



Battle of Neighbourhoods

Coursera IBM Data Science Capstone

Introduction

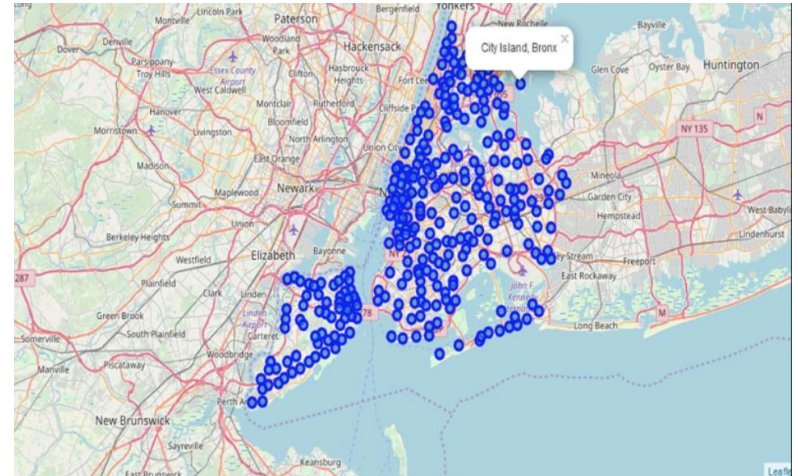
- New York review for ABC Company to set up a restaurant
- Need for an appropriate location

Business Problem

- Selection of neighbourhood
- Low competition
- High demand
- Choice of menu/cuisine served

Success criteria

- Location/neighborhood meeting the above criteria





Facts of New York

- Most populated city in the US
- Diverse in nature and also the financial capital
- Multicultural environment offering several diverse business opportunities
- It is a major hub for businesses in the Finance, retail, fashion, culinary and arts industry



Business Setup

- Highly competitive
- Extremely costly city which makes it expensive to start businesses
- One should strategically targeting the market in order to – -This will help in reduction of risk. -The Return on Investment will be reasonable.
- New business venture or expansion needs to be analysed carefully



Cuisine

- 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
 - Famous 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.



Factors to be studied before choosing location

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



Data description 1

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 description 2

Data 2 : Farmers Markets data - •

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-andFood-Boxes/8vwk-6iz2>

- 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	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

Data description 3

Data 3 : Data from Wikipedia pages 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

	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%	–



Data description 4

- Data from foursquare.com

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	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



Analytic Approach

New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods

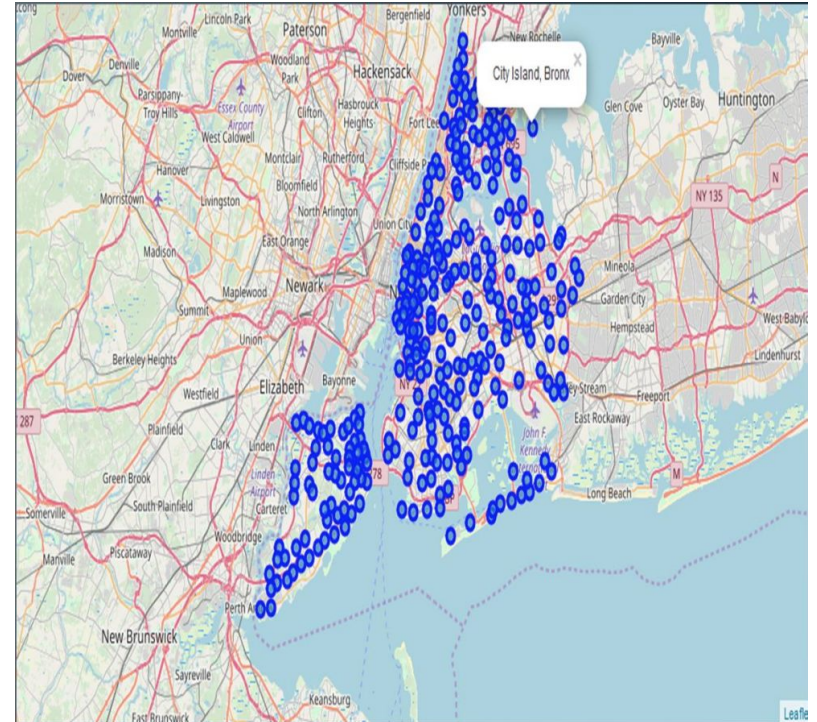
- PART 1 - Clustering of Manhattan and Brooklyn
- PART 2 - Clustering of Bronx, Queens and Staten Island.
- Only restaurant data is filtered from foursquare.com venues data and utilized for this project.
- This is done because of the following Exploratory data analysis.

