THE BATTLE OF NEIGHBORHOODS

CASE STUDY – RESTAURANTS OF NEW YORK

Report by Pranjal Shekhawat

Index

1. Introduction

1.1 Background

The **City of New York** is the most populous city in the United States. New York is the most densely populated major city in the United States. 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. 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. 800 languages are spoken in New York, making it the most linguistically diverse city in the world. New York City is home to more than 3.2 million residents born outside the United States, the largest foreign-born population of any city in the world.

As on July 11, 2019 the population of New York is 19.49 million. The more the population the more is the requirement of food hence restaurants. But at the same time the competitiveness also increases

With all this background about New York, we can infer that the potential of any business is New York is immense. **Here case study will focus on Restaurants in New York**.

1.2 Business Problem

Restaurant is a place where people pay to sit and eat meals that are cooked and served on the premises. The bill can be pre-paid or post-paid. New York is a hub of all sorts of international cuisines because of the diverse population & immigrants in this city. The cuisine broadly consists of Greek, Italian, Eastern European, Brazilian, Egyptian, Arabic, Indian, Pakistani, Chinese, Korean, Jewish, Russian, Uzbek, Jamaican, West Indian, Irish, Puerto Rican, Dominican, Bengali, German and Polish, Bangladeshi, Colombian, Ecuadorian, Peruvian, Korean, Filipino , Mexican, Sea food, African-American, Jamaican, Haitian, Creole, Polish, Ukrainian, mobile food vendors etc.
Finding an optimum location for set up of a New Indian Restaurant depending on the following two factors: -

1. Cuisine Preference of the people residing in that Borough
2. Saturation rate of restaurants at the location

Now while analysing the above problem we need to take into consideration the following also -

- Population of New York City
- Demographics of New York City
- Farmers Markets, Wholesale markets etc. nearby so that the ingredients can be purchased fresh to maintain quality and cost?
- Are there any Shopping places, Gyms, Entertainment zones, Parks etc. nearby where floating population is high etc.
- Who are the competitors in that location?
- Cuisine served / Menu of the competitors
- Segmentation of the Borough
- Untapped markets
- Saturated markets etc.

1.3 Interest

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

2 Data Acquisition & Cleaning

Data 1

New York City Neighborhood Names

NY Spatial Data - https://geo.nyu.edu/catalog/nyu_2451_34572

Description: Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement 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 logitude coordinates of each neighborhood.
Publisher: New York (City). Department of City Planning

GeoJSON file is readily available it is to be converted to .**csv** & used.

	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

Farmer Markets - A farmers' market is a physical retail marketplace intended to sell foods directly by farmers to consumers. Farmers' markets may be indoors or outdoors and typically consisting of booths, tables or stands where farmers sell their homegrown produce, live animals and plants, and sometimes prepared foods and beverages.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2/data

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

	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	Isham 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

Food Boxes - 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.

https://www.grownyc.org/greenmarketco/foodbox

Data 3

- 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

	NewYorkCity'sfiveboroughsvte	Jurisdiction	Population	GrossDomesticProduct	Landarea	Density	Borough	County	Estimate(2017)[207]
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

Data 4

NewYork 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	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

2.2 Data Cleaning

Data downloaded or scraped from multiple sources were combined into one table. There were a lot of missing values, for some data it was removed and for the other it was substituted with NAN for calculation purpose. Some features from the data were removed because they were not required for the analysis. Some data was imported into the excel from Wikipedia and then formatted properly and converted to csv.

Analytic Approach:

New York city 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

2.1 Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modeling or hypothesis testing task.

2.2 Clustering

Clustering is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. In simple words, the aim is to segregate groups with similar traits and assign them into clusters

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 Fouresquare.
5. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.

