Kenyan Restaurant in New York City

IBM Data Science Professional Certificate

Capstone Project

('Battle of neighbourhoods')

Table of contents

1. Introduction

A client wants to explore the opportunities of opening a restaurant in New York City. New York City has five boroughs spread around an area of approximately 784 square kilometres. Like many other cities around the world, New York City is a major attraction for workers and tourists. New York City is the centre of many industries such as financial services, media, real estate and entertainment among others. The client wants to open the restaurant in one of New York City's borough with less competition but a multicultural neighbourhood to benefit from people of different backgrounds and nationalities.

Target audience

In addition to this specific client, the factors we will study in this exercise and all collected information can serve as a basis for making recommendations to other customers who want to open a business in New York City.

Task description

A restaurant is a business that prepares and serves food and drinks to customers. Meals are generally served and eaten on the premises but many restaurants also offer take away and food delivery services. According to the Health Department, New York City has more than 26,000 eateries. With so many restaurants, competition is high and a careful analysis is necessary before opening a restaurant in New York City. Looking at all the 5 boroughs of New York City, which is a recommendable location to open a restaurant?

The location of a restaurant is an important part of its success, thus several factors need to be studied before recommendation can be done. These factors include:

- Rental prices for commercial property

- Menus of competitor restaurants
- Borough constituents
- Population of New York City
- Competitors in a particular location
- Untapped markets
- Location of high traffic facilities that attract high populations e.g. gyms, shopping malls, parks etc.
- Location of Farmers‘ markets where ingredients can be obtained fresh and cheap to maintain quality
- New York city demographics

2. Data description

The following data will be used to address the problem we are trying to solve for RESTAURANT LLC. We will be analysing the city of New York. Data sets to be used include:

Dataset 1

New York has a total of 5 boroughs and 306 neighbourhoods. For best results, we will need a dataset that contains all the 5 boroughs and the neighbourhoods together with the latitude and longitude coordinates.

Such a dataset has been found on the website of the New York University:

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

Dataset 2

We will be using the Farmers Markets dataset as well.

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. Some of the foods found in a farmers' markets include dairy products, vegetables, fish, meat, baked foods and 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

Dataset 3

For analysis of New York City's cuisine and demographics, we will get data from Wikipedia.

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

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

https://en.wikipedia.org/wiki/Portal:New_York_City

Dataset 4

We will use the Foursquare API to explore neighbourhoods of 5 boroughs in New York City. The coordinates of New York City will be utilized as input for the Foursquare API in order to provision venues information for each location. Below is an example of the Foursquare API.

	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

3. Methodology

Our goal is to select a suitable location on behalf of the RESTAURANT LLC to locate a restaurant business.