Methodology 1

Data 1- New York city

Geographical Coordinates Data.

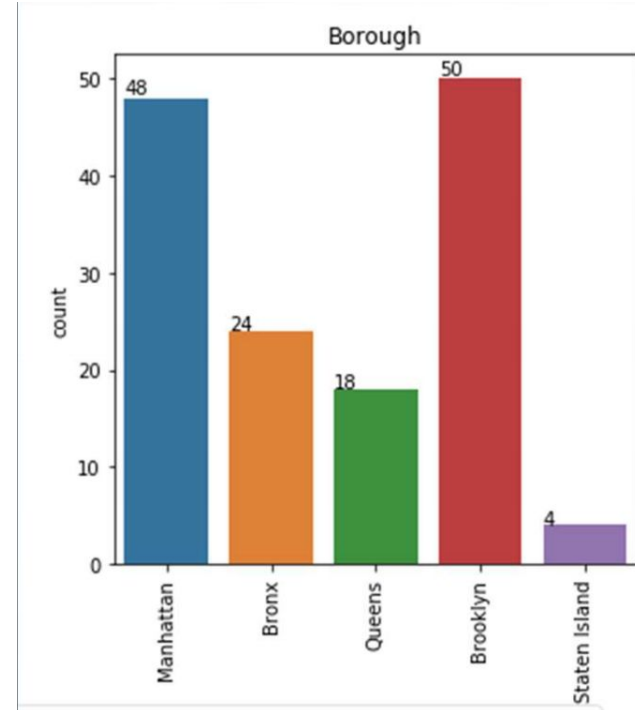
- Load the data and explore data - newyork_data.json
- Transform the data of nested python dictionaries into a pandas dataframe.
- Dataframe contains the geographical coordinates of New York city neighborhoods.
- Data will be used to get Venues data from Foursquare.
- Geopy and folium libraries used to create a map of New York city with neighborhoods superimposed on top



Methodology 2

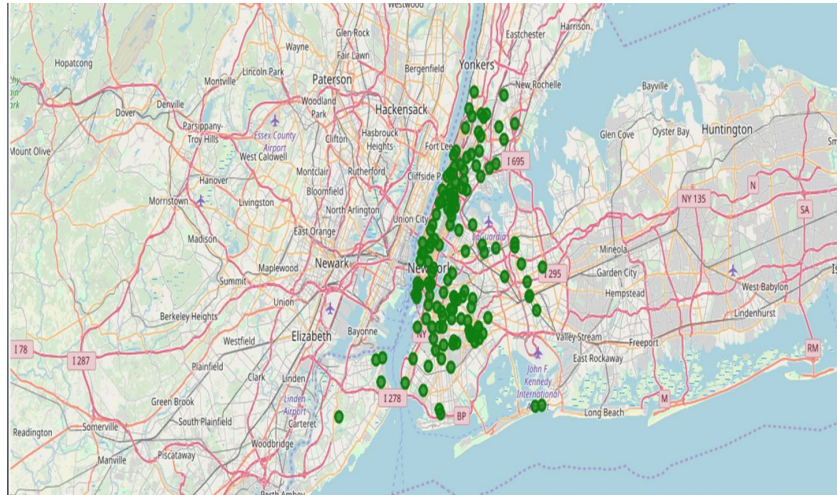
Data 2- DOHMH Farmers Markets and Food Boxes dataset.

- In this we will be using the data of Farmers Markets
- 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.



Methodology 2

Farmers Market visualisation-New York City





Methodology 3

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




Methodology 3 -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 (27,812/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



Methodology 3 - 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,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 India, Africa, 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%	—

Cuisine

This data has been manually prepared. Data is taken from Wikipedia page -

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

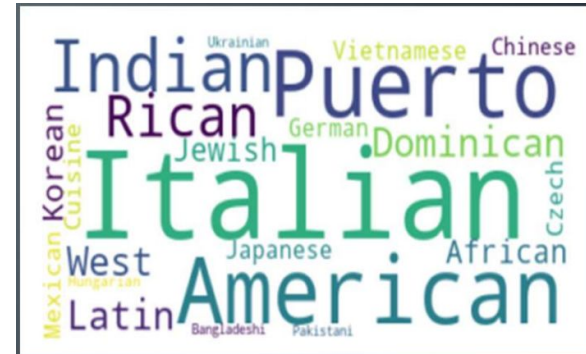
- Data used to create word cloud. NEW YORK CITY CUISINE : Most Preferred Food in New York City –Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani & Dominican



