Report: The Battle of the Neighbourhoods

1. Introduction and Business Problem

1.1 Background

New York is the most populous city in the United States. It is also the financial capital of USA. It is a global hub of business and commerce. The city is a major centre for banking and finance, retailing, tourism, real estate, media, legal services, accountancy, insurance etc. in the United States. This also means that the market is highly competitive and hence the cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from this analysis will give comprehensive understanding of the NY city landscape and the neighbouring places where businesses can consider for their expansion or setting up. This will help in risk mitigation and achieve better return on investment for their investments.

1.1 Problem Description

A restaurant is a business which offers food to customers in return for money. The City of New York is famous for its excellent cuisine. New York has some of the best fine dining restaurants which are highly rated for their range, taste, ambience and service. The city is home to "nearly one thousand of the finest and most diverse 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 and invest. Various factors need to be studied to zero in on the location of setting up restaurant and some of them are

1.2 Target Audience

The objective is to locate and recommend which neighbourhood of New York City will be best choice to start a restaurant. The recommendation of best location to set up restaurant is arrived by analysing various factors that can have influence on the success of the restaurant. This would interest anyone who wants to start a new restaurant in New York City.

2. Data :

We will be using the below datasets for analysing Newyork city

Data 1 : Neighbourhood has a total of 5 boroughs and 306 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 the latitude and longitude coordinates of each neighborhood.
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 2 : Second data is the data of Farmers Markets.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-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 3 : For the below analysis we will get data from Wikipedia as given below :
1. New York Population
2. New York City Demographics
3. 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 Data

	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 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 neighbourhoods in New York City. The below is image of the Foursquare API data.

	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	Rite Aid	40.896521	-73.844680	Pharmacy
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898283	-73.850478	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Dunkin Donuts	40.890631	-73.849027	Donut Shop

3.Methodology :

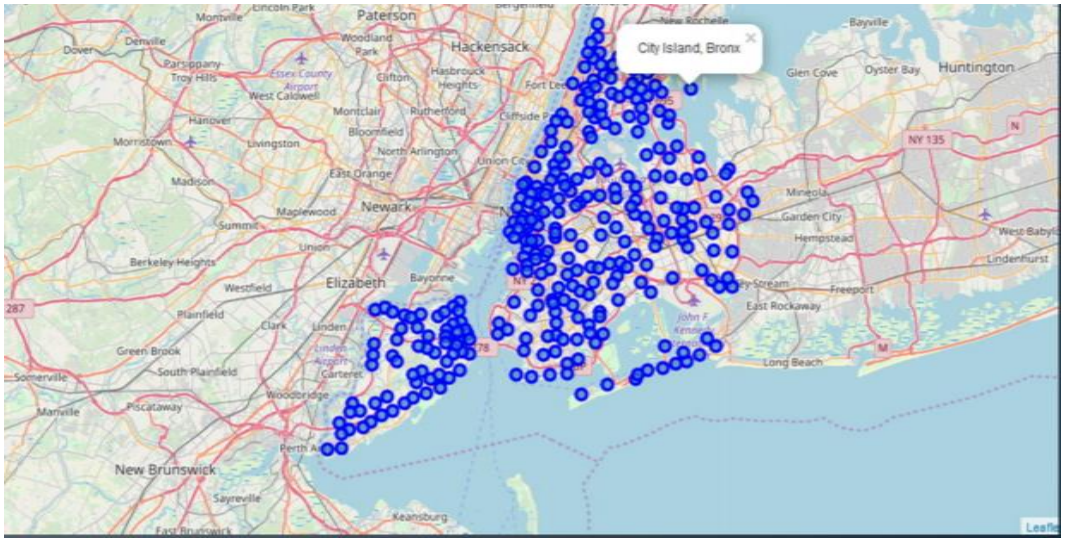
Business Understanding : Our main goal is to get optimum location for new restaurant business in New York City.

3.1 Analytical Approach :

New York city neighbourhood 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.

