

Capstone Final Project

Introduction

- Consider New York as a place to open a restaurant offering delivery services and high quality products, due to increased demand during the pandemic.
- Consider different boroughs and neighbourhoods in New York.
- Consider famous cuisines and corresponding competition per area.
- Find the best location as per the above criteria

New York City

- Most famous in USA.
- Financial capital and major center of several sectors like banking, finance, tourism.
- Multicultural (variety of cuisines).
- High number of population.
- Highly competitive market.
- Different nationalities that have different food tastes.

Factors to consider for restaurant location

- New York population.
- New York demographics.
- Proximity of farmers market, for higher quality products.
- Different type of venues in close radius.
- Number of competitors per area and menu served.
- Saturated markers.
- Segmentation of the markets.

Data Description - Data 1

- New York has several boroughs which are further divided into several neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

Data Description - Data 2

- Second data which will be used is the DOHMH Farmers Markets. In this we will be using the data of Farmers Markets.
- A *farmers' market* is a public site used several local producers for the direct sale of farm products to consumers.

Data Description - Data 3

- Wikipedia data will be used for the below:
 - New York Population
 - New York City Demographics
 - Cuisine of New York city

Data Description - 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 neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City.

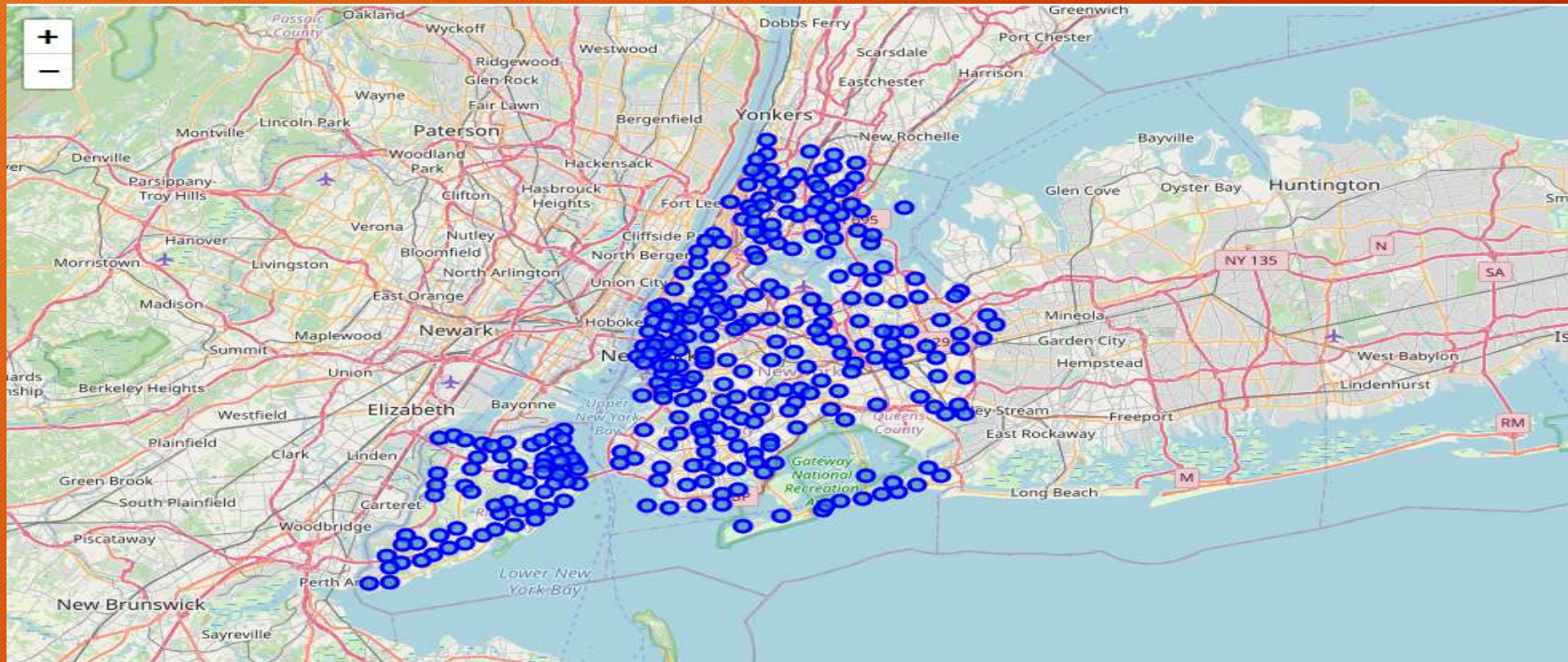
Analytical Approach

- New York city is divided into 5 boroughs and several neighbourhoods.
- First part is clustering of Manhattan and Brooklyn .
- Second part is clustering of Bronx, Queens and Staten Island.
- Above approach will be explained in the following slides.

Explanatory Analysis - Data 1

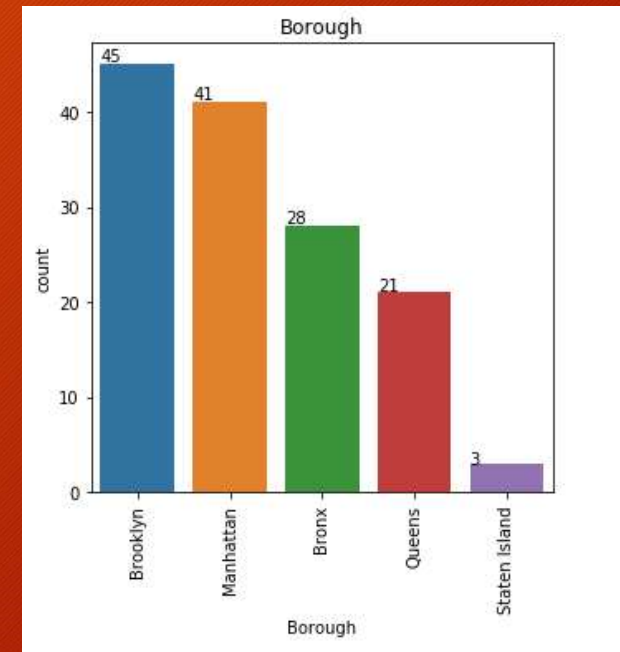
- 1. In this we load the data and explore data from `newyork_data.json` file.
- 2. Transform the data of nested python dictionaries into a pandas dataframe.
- 3. This dataframe contains the geographical coordinates of New York city neighbourhoods.
- 4. This data will be used to get Venues data from Foursquare.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