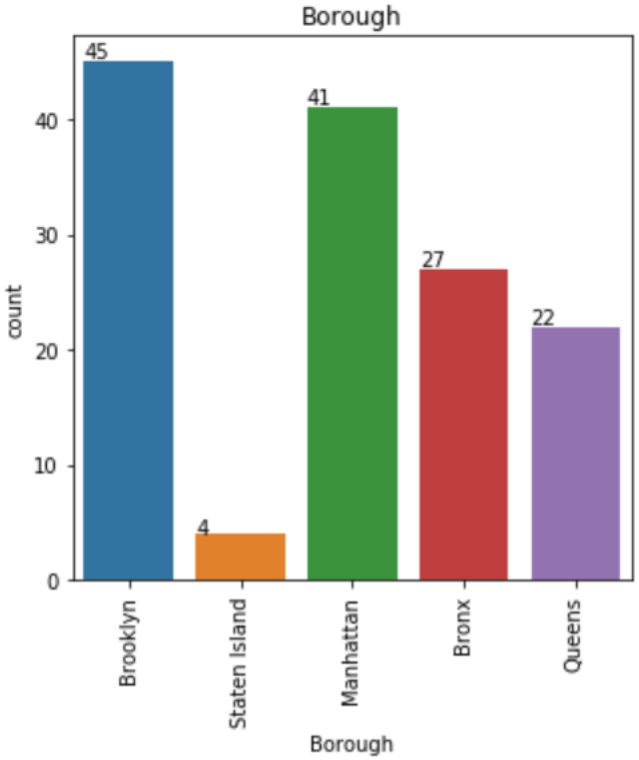
4 New York Neighborhood Visualization in 5 Boroughs

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.

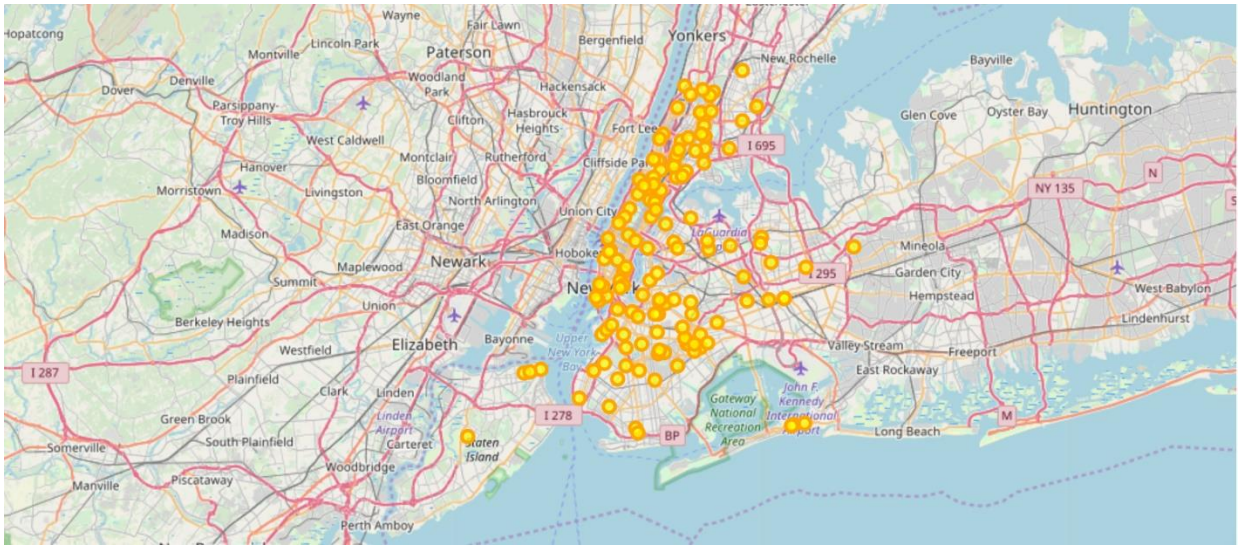
5 Farmer Markets

There are totally 139 Farmers Markets in New York City. Using geopy and folium libraries we draw a map to visualize farmers markets of the New York City.
From the bar chart, we can see the order is as follows:

- Brooklyn - 45
- Manhattan - 41
- Queens - 27
- Bronx - 22
- Staten Island - 4



5.1 Farmers Market Visualization - New York City



Data 3: To analyze New York City Population, Demographics and Cuisine, the data from Wikipedia pages given above in the data section were used. BeautifulSoup python library is used. 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 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. Cuisine of New York City: The data was extracted from Wikipedia & manually cleaned and brought in proper format for analysis.

Wikipedia page - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City . Using this data, word cloud was generated.

