The Battle of Neighborhoods

Manuel José Naranjo

April 16, 2021

1. Introduction

1.1   Background

New York City is one of the most cosmopolitan cities in the world, full of different nationalities and with constant tourism. One of the attractions of New York is that you can find a variety of Restaurants with different cuisines and different purposes such as fast food, gourmet food, take and go, etc. Being such a big city and with many different people makes it an interesting place to open a new restaurant.

1.2   Problem

As mentioned before, New York is an interesting place to open a new restaurant for any restaurant owner. The thing is that one of the biggest problems when deciding to start a new business and mostly common on the food industry, is deciding where you should open your restaurant. Taking a good or bad decision will significantly impact your business depending on the place you set it.

1.3   Interest

The principal interest of the project developed in this report is that it will help any food entrepreneur to make the correct decision of where to set up his business considering different factors that can impact the development of the business.

2. Data Acquisition and Cleaning

2.1 Data Sources

For the development of this project, we will be using only data about New York, US. The data that we will be using will consider the different neighborhoods, boroughs and its latitude and longitude respectively. Also, we will be using the venues data retrieved from Foursquare to analyze how the consumer is qualifying the different Restaurants in every area of study.

2.2   Data Sets

2.2.1 New York Data (Neighborhoods and Boroughs)

This data set was given to us in Week 3 of the course, it contains a lot of different data of New York City. For our project we will only be using the data about the Neighborhoods and Boroughs of the City. Once we processed and clean the data, we were able to generate the Data Frame composed with 5 different Boroughs and 306 different Neighborhoods, the first 5 rows are presented on Table 1.

For the purpose of the project we segmented this data in 5 different Data Frames corresponding for each Borough. The reason to do this was mainly because this way we can separate the borough preferences for any new owner at the moment of deciding the location.

	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

Table 1: New York Data Frame.

2.2.2 Venues Data (Foursquare)

Once we have all the data segmented for New York City, we used the Foursquare API to get the information of every neighborhood venue. This way we were able to analyze the customers preferences in each location so we can give the new owner a precise recommendation of a good location for its new restaurant depending on different factors.

3. Exploring Data Sets

As we mentioned before, the New York Data Set is composed by 5 different boroughs and 306 different neighborhoods, each neighborhood has his own latitude and longitude. We segmented this data frame in 5 new Data Frame containing the information for each Borough. The Result was a total of 40 neighborhoods for Manhattan, 81 neighborhoods for Queens, 52 neighborhoods for Bronx, 70 neighborhoods for Brooklyn and 63 neighborhoods for Staten Island.

Once we segmented the data, we called the Foursquare API to get the top 10 most common venues for each Neighborhood in each Borough. an example of the resulting data frame is presented in table 2.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Battery Park City	Park	Coffee Shop	Hotel	Memorial Site	Playground	Gym	Clothing Store	BBQ Joint	Burger Joint	Beer Garden
1	Carnegie Hill	Coffee Shop	Café	Yoga Studio	Bookstore	French Restaurant	Cosmetics Shop	Gym	Pizza Place	Wine Shop	Gym / Fitness Center
2	Central Harlem	Public Art	Gym / Fitness Center	Seafood Restaurant	Bar	Fried Chicken Joint	Art Gallery	French Restaurant	Chinese Restaurant	African Restaurant	American Restaurant
3	Chelsea	Coffee Shop	Bakery	French Restaurant	American Restaurant	Art Gallery	Nightclub	Italian Restaurant	Wine Shop	Seafood Restaurant	Ice Cream Shop
4	Chinatown	Bakery	Chinese Restaurant	Cocktail Bar	Hotpot Restaurant	American Restaurant	Spa	Dessert Shop	Salon / Barbershop	Ice Cream Shop	Boutique

Table 2: Top 10 most common venues for Manhattan neighborhoods.

3.1 K-Means clustering model

Now that we have all the necessary data, we decided to run a K-Means clustering model.

The reason of using a K-Means model is because it is an unsupervised model that will give us a solution looking to minimize the intra-cluster distance and maximize de inter-cluster distance. This solution will give us the neighborhoods that seems similar to each other considering the interest of the costumer venues for each neighborhood. This will make us notice which neighborhoods are more interesting to open a new restaurant depending on its characteristics (type of cuisine, fast food or gourmet, etc.)

We set the k parameter to 5 so we get 5 different cluster for each borough. An example of the results of the model is displayed on table 3.