Explanatory Analysis - Data 1



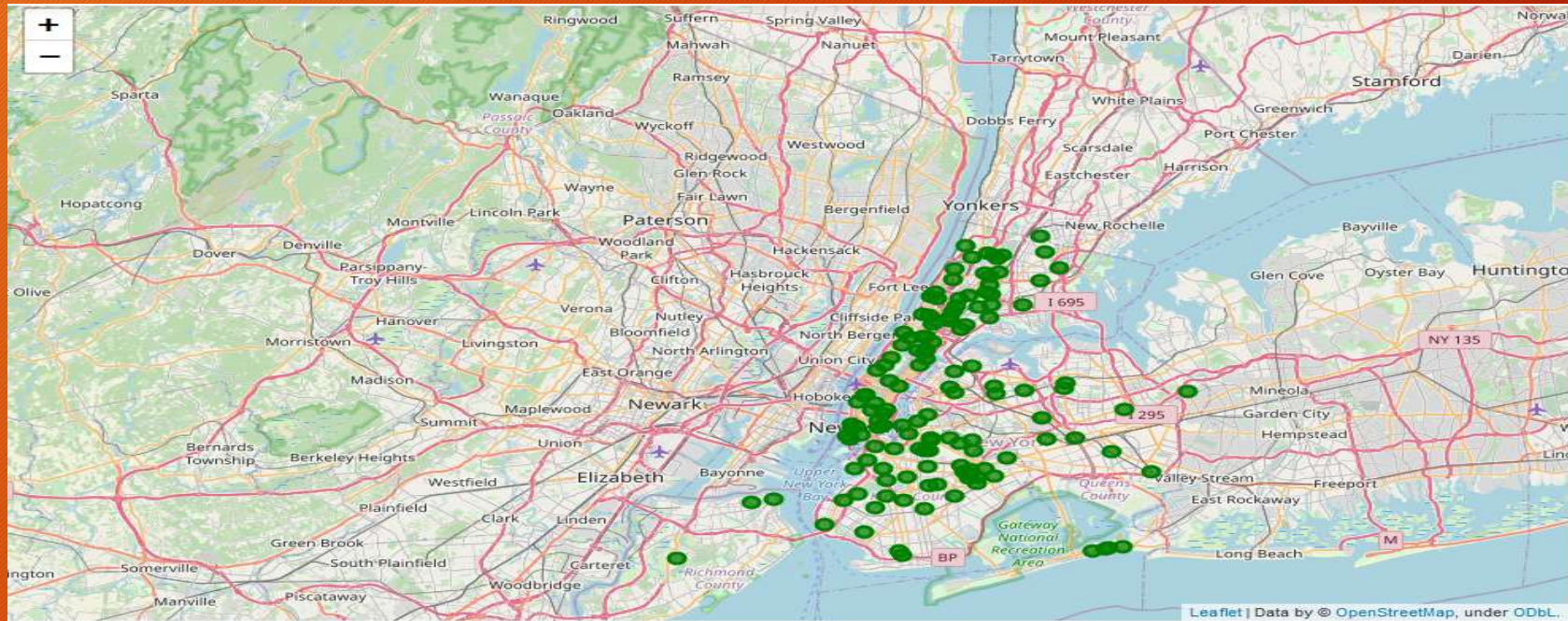
Explanatory Analysis - Data 2

- The data of Farmers Markets will be used.
- There are totally 138 Farmers Markets in New York city.
- Highest number are in Brooklyn(45) and Manhattan(41).
- And lowest in Bronx(28), Queens(21), and Staten Island(3).



Explanatory Analysis - Data 2

- Farmers Market Visualisation in New York.



Explanatory Analysis - Data 3

- New York city Population, Demographics and Cuisine, data from Wikipedia were used.
- BeautifulSoup python library was used in some occasions and in some other the data was gathered manually.

Explanatory Analysis - Data 3 - Population

- Manhattan (New York County) is the geographically smallest but the borough with highest population density.
- Manhattan's (New York County's) population density is 71.341 people per square mile (27,544/km²) in 2019.
- Brooklyn is the New York's most populous borough with 2,56m people.
- Queens is geographically the largest borough

	Borough	County	Estimate_2019	GrossDomesticProduct	GDP/capita	square miles	square km	squarekm	persons/mi2	persons/km2
0	The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04	NaN	33867	13006
1	Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42	NaN	36147	13957
2	Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13	NaN	71341	27544
3	Queens	Queens	2,253,858	93.310	41,400	108.53	281.09	NaN	20767	8018
4	Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18	NaN	8157	3150
5	City of New York		8,336,817	842.343	101,000	302.64	783.83	27,547	NaN	27547
6	State of New York		19,453,561	1,731.910	89,000	47,126.40	122,056.82	412	NaN	412

Explanatory Analysis - Data 3 - Demographics

- New York City has 8,3m residents as of 2019.
- The racial composition of the above population is set below. For this reason there is a wide variety of restaurants offering different cuisines (Italian, Greek, Indian etc).