Types of Cuisine in Borough

Out[16]:

	Borough
Brooklyn	19
Queens	17
Manhattan	12
Bronx	9
Staten Island	5

6 New York City Cuisine

Most Preferred Cuisine in New York City are

- Italian
- Puerto Rican
- Mexican
- Jewish
- Russian
- Dominican
- Indian
- Chinese
- Pakistani

6.1 Brooklyn Cuisine

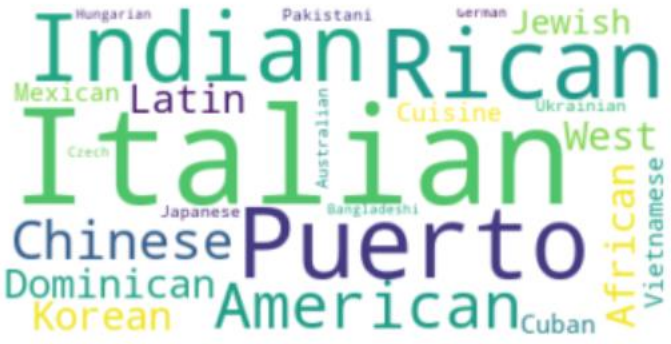
Most Preferred Cuisine in Brooklyn is –

- Italian
- Puerto Rican
- Mexican
- Jewish
- Jamaican
- Russian

6.2 Manhattan Cuisine

Most Preferred Cuisine in Manhattan

- Italian
- Puerto Rican
- Indian
- American
- Chinese



6.3 Queens Cuisine

Most Preferred Cuisine in Queens is –

- Indian
- Italian

- Mexican
- Jewish
- Irish
- Uzbek
- Pakistani

6.4 Bronx Cuisine

Most Preferred Cuisine in Bronx

- Italian
- Puerto Rican
- Albanian
- Dominican

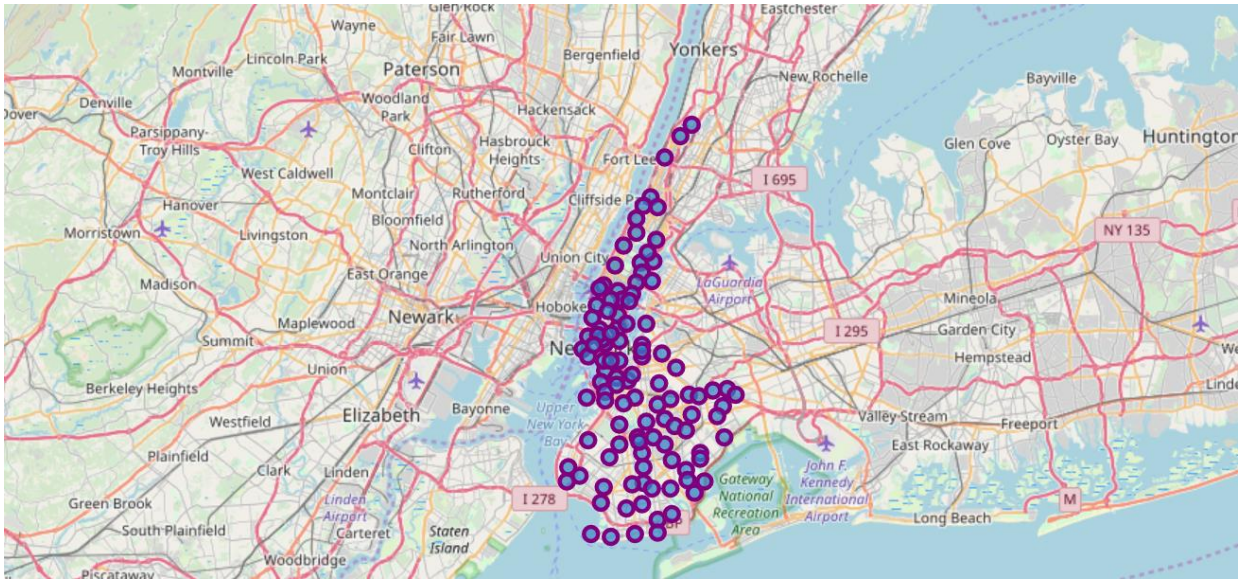
6.5 Staten Island Cuisine

Most preferred cuisine in Staten Island

- Italian
- Arab
- Polish
- Russian
- Indian

Data 4: New York City geographical coordinates data has to 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.