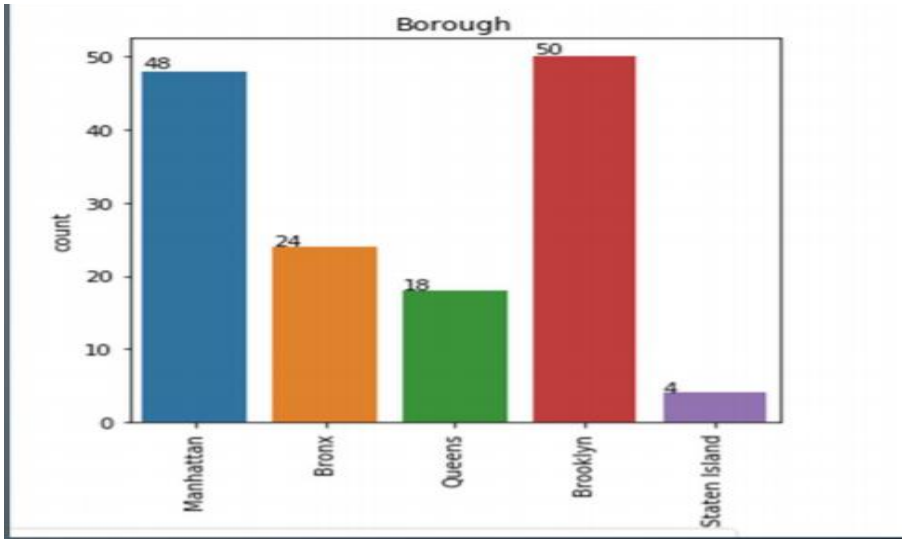
Exploratory Data Analysis :

Data 1- New york city Geographical Coordinates Data.
1. In this we load the data and explore data from newyork_data.json file.
2. Transform the data 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.

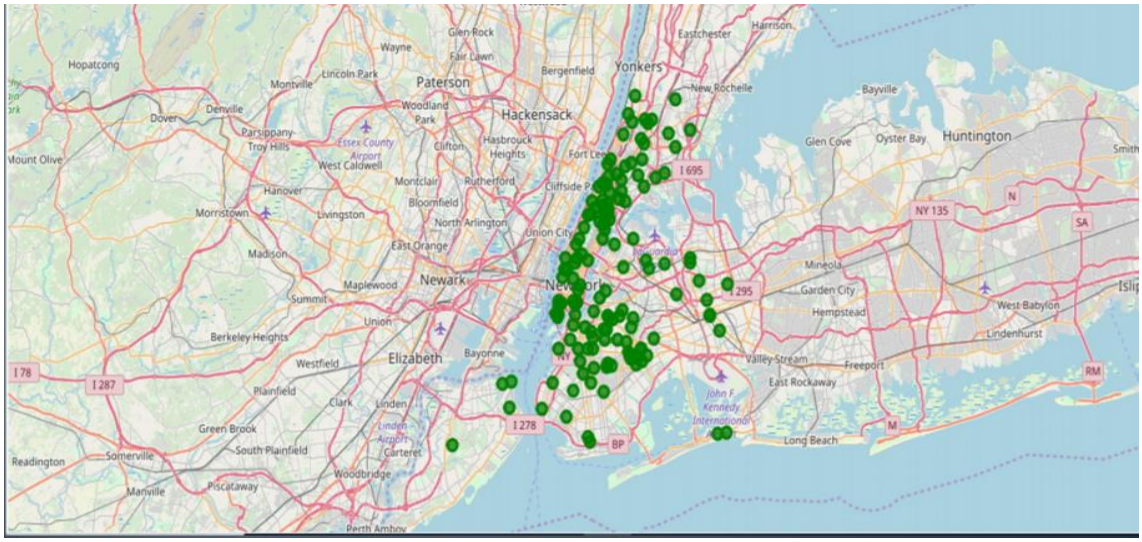


Data 2-

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 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn. And lowest in Queens, Bronx and Staten Island.