Analytical approach

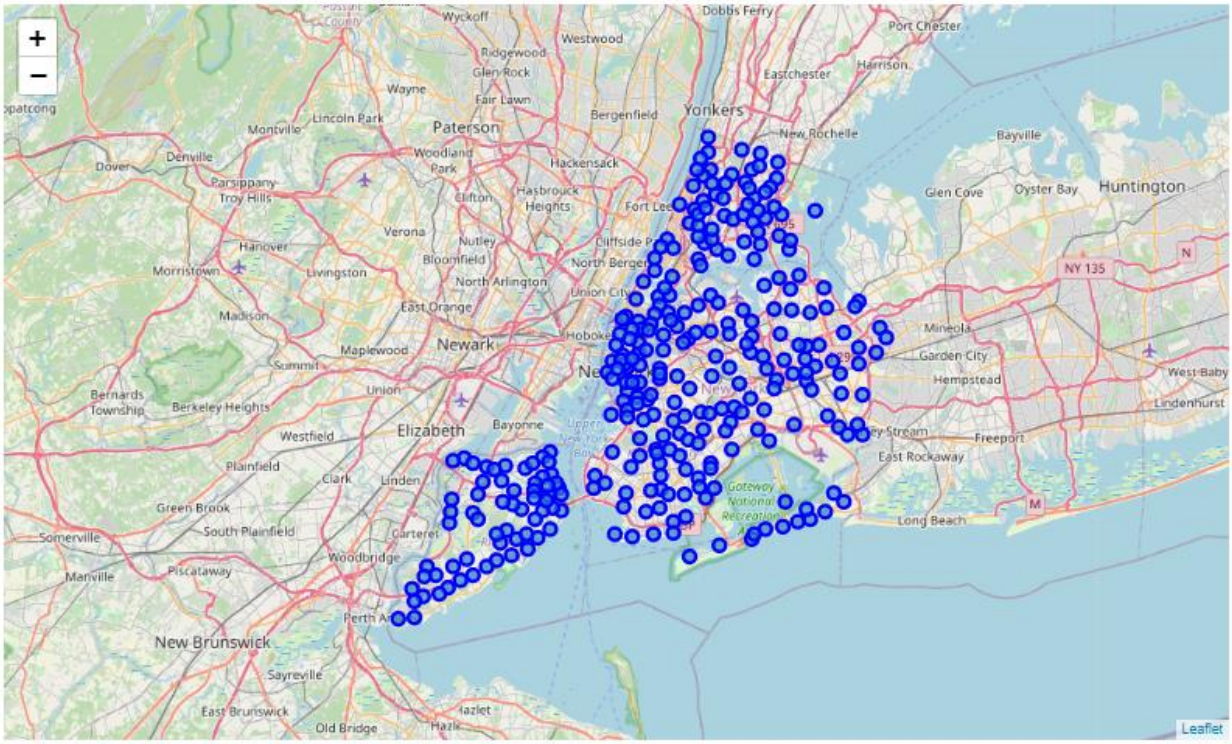
New York City has 5 boroughs and 306 neighbourhoods. Our approach will involve the clustering of Bronx, Queens, Staten Island, Brooklyn and Manhattan.

Data analysis

Dataset 1 – New York City Geographical coordinates data

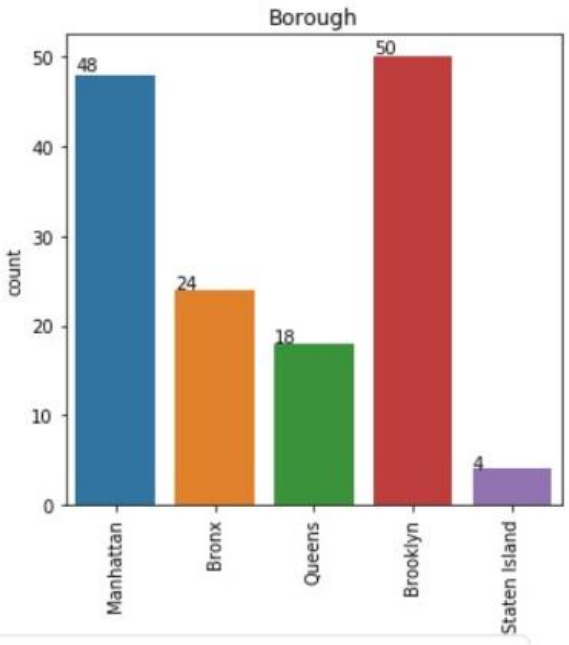
We load the data and explore data from New York University's website (https://geo.nyu.edu/catalog/nyu_2451_34572). We then transform the data of nested python dictionaries into a pandas dataframe. In this dataframe you will find geographical coordinates of New York City neighbourhoods. We will use this data to get venues data from Foursquare. In the visualization on the next page, we used geopy and folium libraries to create a map of New York city with neighbourhoods superimposed on top.

New York City neighbourhoods



Dataset 2 – Farmers Markets in New York City

We are also going to use the Farmers Markets dataset. According to the bar graph below the number of Farmers Markets in each Borough are as follows: Manhattan (48), Brooklyn (50), Bronx (24), Queens (18) and Staten Island (4).



Dataset 3 – Analysis of New York City Population, Demographics and Cuisine

We used a python library called Beautiful Soup. 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 for web scraping.

New York Population

Queens (2.25m) and Brooklyn (2.6m) are the most densely populated Boroughs respectively.

Queens is geographically the largest Burough (108.53 Square miles).

