



THE BATTLE OF THE NEIGHBORHOODS

Applied Data Science Capstone

Realized by: **EL GHOUZAL SAAD**
January 2021

INTRODUCTION

- New York city review for ABC Company.
- Optimum location for new Restaurant business
- Business Problem :
 - Choice of first neighborhood to start restaurant business.
 - Easy to replicate.
 - Low competition
 - High demand
 - Choice of Menu
- Success Criteria :
 - Best neighborhood which meets above criteria.

NEWYORK CITY - FACTS

- Most populous city in the United States.
- It is diverse and is the financial capital of USA .
- It is multicultural .
- Provides lot of business opportunities .
- Business friendly environment .
- Attracted many different players into the market .
- 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, theatre, fashion, and the arts in the United States.



NEWYORKCITY - BUISSNESS ENVIRONMENT

- Market is highly competitive.
- Highly developed city so cost of doing business is also one of the highest.
- New business venture or expansion needs to be analyzed carefully.
- One should strategically targeting the market in order to :
 - This will help in reduction of risk.
 - The Return on Investment will be reasonable.



NEWYORK CITY- 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.



NEWYORK CITY- CUISINE

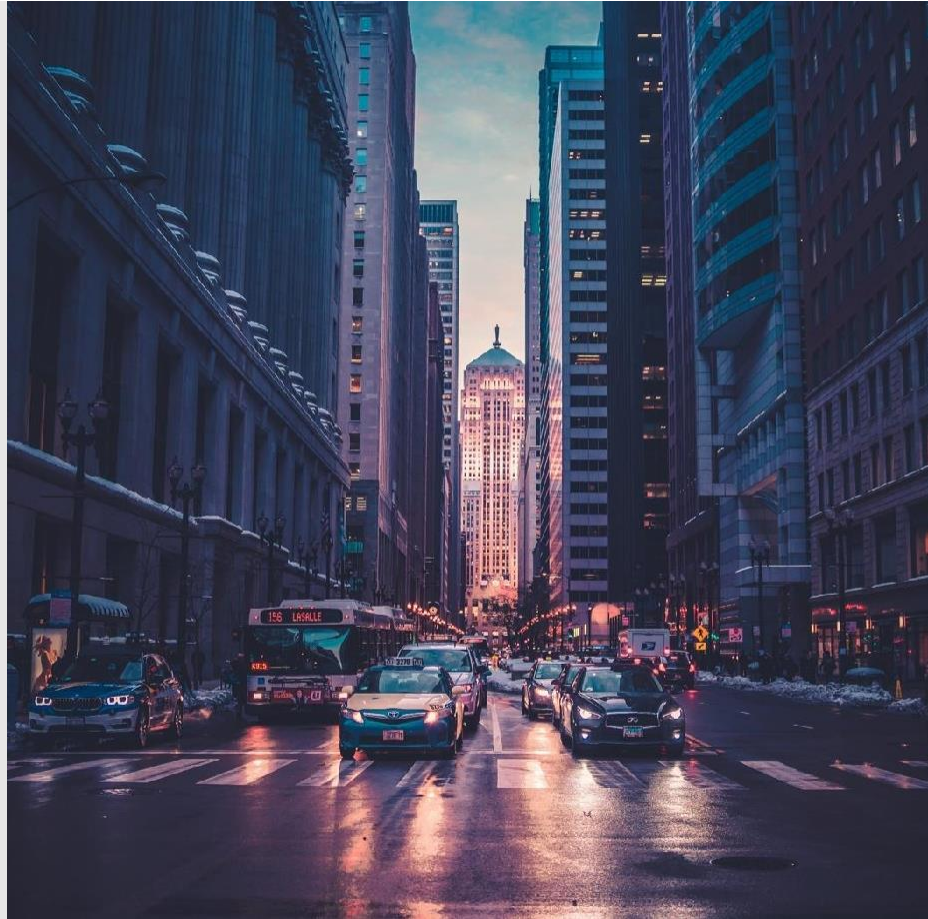
- Middle Eastern foods such as falafel and kebabs examples of modern New York street food.
- Famous for fine dining Michelin starred restaurants. The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin.
- Soit is evident that to survive in such competitive market it is very important to strategically plan.
- Mobile food vendors - Some 4,000 licensed by the city.