6.6 Brooklyn and Manhattan Visualization:

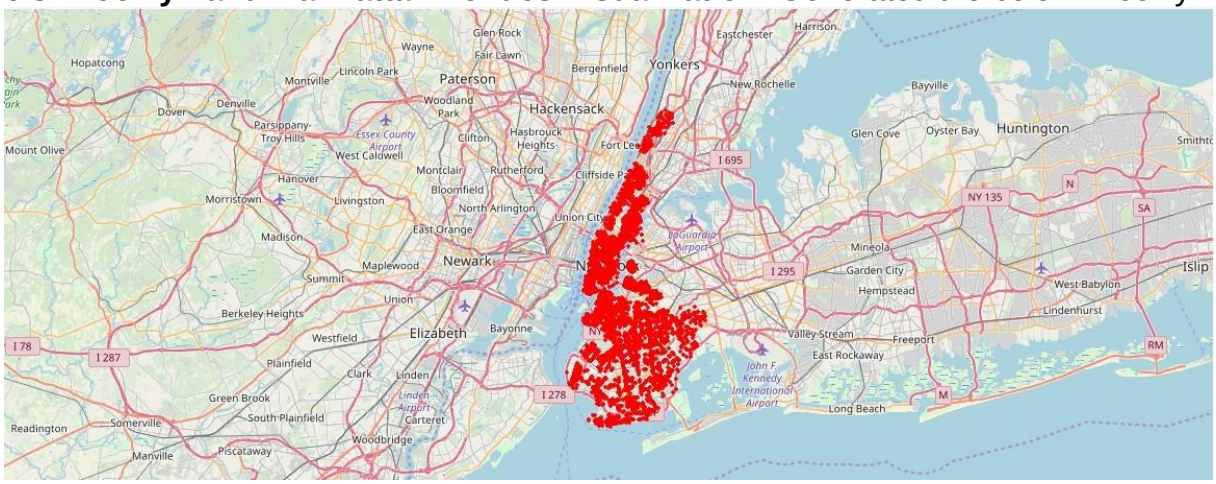


Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below:

6.7 Brooklyn and Manhattan Venues:

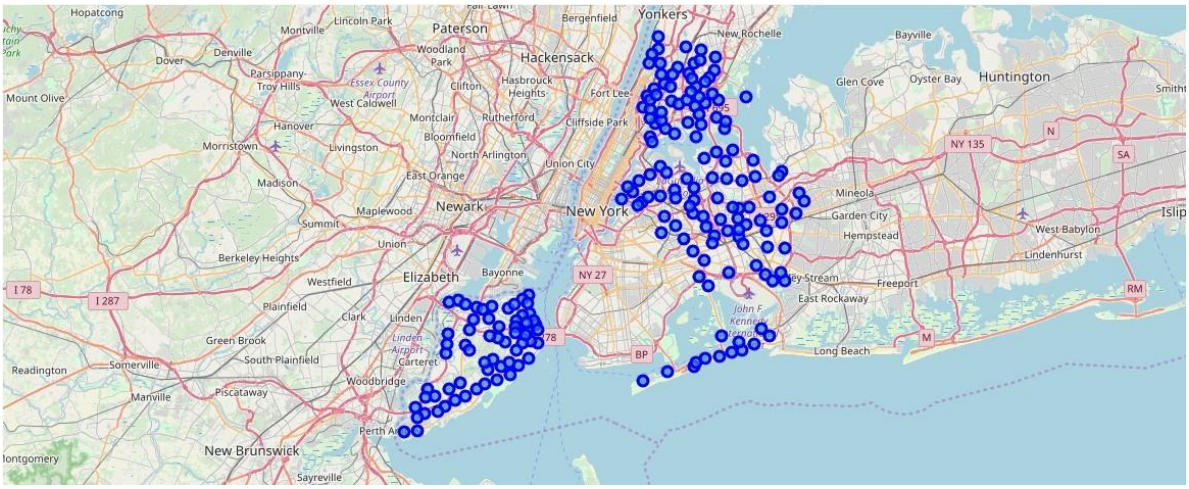
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett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

6.8 Brooklyn and Manhattan Venues Visualization: Generated the below Brookly



n and Manhattan Venues Visualization. The "BM_venues" dataframe has 9747 venu es and 402 unique venue type

6.9 Bronx, Queens and Staten Island Neighborhoods Visualization :



6.10 Bronx, Queens and Staten Island Venues Visualization: The "BQS_venues" dataframe has 11131 venues and 395 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	-73.844680	Pharmacy