	Borough	County	Estimate_2019	GrossDomesticProduct	square_miles	square_km	persons_sq_mi	squarekm	persons/sq.mi	persons/l
0	The Bronx	Bronx	1,418,207	42.695\n	30,100	42.10	109.04			
1	Brooklyn	Kings	2,559,903	91.559\n	35,800	70.82	183.42			
2	Manhattan	New York	1,628,706	600.244\n	368,500	22.83	59.13			
3	Queens	Queens	2,253,858	93.310\n	41,400	108.53	281.09			
4	Staten Island	Richmond	476,143	14.514\n	30,500	58.37	151.18			
5	City of New York	8,336,817	842.343	101,000	302.64	783.83	27,547			
6	State of New York	19,453,561	1,731.910	89,000	47,126.40	122,056.82	412			
7	Sources: [14] and see individual borough articles									

New York City demographics

New York is the most populous city in the US with a population of 8,336,817 residents in 2019. The racial composition of the New York City is described in the diagram below.

	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 in New York City

We have taken data from Wikipedia about cuisine in New York City.
Source: https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

We used the data to prepare wordclouds for each Borough.

Manhattan Cuisine

The most popular cuisine in Manhattan is: Italian, American, Puerto Rican and Indian.

Brooklyn Cuisine

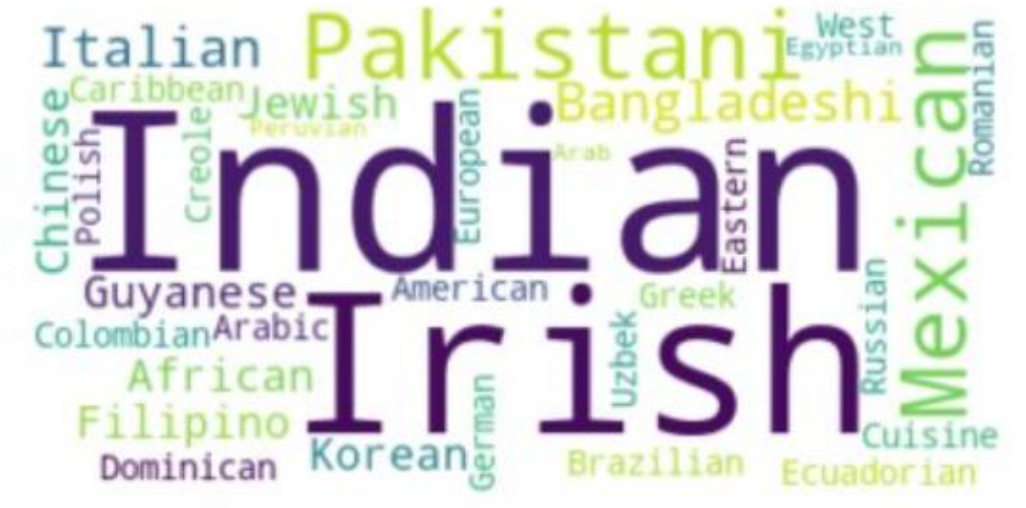
The most popular cuisine in Brooklyn are: Italian, Mexican, Puerto Rican, African and American.

The Bronx Cuisine

The most popular cuisine in the Bronx are: Italían, Puerto Rican, Dominican, Jamaican and Albanian.

Queens Cuisine

The most popular cuisine in Queens are: Indian, Irish, Mexican, Pakistani and Italian.