FACTORS TO STUDY



- 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
 - Segmentation of the Borough
 - Untapped markets
 - Saturated markets etc..
- The list can go on...

DATA DESCRIPTION

- 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

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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	Starbucks	40.877531	-73.905582	Coffee Shop
4	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop

DATA DESCRIPTION

• DATA 2 : 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

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,418,207	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	368,500	22.83	59.13
3	Queens	Queens	2,253,858	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	30,500	58.37	151.18
5		City of New York	8,336,817	842.343	302.64	783.83
6		State of New York	19,453,561	1,731.910	47,214	122,284

DATA DESCRIPTION

- DATA FROM [FOURSQUARE.COM](https://www.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	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



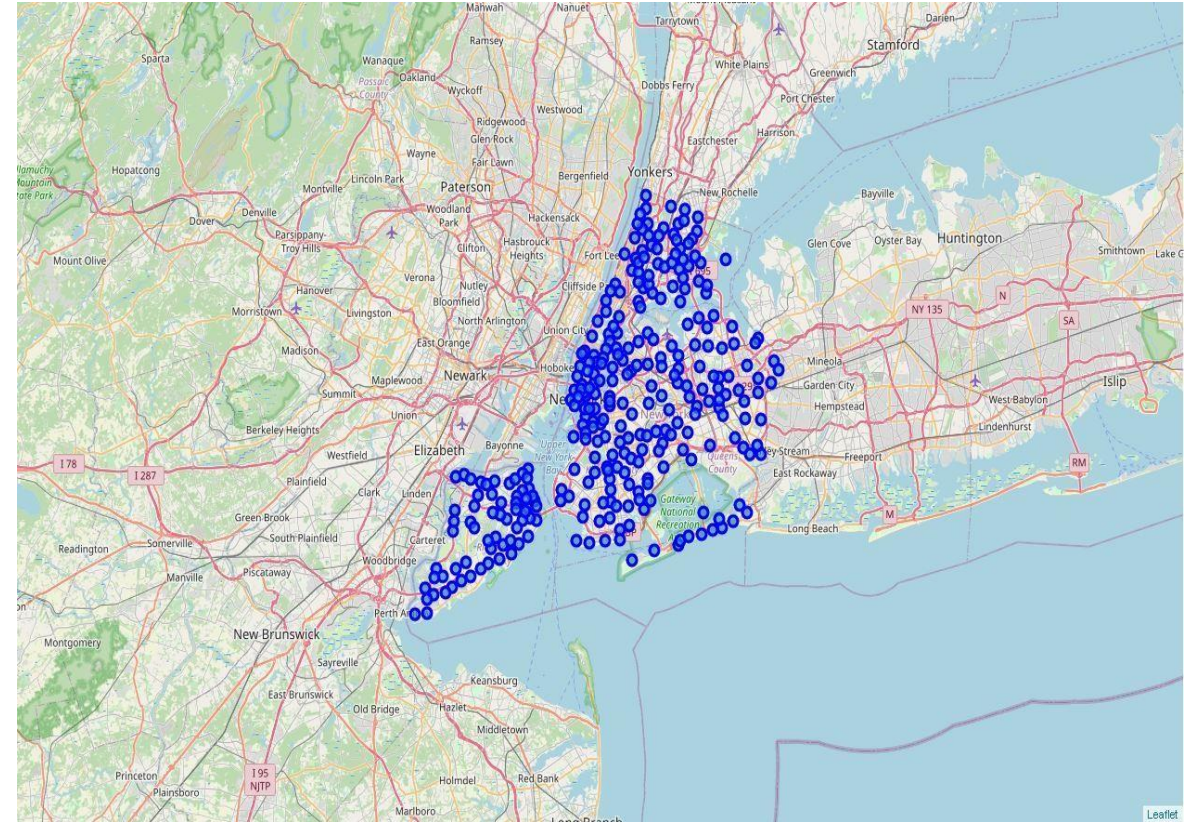
ANALYTIC APPROACH

- New York city neighborhood 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

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

Data 2 :

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

- Manhattan (New York County) is the geographically smallest and most densely populated borough.
- Manhattan's (New York County's) population density of 368,500 people per square mile (22.83/km²) in 2017 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
0	The Bronx	Bronx	1,418,207	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	368,500	22.83	59.13
3	Queens	Queens	2,253,858	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	30,500	58.37	151.18
5		City of New York	8,336,817	842.343	302.64	783.83
6		State of New York	19,453,561	1,731.910	47,214	122,284

METHODOLOGY

- New York City is the most populous city in the United States, with an estimated record high of 8,336,817 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 Indian, African, Japan etc. This also increases the scope for restaurants business in New York City.

