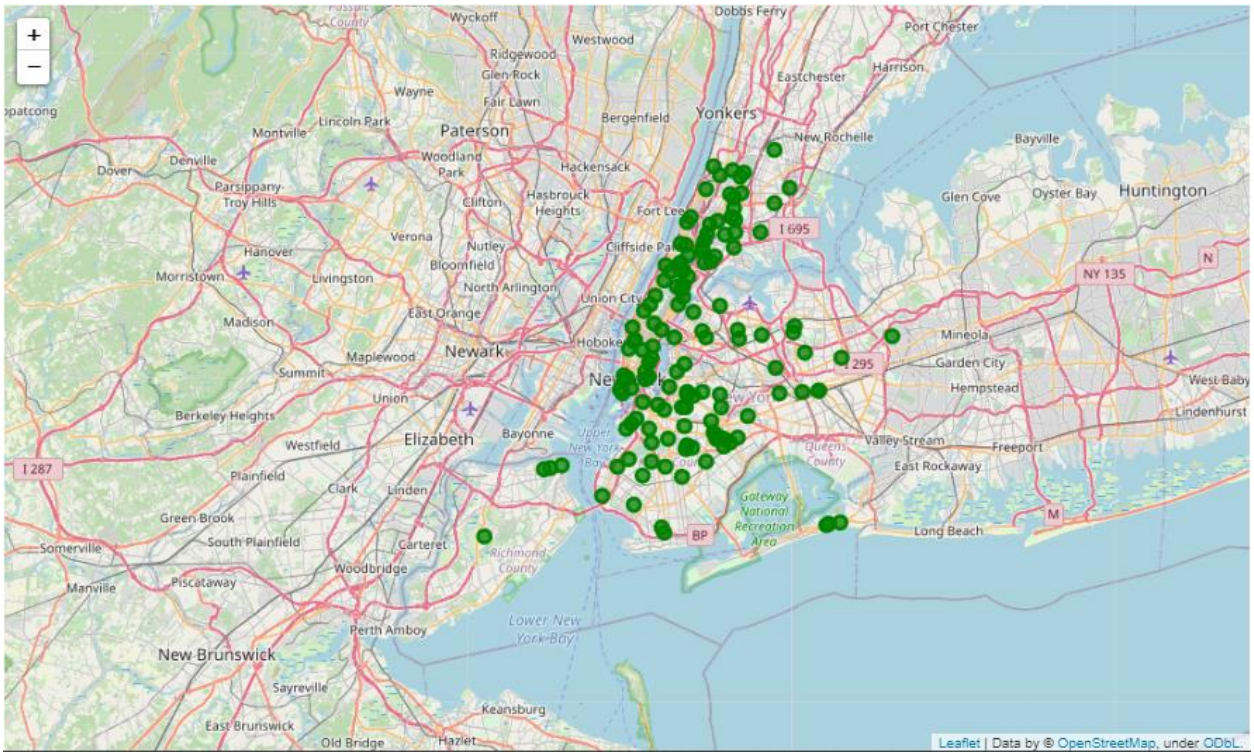
Data 2:

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 138 Farmers Markets in New York city. Highest number are in Manhattan-41 and Brooklyn-44. And lowest in Queens-22, Bronx-27 and Staten Island-4. The proof of this is as given below.



We used geopy and folium libraries to create a map to visualize farmers markets of New York city.

Farmers Market visualization-New York City



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. Beautiful Soup 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 71,341 people per square mile (27,544/km²) in 2019 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_2019	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,418,207	42.695	109.04	33,867	13,006
1	Brooklyn	Kings	2,559,903	91.559	183.42	36,147	13,957
2	Manhattan	New York	1,628,706	600.244	59.13	71,341	27,544
3	Queens	Queens	2,253,858	93.310	281.09	20,767	8,018
4	Staten Island	Richmond	476,143	14.514	151.18	8,157	3,150
5		City of New York	8,336,817	842.343	101,000	27,547	10,636
6		State of New York	19,453,561	1,731.910	89,000	412	159

2.New York City Demographics:

New York City is the most populous city in the United States, with an estimated record high of 8,336,817 residents as of 2019, 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	44.00%	52.30%	76.60%	93.60%
1	Non-Hispanic	33.30%	43.20%	62.90%	92.00%
2	Black or African American	25.50%	28.70%	21.10%	6.10%
3	Hispanic or Latino (of any race)	28.60%	24.40%	16.20%	1.60%
4	Asian	12.70%	7.00%	1.20%	NaN

3.Cuisine of New York city:

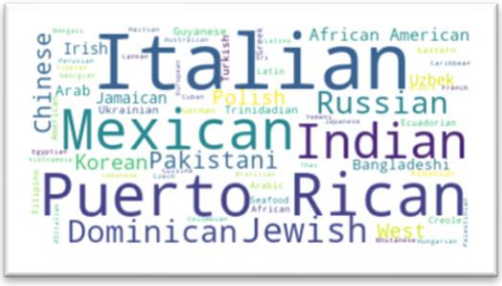
This data has been manually prepared.