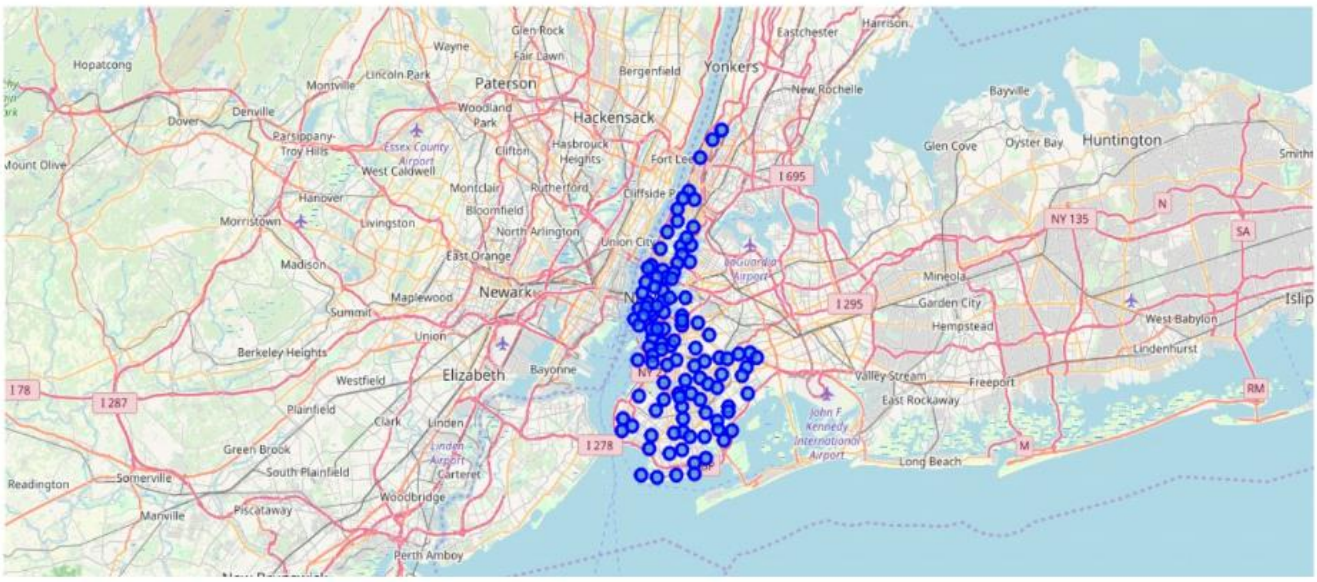


Staten Island has very limited data hence we could not prepare a word cloud.

Dataset 4 – Foursquare API

New York City geographical coordinates data has been used as input for the Foursquare API in order to obtain the venues of each neighbourhood. The Foursquare API has been used to explore neighbourhoods in New York City.

Brooklyn and Manhattan



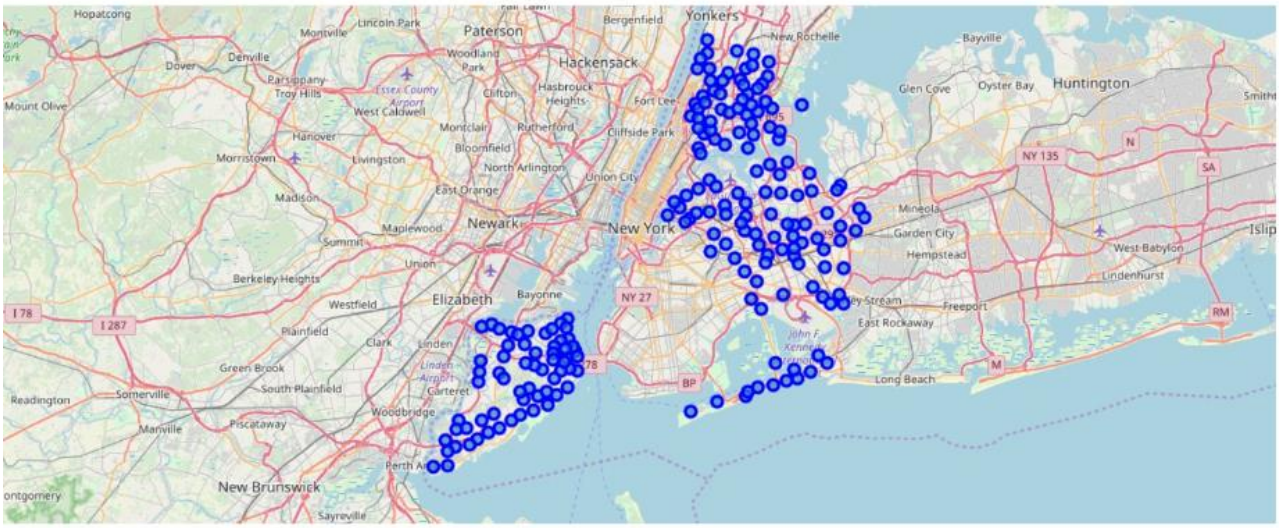
We used the Foursquare API to make calls and obtain the top 100 venues in a radius of 1000 meters as shown below.

	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

Manhattan and Brooklyn Venues

We have generated a visualization of venues in Manhattan and Brooklyn. We found 9708 venues and 397 unique venue types.