Cuisine

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

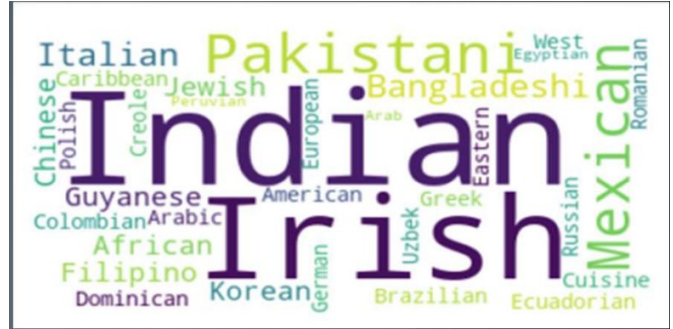
- MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.



Cuisine

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.





Methodology 4 - Foursquare

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

PART – 1 Brooklyn and Manhattan

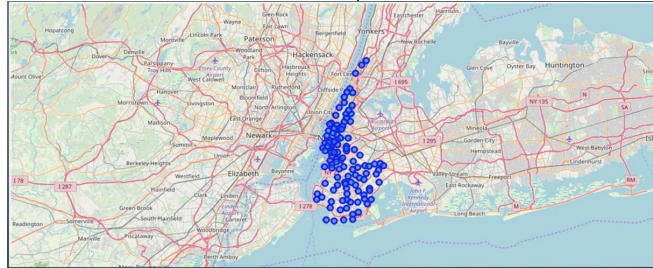
PART – 2 Bronx, Queens and Staten Island

Method 4- Part 1 Brooklyn & Manhattan

Brooklyn and Manhattan Venues 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

Brooklyn and Manhattan Visualization

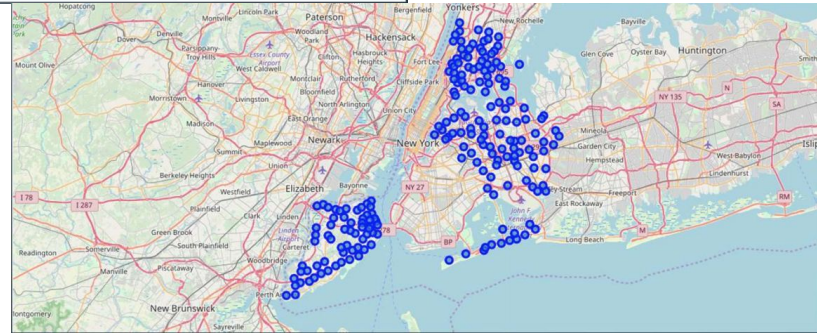


Method 4- Part 2 (Bronx, Queens & Staten Islands)

Bronx, Queens and Staten Island Venues data

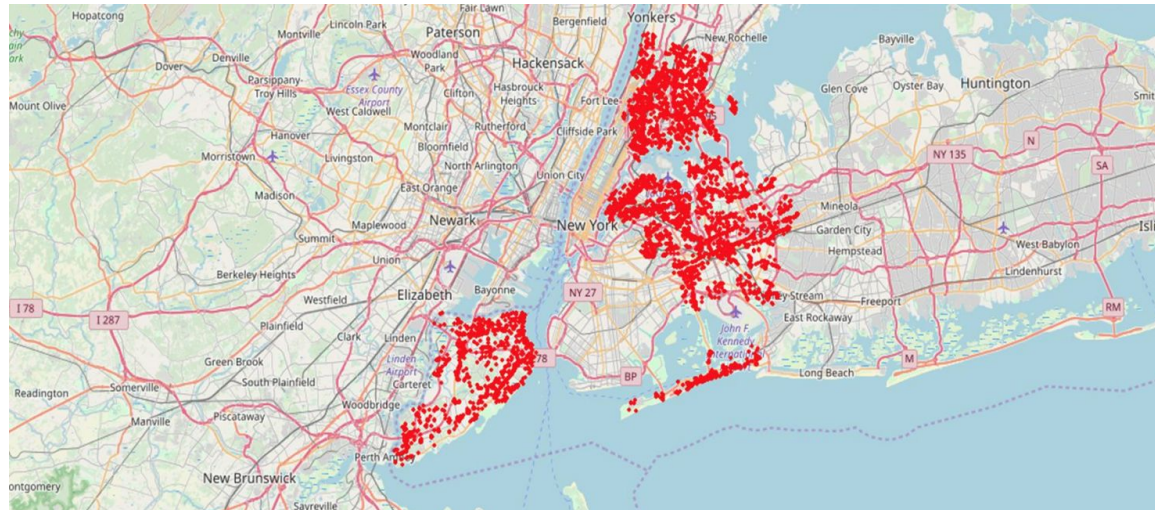
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-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