7 Result :

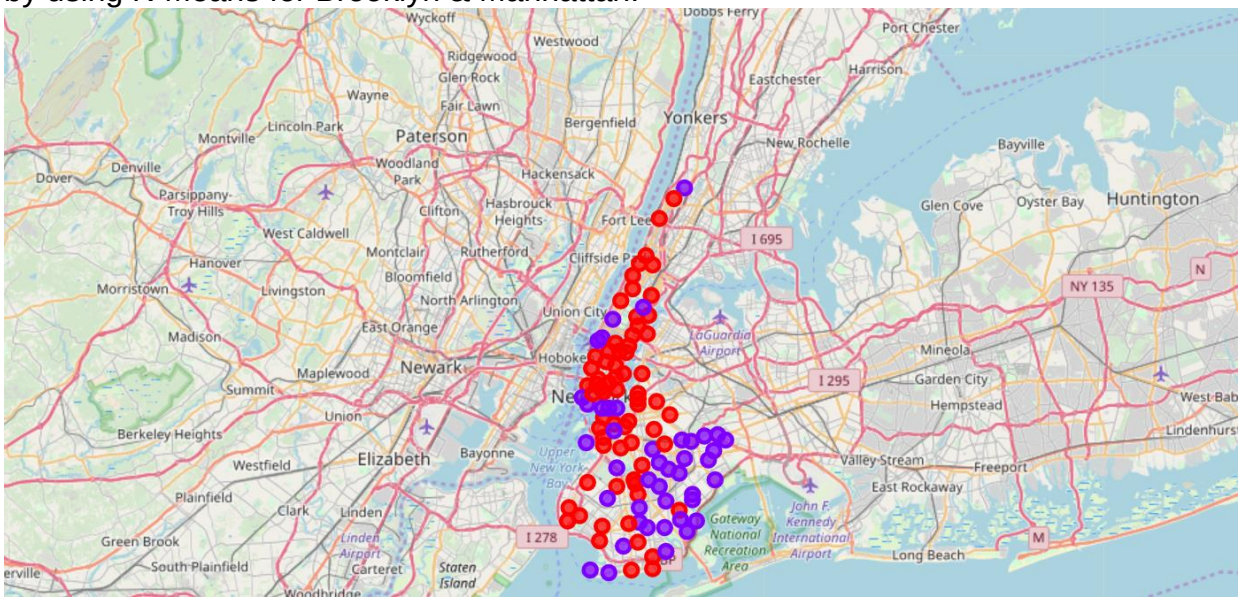
From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we are focusing only on restaurants business.

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.