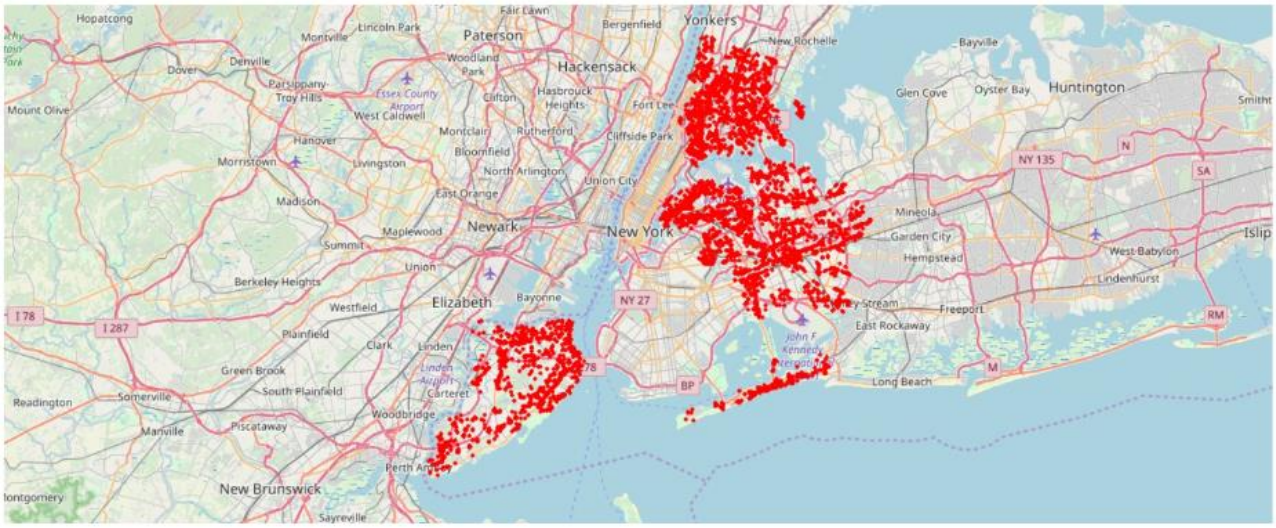
Staten Island, Queens and Bronx Neighbourhoods



Staten Island, Queens and Bronx Venues

	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

Staten Island, Queens and Bronx Venues



4. Results

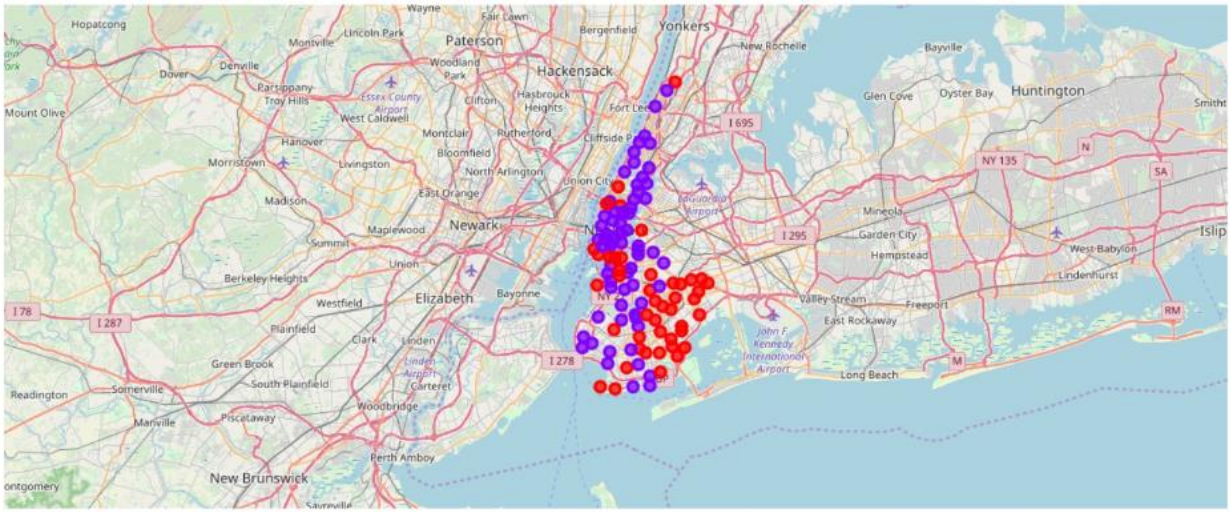
We have filtered and used venues data to create two cluster for: (1) Staten Island, Queens and the Bronx and (2) Brooklyn and Manhattan. Our focus has been only restaurant businesses.