Racialcomposition	2010	1990	1970	1940
White	0.440	0.523	0.766	0.936
—Non-Hispanic	0.333	0.432	0.629	0.920
Black or African American	0.255	0.287	0.211	0.061
Hispanic or Latino (of any race)	0.286	0.244	0.162	0.016
Asian	0.127	0.070	0.012	0.000

Explanatory Analysis - Data 3 - NY Cuisine

- Data is taken from Wikipedia page - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City .
- Using this data word cloud was prepared so as to find the most popular cuisine per area.
- Most Preferred Food in New York City -Mexican, Indian, Italian, Jewish, Chinese, Puerto Rican.



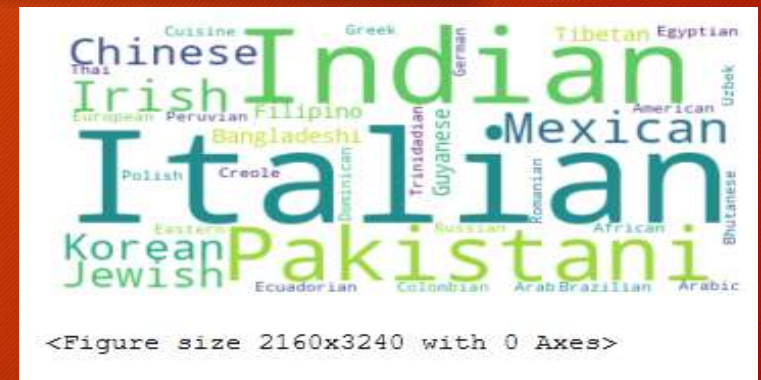
Explanatory Analysis - Data 3 - Cuisines

- **BROOKLYN CUISINE** -Most Preferred Food in Brooklyn is - Italian, Puerto Rican, Mexican
- **MANHATTAN CUISINE** - Most Preferred Food in Manhattan is - Italian, Chinese, Puerto Rican and Indian.



Explanatory Analysis - Data 3 - Cuisines

- QUEENS CUISINE - Most Preferred Food in Queens is - Indian, Italian, Pakistani.
- THE BRONX CUISINE - Most Preferred Food in The Bronx is - Italian, Puerto Rican, Dominican and Albanian.



Explanatory Analysis - Data 4 - Foursquare.com

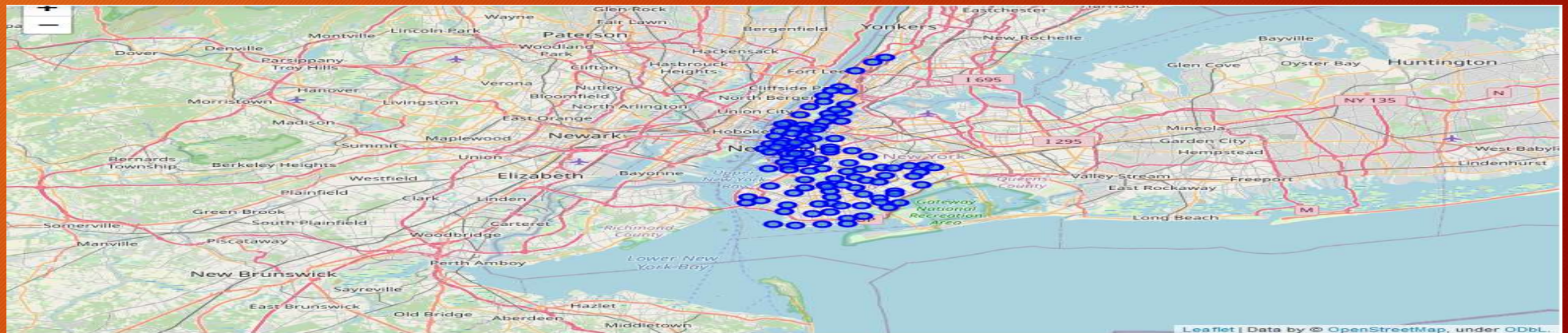
- New York city geographical coordinates data has been utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City.
- Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

Analysis was split in two parts:

- Brooklyn and Manhattan
- Bronx, Queens and Staten Island

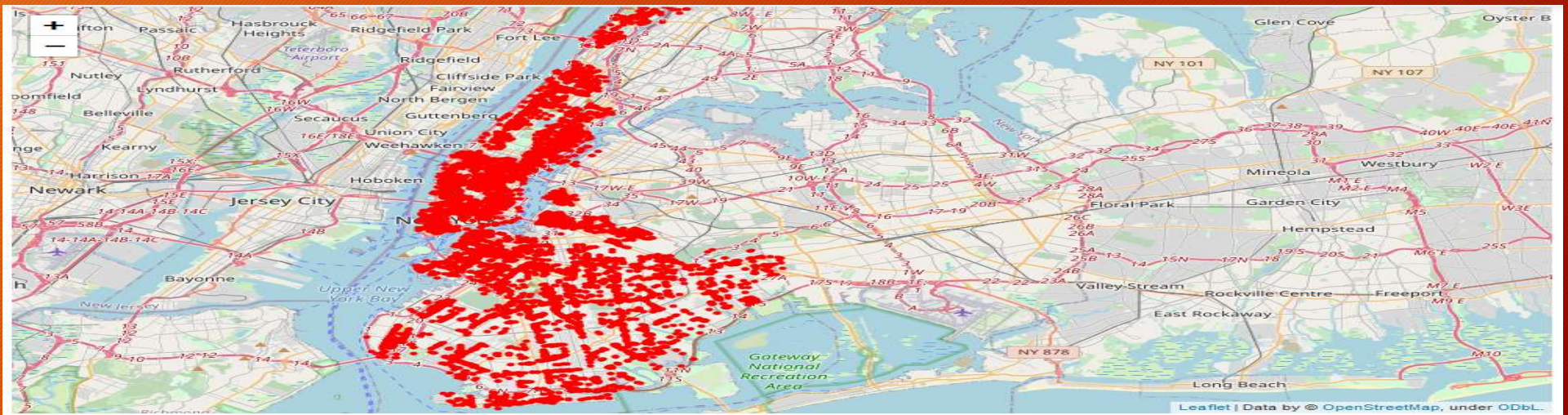
Explanatory Analysis - Data 4 - Brooklyn & Manhattan

	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



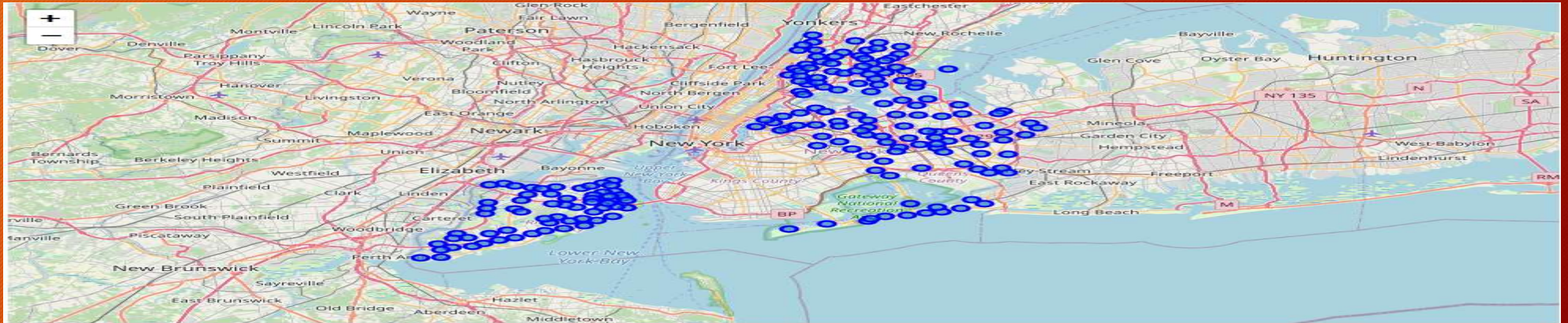
Explanatory Analysis - Data 4 - Brooklyn & Manhattan

- Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9581 venues and 409 unique venue types.