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,418,207	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	368,500	22.83	59.13
3	Queens	Queens	2,253,858	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	30,500	58.37	151.18
5		City of New York	8,336,817	842.343	302.64	783.83
6		State of New York	19,453,561	1,731.910	47,214	122,284

METHODOLOGY

Data 3 :

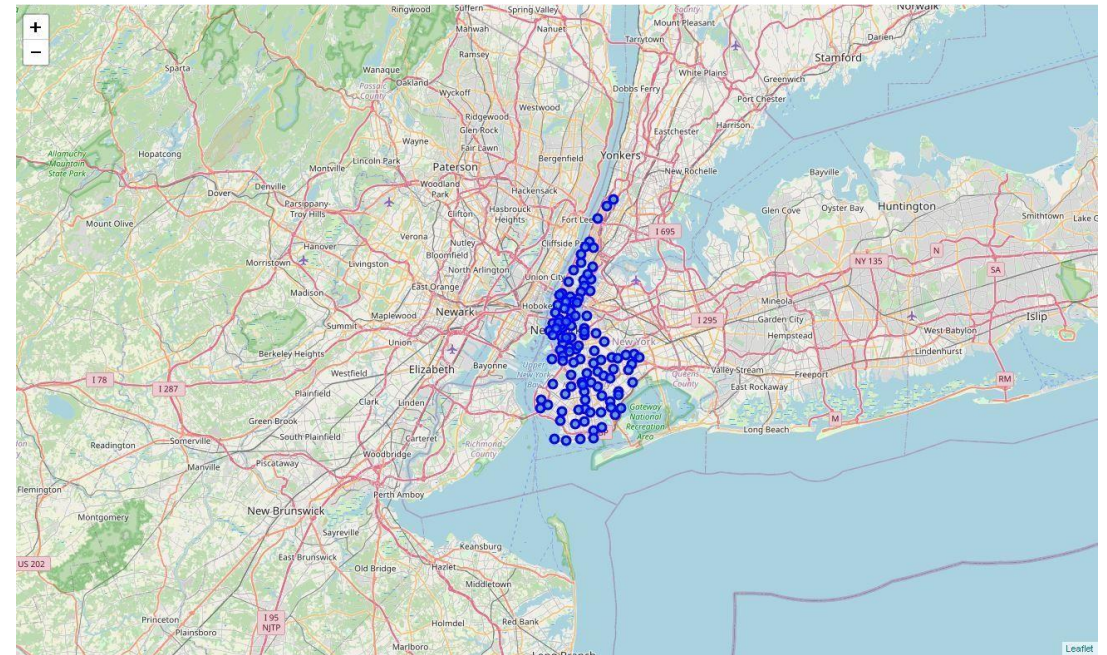
- New York city geographical coordinates data has been utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood.
- We used the Foursquare API data to explore neighborhoods in New York City.
- Using the geographical coordinates of each neighborhood 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.

METHODOLOGY

- **PART 1 :Brooklyn and Manhattan**
- Brooklyn and Manhattan Venues 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