Neighbourhood K-Means clustering based on mean occurrence of venue category

In order 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 an iterative refinement approach.

Manhattan and Brooklyn

In the visualization below, we can see different types of clusters created using K-Means for Manhattan and Brooklyn.



Cluster0

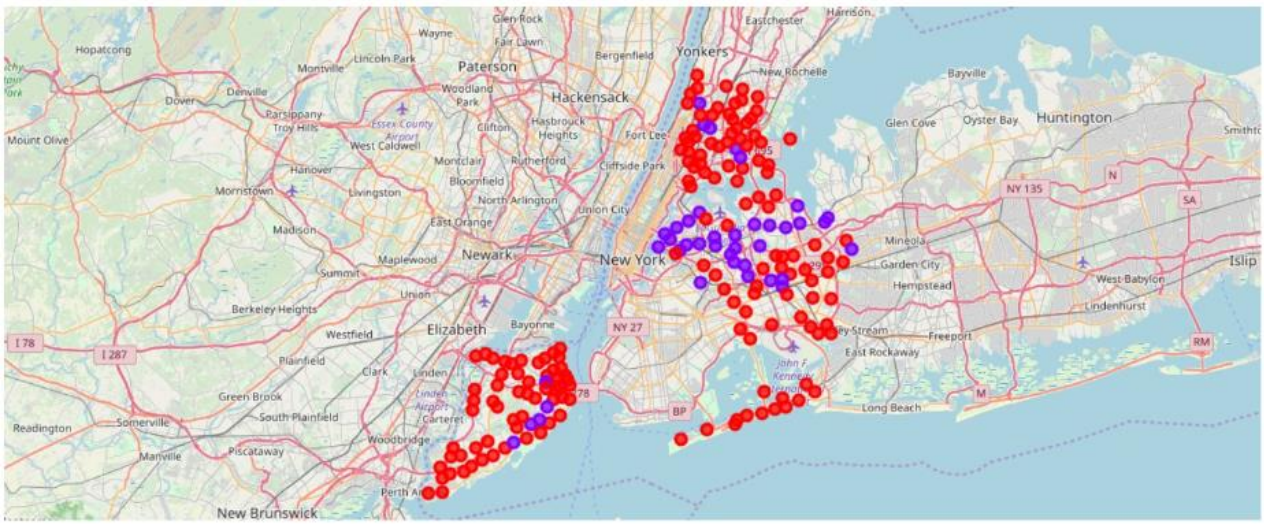
The Total and Total Sum of cluster0 has the smallest value. It shows that the market is not saturated.

Cluster1

The Total and Total Sum of cluster1 has the highest value. It shows that the market is saturated with a high number of restaurants. Our conclusion is that all markets in Manhattan and Brooklyn have been tapped.

Staten Island, Queens and the Bronx

Below we can see different clusters created using K-Means for Queens, the Bronx and Staten Island.



Cluster0

Total and Total Sum of cluster0 has the smallest value. This shows that the market is not saturated thus it has untapped neighbourhoods. Some of the untapped neighbourhoods include:

	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

Cluster1

The Total and Total Sum of Cluster1 has the highest value that shows that the market is saturated with a high number of restaurants.

5. Conclusion & discussion

There is room for increasing the number of Farmers Markets in Bronx, Queens and Staten Island. Additionally, there is scope for exploring cuisines of different nations in Staten Island, Queens and the Bronx. Another conclusion is that Manhattan and Brooklyn have many restaurants with cuisines from multiple countries and nations. There is clear evidence that people in all 5 Boroughs love to eat cuisines from different countries. It can be concluded from the data that opening a restaurant in Manhattan and Brooklyn would be a risky venture prone to failure given the high number of restaurants.

We performed this analysis on a limited data sample. If there is a large enough dataset, we can assume that we would get more precise results. We would recommend to the management of RESTAURANT LLC to explore one of the neighbourhoods in Queens, Staten Island and the Bronx as a suitable location for an East African restaurant.