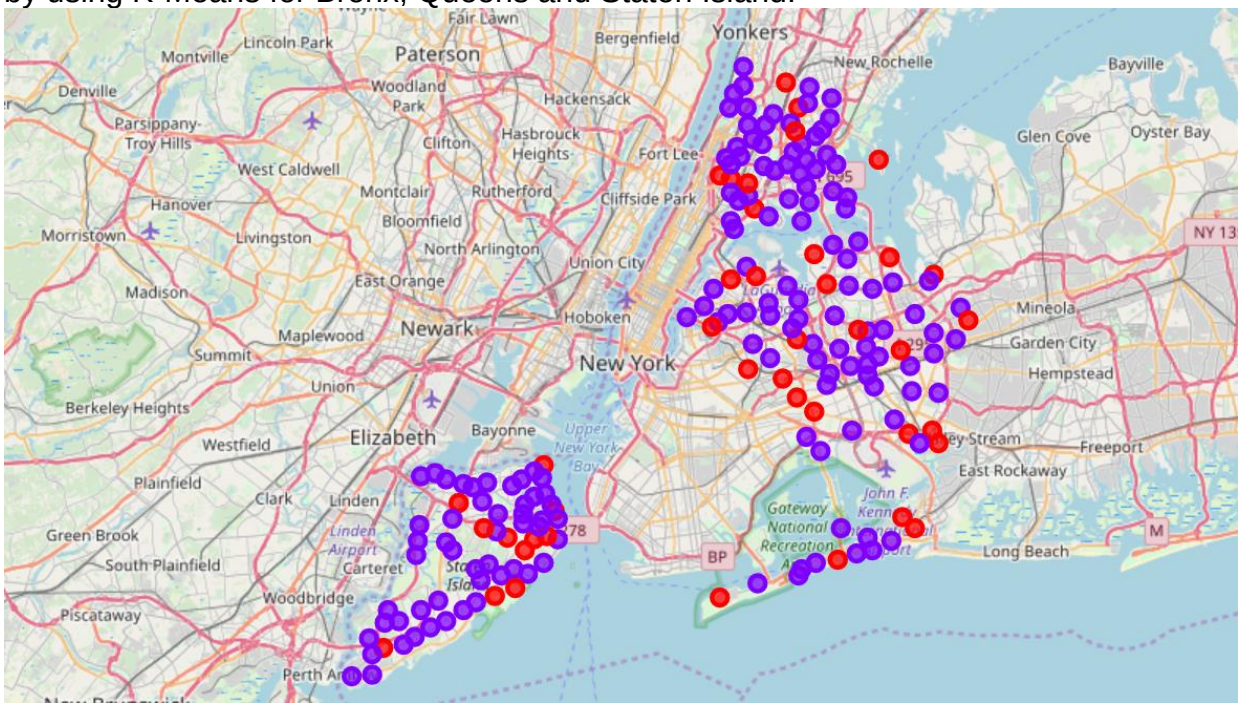


Cluster0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
Cluster1: The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Manhattan	Marble Hill	40.876551	-73.910660	12	0
1	Brooklyn	Gravesend	40.595260	-73.973471	6	0
2	Brooklyn	Manhattan Terrace	40.614433	-73.957438	20	0
3	Brooklyn	Crown Heights	40.670829	-73.943291	20	0
4	Brooklyn	East Flatbush	40.641718	-73.936103	8	0
5	Brooklyn	Brownsville	40.663950	-73.910235	13	0
6	Brooklyn	Red Hook	40.676253	-74.012759	11	0
7	Brooklyn	Cypress Hills	40.682391	-73.876616	12	0
8	Brooklyn	East New York	40.669926	-73.880699	5	0
9	Brooklyn	Starrett City	40.647589	-73.879370	8	0
10	Brooklyn	Canarsie	40.635564	-73.902093	9	0
11	Brooklyn	Flatlands	40.630446	-73.929113	10	0
12	Brooklyn	Mill Island	40.606336	-73.908186	2	0
13	Brooklyn	Coney Island	40.574293	-73.988683	9	0
14	Brooklyn	Borough Park	40.633131	-73.990498	6	0

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.
The entire list of 157 neighborhoods is generated where no of total varies from min 2 to max 19. Sample of the same is shown below:-

Though there are many potential areas, the three Neighborhoods with lowest no of restaurants are mentioned below -

S.No	Neighborhood	Borough	No of Restaurants
1	Todt Hill	Staten Island	4
2	Port Ivory	Staten Island	7
3	Breezy Point	Queens	9

Etc.

	Neighborhood	Total	Cluster_Labels
0	Allerton	13	1
1	Annadale	4	1
2	Arden Heights	3	1
3	Arlington	2	1
4	Arrochar	4	1
5	Arverne	2	1
6	Astoria Heights	11	1
7	Bay Terrace	18	1
8	Baychester	15	1
9	Bayswater	1	1
10	Bedford Park	9	1
11	Beechhurst	11	1

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

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

9. Conclusion:

- The analysis is performed on limited data which is freely available online. If large data is available there is scope to come up with better and finer results.
- Brooklyn and Manhattan have higher density of restaurant business & more competitive markets compared to Bronx, Queens and Staten Island.

So, this can be explored.

As mentioned, while describing the problem to find a suitable place for an up coming up with an Indian restaurant, we will leave the stake owners with the following inputs

S.No	Borough	Inclination towards Indian Cuisine	Competition in general
1.	Staten Island	medium	low
2.	Queens	high	Medium
3.	Bronx	low	low
4.	Brooklyn	low	high
5.	Manhattan	high	high

10 References

https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets/8vwk-6iz2/data
https://en.wikipedia.org/wiki/Cuisine_of_New_York_City
https://www.grownyc.org/greenmarketco/foodbox
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://geo.nyu.edu/catalog/nyu_2451_34572
www.coursera.org
https://github.com/Applied-Data-Science-Capstone

Special Thanks to all the professors (Instructors- Course 1 to Course 9), Mentors & Peers including the one who reviews this final assignment, for constant support, motivation & guidance.
Learned excellent & updated stuff and also made some friends.