Data is taken from Wikipedia page –

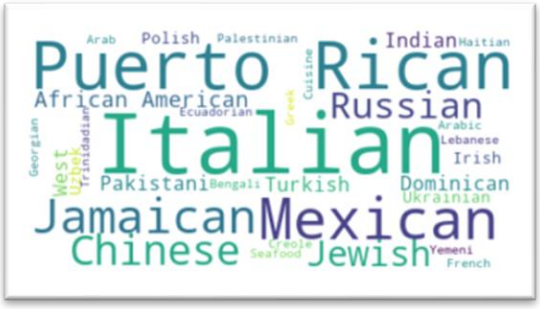
https://en.wikipedia.org/wiki/Cuisine_of_New_York_City .

Using this data, we did word cloud.

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



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



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

QUEENS CUISINE - Most Preferred Food in Queens is – Indian, Italian, Jewish.

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

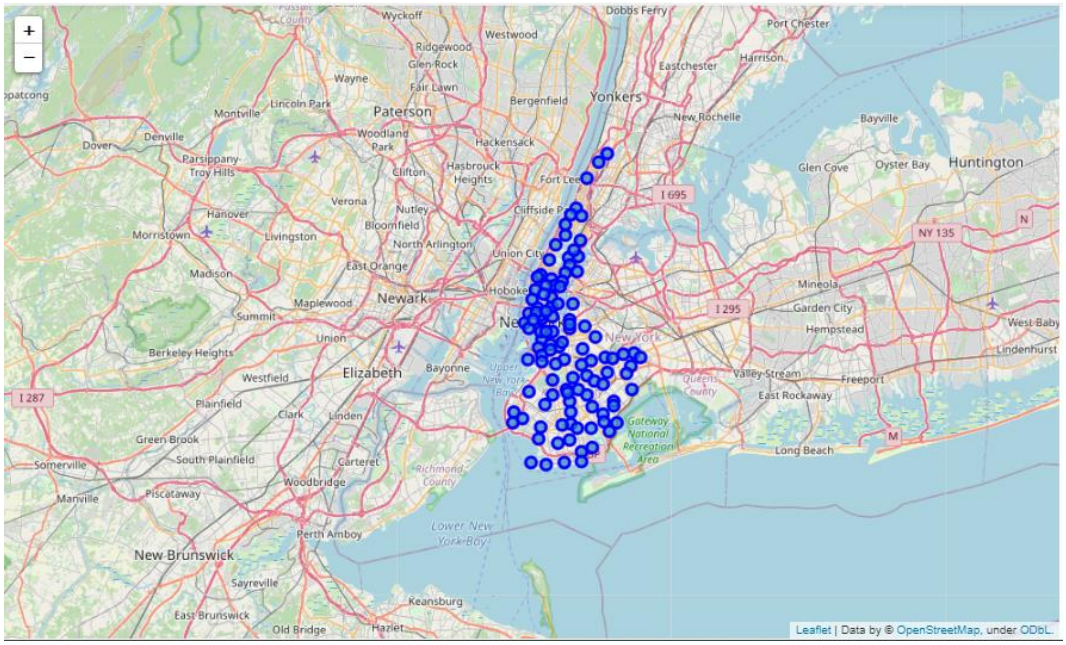
STATEN ISLAND CUSINE - Most Preferred Food in The Bronx is – Italian, Arab, Polish, Russian.

Data 4:

New York city geographical coordinates data has been 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:

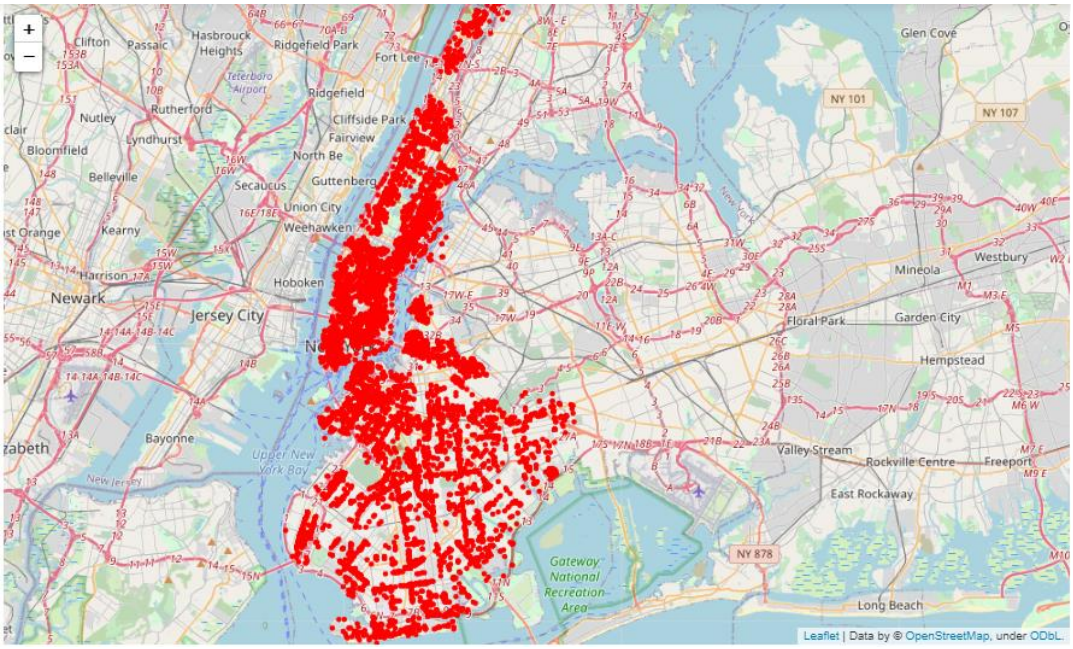
Brooklyn and Manhattan Visualization:



Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

Brooklyn and Manhattan Venues:

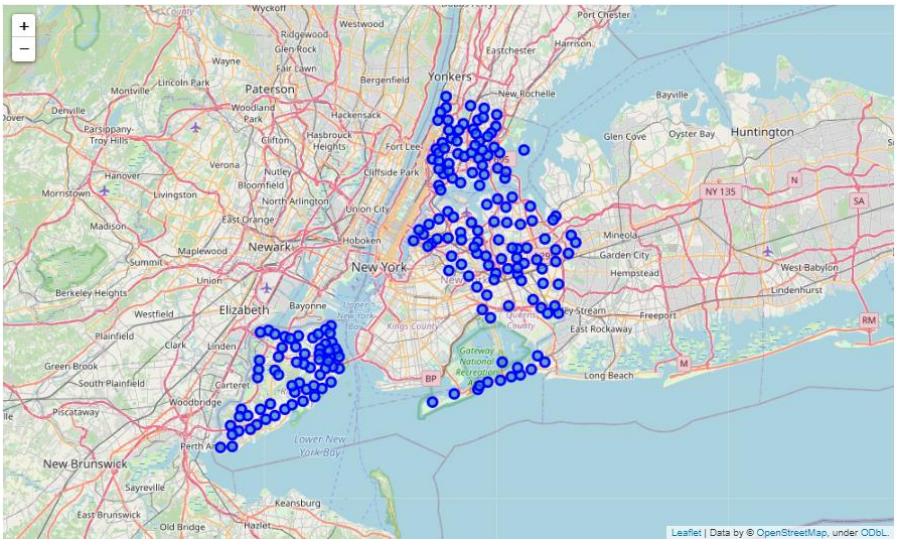
Brooklyn and Manhattan Venues Visualization:



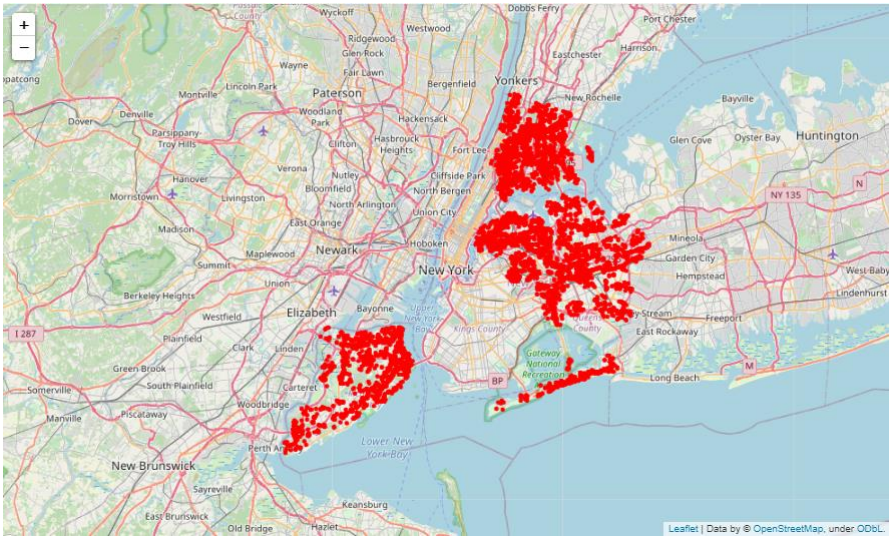
Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9627 venues and 407 unique venue types.

Bronx, Queens and Staten Island:

Bronx, Queens and Staten Island Neighborhoods Visualization:



Bronx, Queens and Staten Island Venues Visualization:



The "BQS_venues" dataframe has 11003 venues and 388 unique venue types.

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 focused 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.

1. 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 a very few untapped neighborhoods in Brooklyn and Manhattan.

2. 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.
   - Cluster1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

DISCUSSION:

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

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 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 neighborhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.