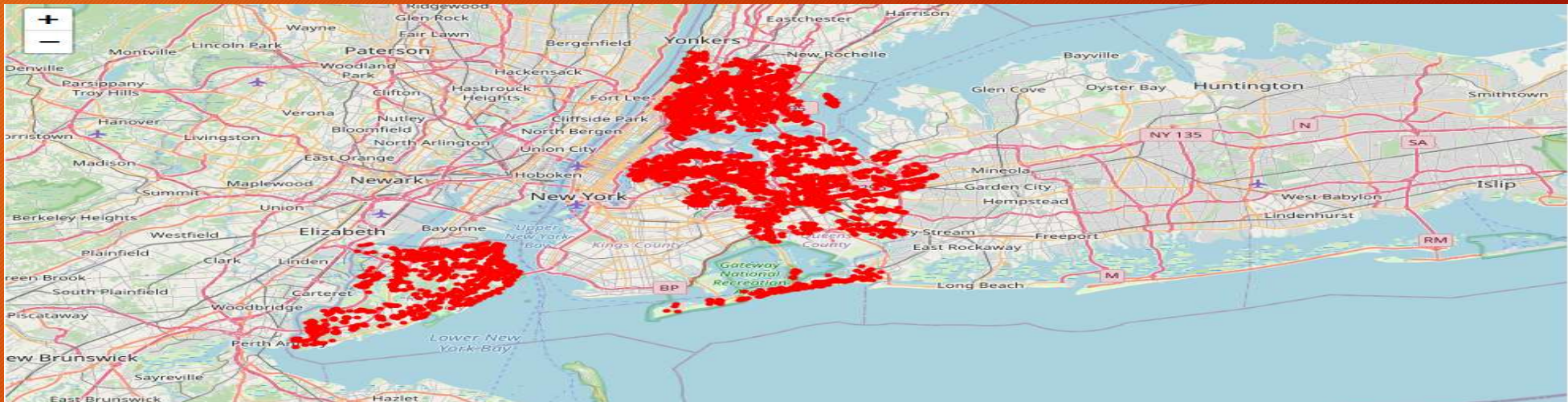
Explanatory Analysis - Data 4 - Bronx, Queens and Staten Island

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
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	40.896649	-73.844846	Pharmacy



Explanatory Analysis - Data 4 - Bronx, Queens and Staten Island

- **Bronx, Queens and Staten Island Venues Visualization :** The "BQS_venues" dataframe has 10899 venues and 394 unique venue types.