Farmer's Market Visualization



Data 3 :

- **To analyze New York city Population, Demographics and Cuisine , scrapped the data from Wikipedia pages listed in the data section and used BeautifulSoup python library**
- **Beautiful Soup is a Python package for parsing HTML and XML documents**
- **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

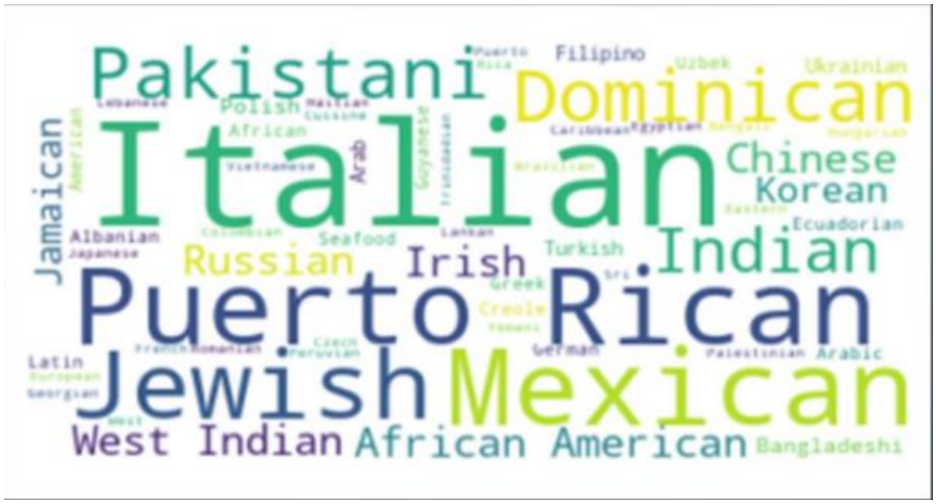
- **Manhattan (New York County) is the geographically smallest and most densely populated borough**
- **Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual 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

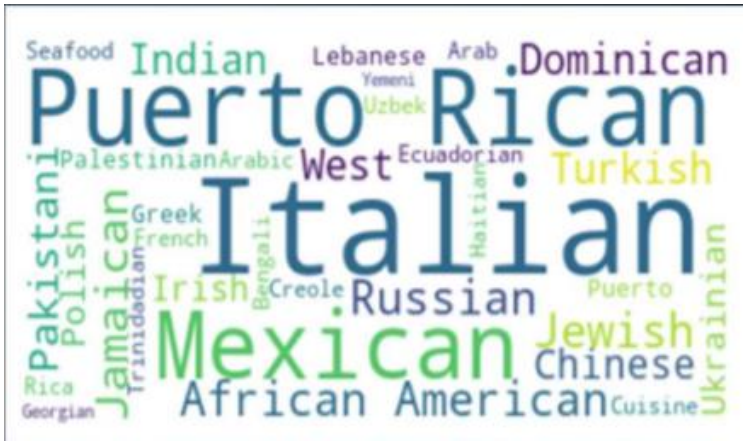
- **New York City is the most populous city in the United States , with an estimated record high of 8,622,698 residents as of 2017,incorporating more immigration into the city than outmigration since the 2010 United States Census**
- **The racial composition is as given below. This is the reason New York city has restaurants serving cuisines as diverse from Chinese, Indian, Pakistani, Jewish, Italian, Dominican, Mexican, Puerto Rican etc. This also increases the scope for restaurants business in New York City.**

Word cloud – Famous Cuisines of NY city



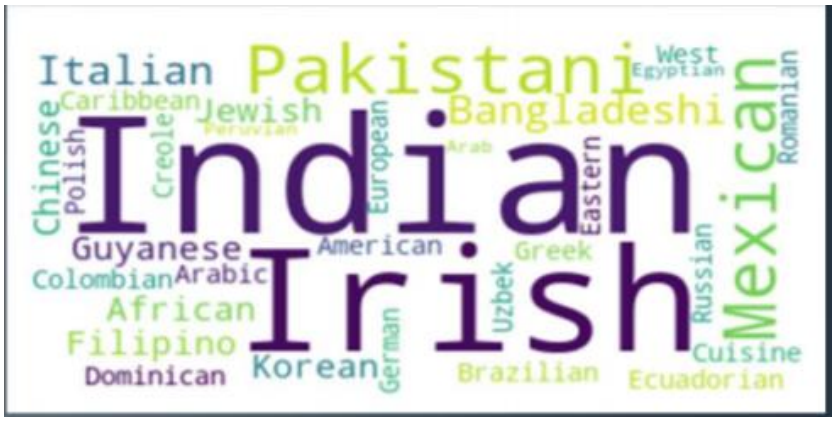
Segment 1: Brooklyn and Manhattan cuisine preferences