Brooklyn and Manhattan Visualization

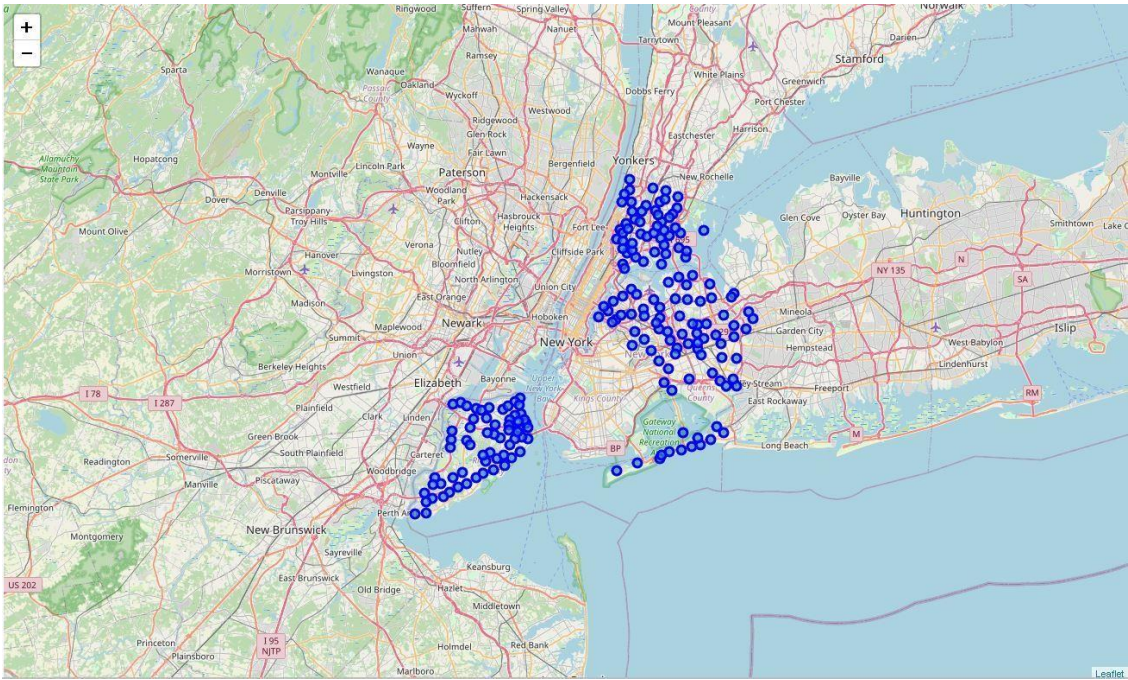


METHODOLOGY

- **PART 2** : Bronx, Queens and Staten Island.
- Bronx, Queens and Staten Island Venues data

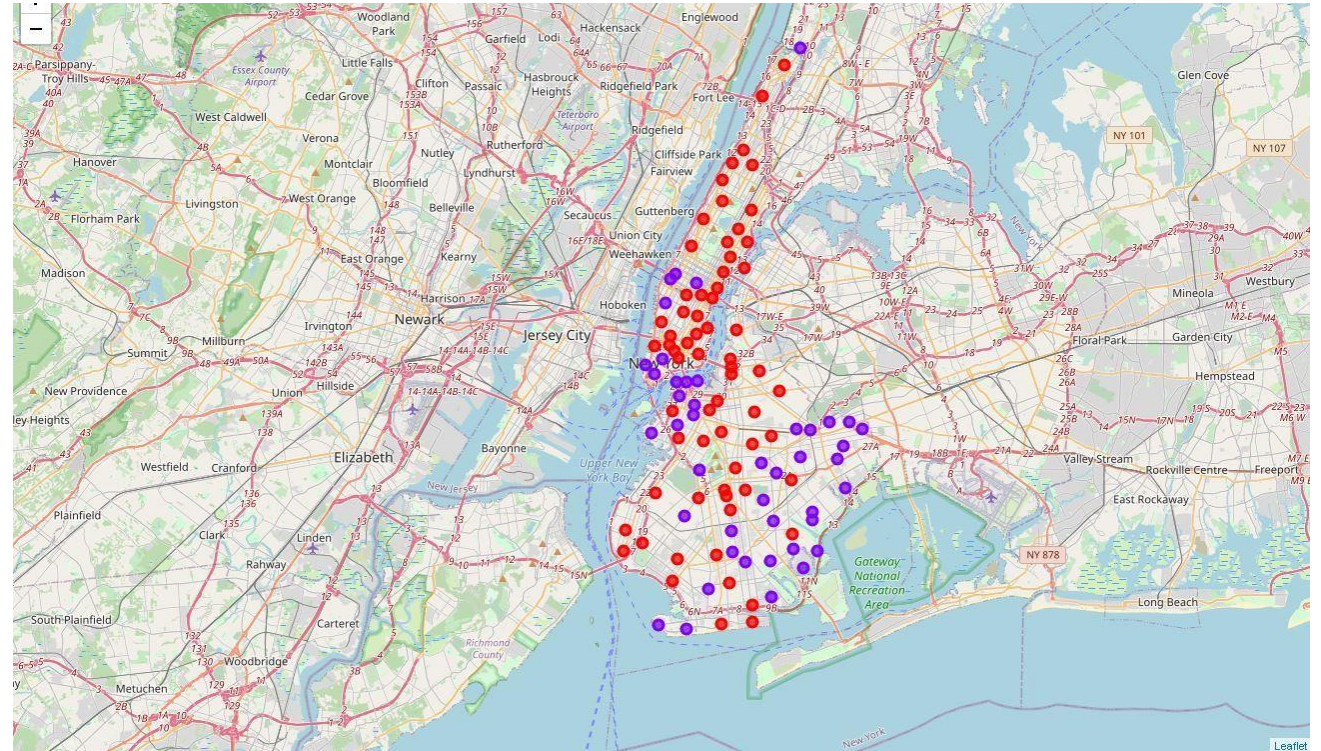
	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Allerton	66	66	66	66	66	66
Annadale	19	19	19	19	19	19
Arden Heights	22	22	22	22	22	22
Arlington	22	22	22	22	22	22
Arrochar	23	23	23	23	23	23
Arverne	34	34	34	34	34	34
Astoria	100	100	100	100	100	100
Astoria Heights	76	76	76	76	76	76