Bronx, Queens and Staten Island Visualization



Method 4- Part 2 (Bronx, Queens & Staten Islands)

Bronx, Queens and Staten Island Venues Visualization : The "BQS_venues" dataframe has 10805 venues and 387 unique venue types.



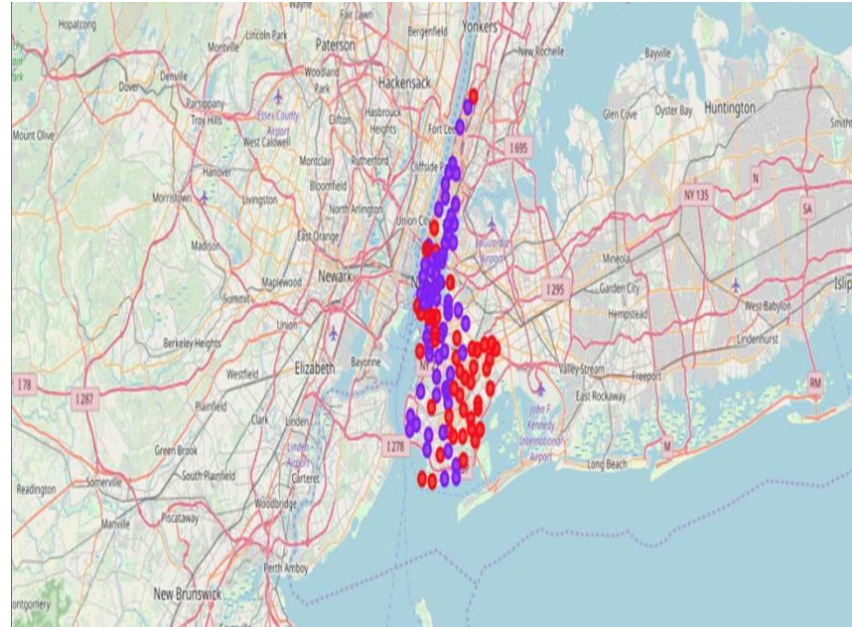
Results - Part 1 (Brooklyn & Manhattan)

Segmenting and Clustering Neighborhoods

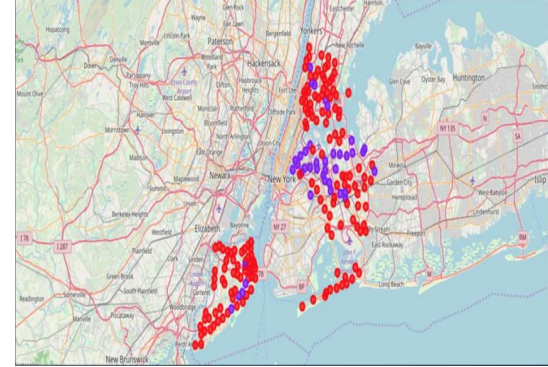
-

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 is very high. Lot of competition. Saturated neighborhoods.



Results - Part 2 (Bronx, Queens & Staten Island)



Segmentation and Clustering Neighborhoods -

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

Cluster1 : The Total and Total Sum is very high. Lot of competition. Saturated neighborhoods.

	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



Discussion

Scope to explore cuisines of various countries in Bronx, Queens and Staten Island

- In Manhattan and Brooklyn restaurants , cuisines of many countries are part of their Menu. Risk can be taken with great menu on board. It also shows people love and explore cuisines of various countries.
- Scope to increase Farmers markets in Bronx, Queens and Staten Island.

Region - Cluster0

Neighborhood – Staten Island

Tod Hill, Port Ivory, Bloomfield



Conclusion

Analysis performed on limited data

- Re-run program with updated information
- 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 neighbourhood or restaurant type mentioned like Indian Restaurant, analysis can be checked. A venue with lowest risk and competition can be identified