Brooklyn Cuisine Preference



Manhattan Cuisine Preference

Queens Cuisine preferences



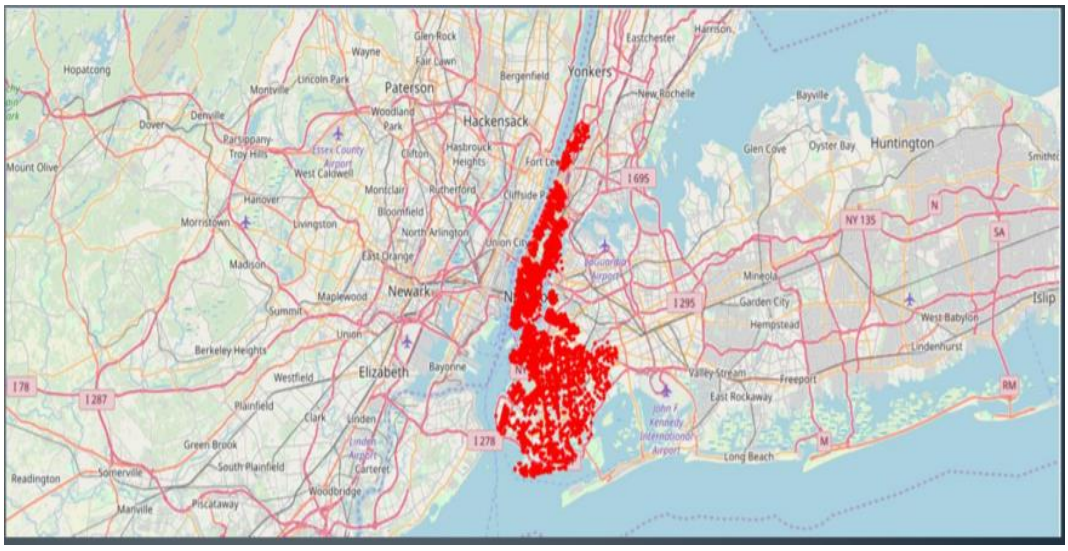
Bronx Cuisine preferences

Methodology 4 – Foursquare coordinates and venues data

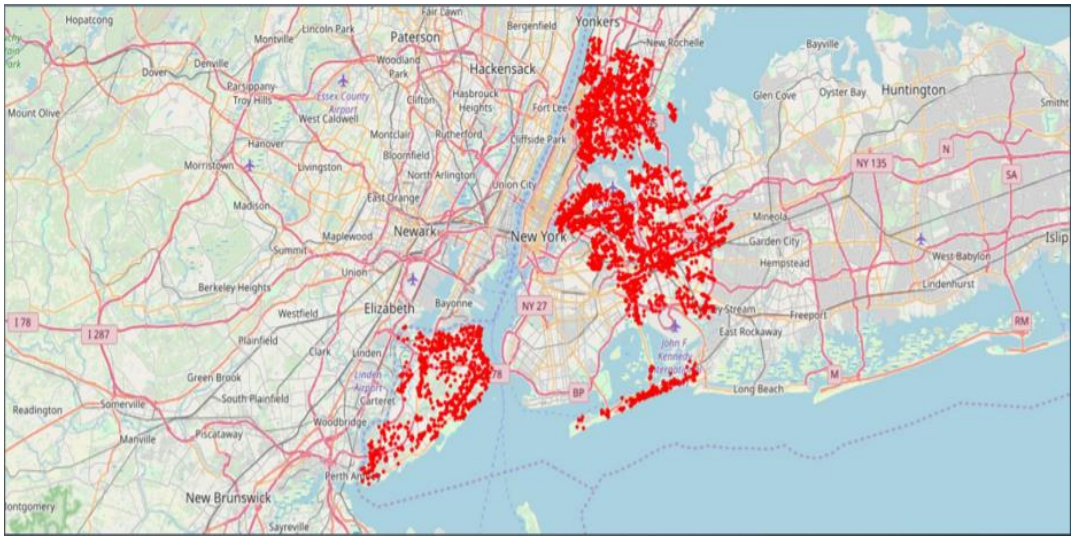
- New York city geographical coordinates data has be utilized as input for the Foursquare API to explore neighbourhoods in New York City
- Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters
- PART – 1 Brooklyn and Manhattan
- PART – 2 Bronx, Queens and Staten Island