Bronx, Queens and Staten Island.



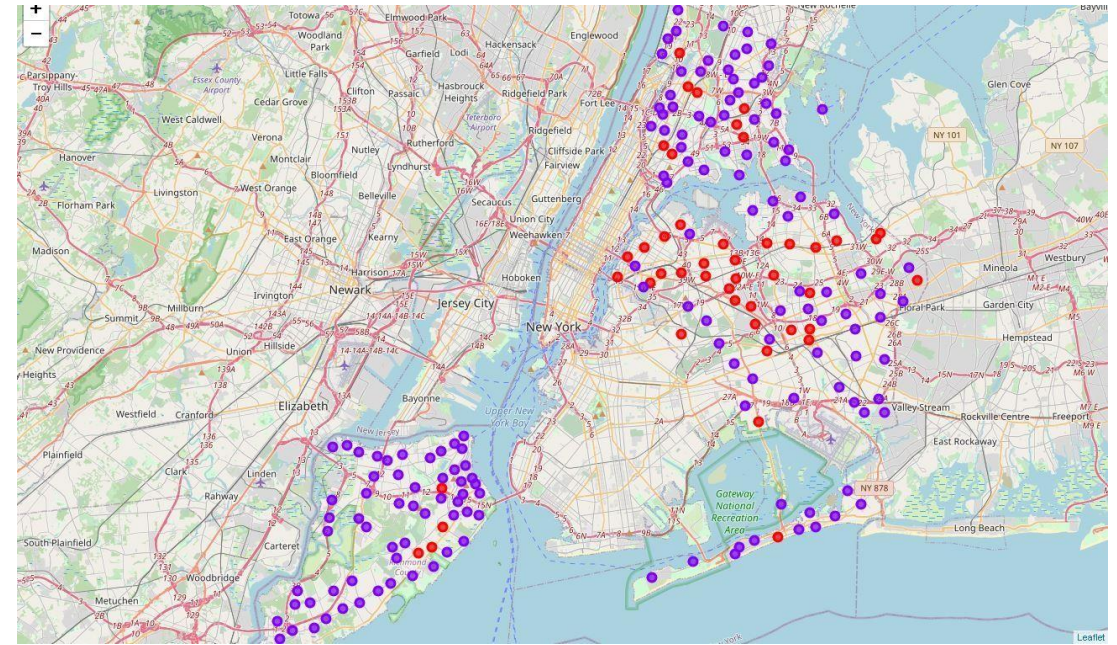
RESULT

- **PART 1 :Brooklyn and 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.



RESULT

- **PART 2 : Bronx, Queens and Staten Island.**
- Segmenting 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	1
1	Staten Island	Port Ivory	40.639683	-74.174645	0	1
2	Staten Island	Butler Manor	40.506082	-74.229504	0	1
3	Staten Island	Bloomfield	40.605779	-74.187256	0	1

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, Butler Manor.

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



THANK YOU

-EL GHOUZAL SAAD 2021-