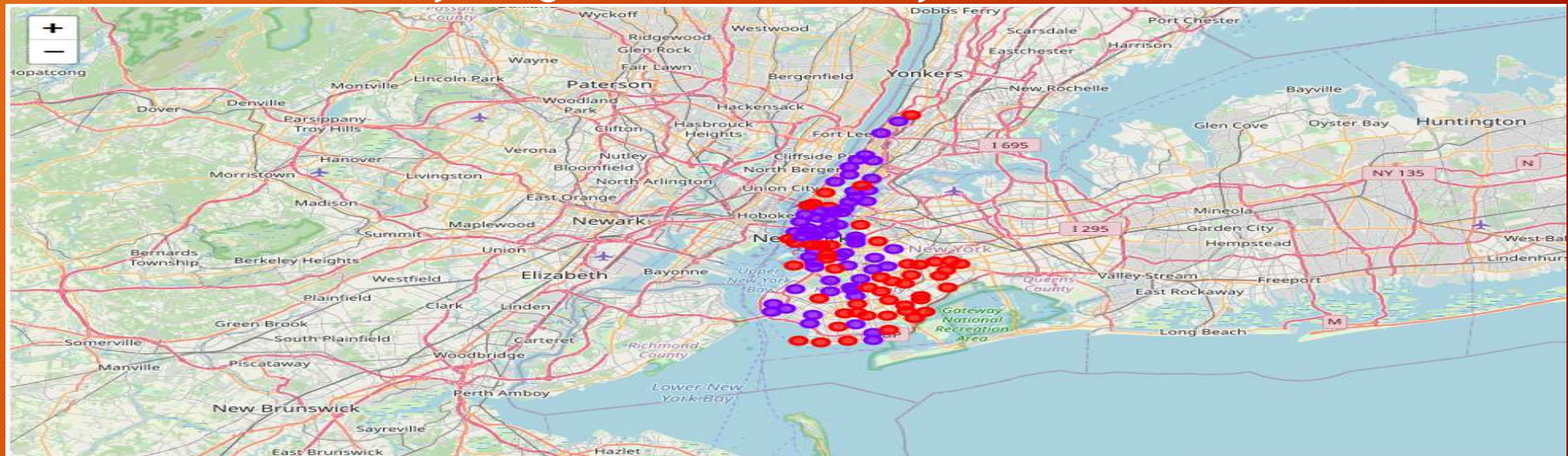


Results - Brooklyn Manhattan

- To cluster the neighbourhoods 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.
- **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.

Results - Brooklyn Manhattan

- In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.

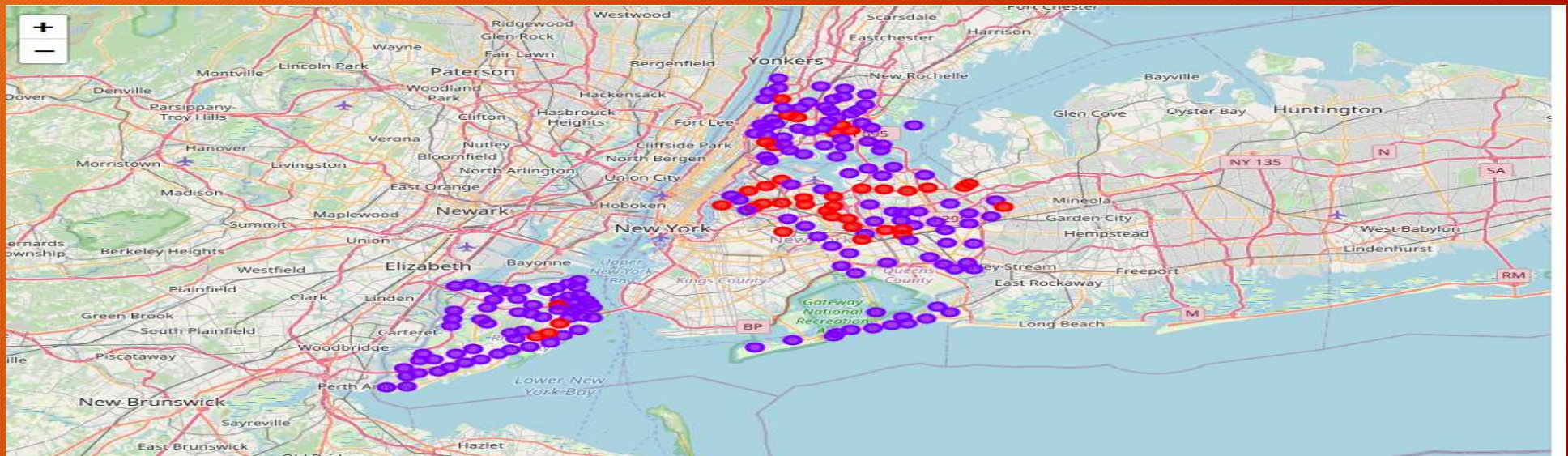


Results - Bronx, Queens and Staten Island

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

Results - 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.



Discussion

- 1. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
- 2. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. This is also an indication that people love eating cuisines of various countries.

Conclusion

- Brooklyn and Manhattan has high concentration of restaurant business and this also indicates a highly competitive market.
- The market in Bronx, Queens and Staten Island is not saturated and although there are several restaurants in the area and therefore this area could be further examined for a business opportunity.