Methodology 4 – Visualizing Segment1 and Segment 2 venues

- Brooklyn and Manhattan venues visualization
- The segment#1 has 9708 venues and 397 unique venue types

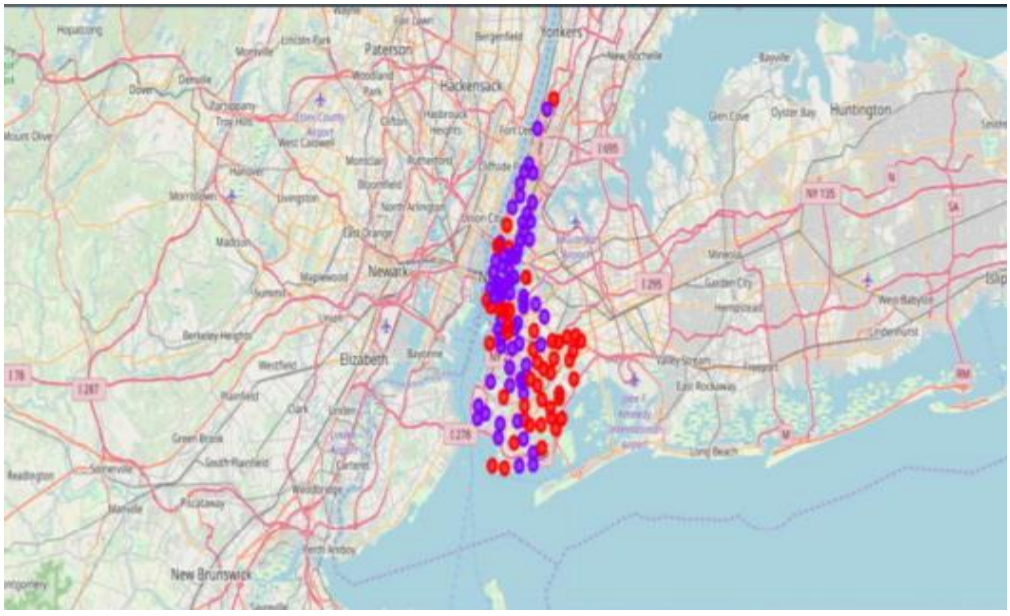


- Bronx, Queens and Staten Island venues visualization
- The segment#2 has 10805 venues and 387 unique venue types



Result – Part 1 (Brooklyn and Manhattan)

- Segmenting and Clustering Neighbourhoods
- Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated and relatively untapped compared to Cluster1
- Cluster1 : The Total and Total Sum is very high. It shows that the market is highly competitive and also saturated



Result – Part 2 (Bronx, Queens and Staten Island)

- Segmenting and Clustering Neighbourhoods
- Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated and relatively untapped compared to Cluster1
- Cluster1 : The Total and Total Sum is very high. It shows that the market is highly competitive and also saturated

Recommendation/Results:

- Brooklyn and Manhattan has high concentration of restaurant business and is highly competitive market
- Bronx, Queens and Staten Island seems to have lesser number of restaurants than required
- Scope to explore cuisines of various countries in Bronx, Queens and Staten Island
- Within the Bronx, Queens and Staten Island, neighbours such as Todt Hill, Port Ivory and Bloomfield can be considered for opening up new restaurants