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bronx	Wakefield	40.894705	-73.847201	1	Pharmacy	Dessert Shop	Deli / Bodega	Ice Cream Shop	Laundromat	Donut Shop	Sandwich Place	Accessories Store	Outlet Store	Paper / Office Supplies Store
1	Bronx	Co-op City	40.874294	-73.829939	2	Pizza Place	Accessories Store	Basketball Court	Donut Shop	Restaurant	Fast Food Restaurant	Pharmacy	Park	Bus Station	Discount Store
2	Bronx	Eastchester	40.887556	-73.827806	2	Bus Station	Caribbean Restaurant	Diner	Deli / Bodega	Seafood Restaurant	Food & Drink Shop	Bus Stop	Bowling Alley	Pizza Place	Platform
3	Bronx	Fieldston	40.895437	-73.905643	0	Plaza	River	Bus Station	Medical Supply Store	Accessories Store	Peruvian Restaurant	Nightclub	Other Great Outdoors	Outlet Store	Paper / Office Supplies Store
4	Bronx	Riverdale	40.890834	-73.912585	1	Bus Station	Park	Plaza	Playground	Gym	Home Service	Medical Supply Store	Baseball Field	Bank	Accessories Store

Table 3: k-means cluster made on Bronx neighborhoods.

With this new data frames, we were able to display a map of the clusters made by the model. One example of this is displayed on section 4.

The last thing we did with our data frame was to generate a table to display the top 10 most common venues on each Borough. To realize this, we make a count for each venue. The example of this is displayed on section 4.

4. Results

As you can see on section 3 of this paper, one of the results that we obtain was the cluster for each Borough of our data set. We display this cluster on each Borough map represented below.
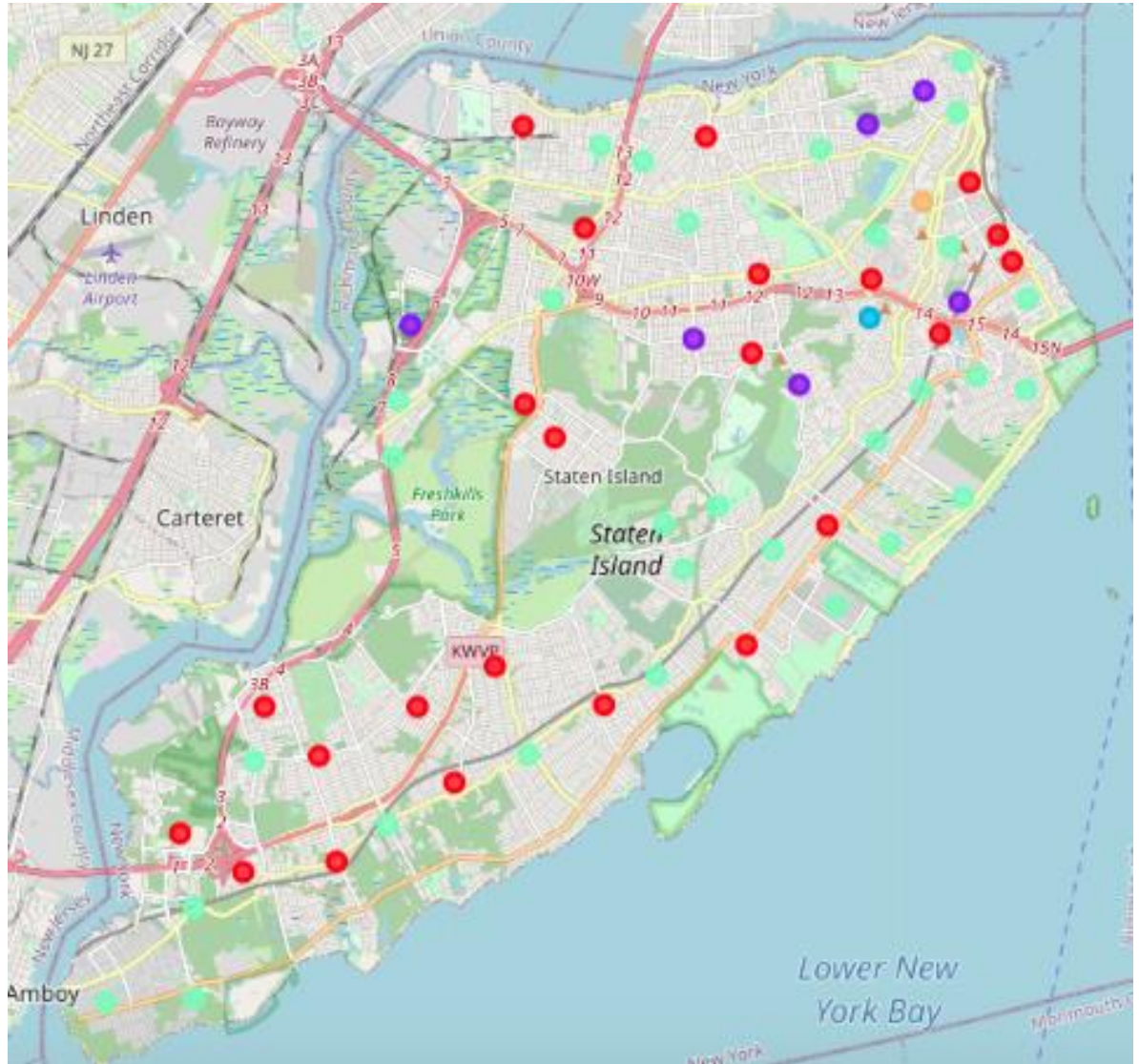

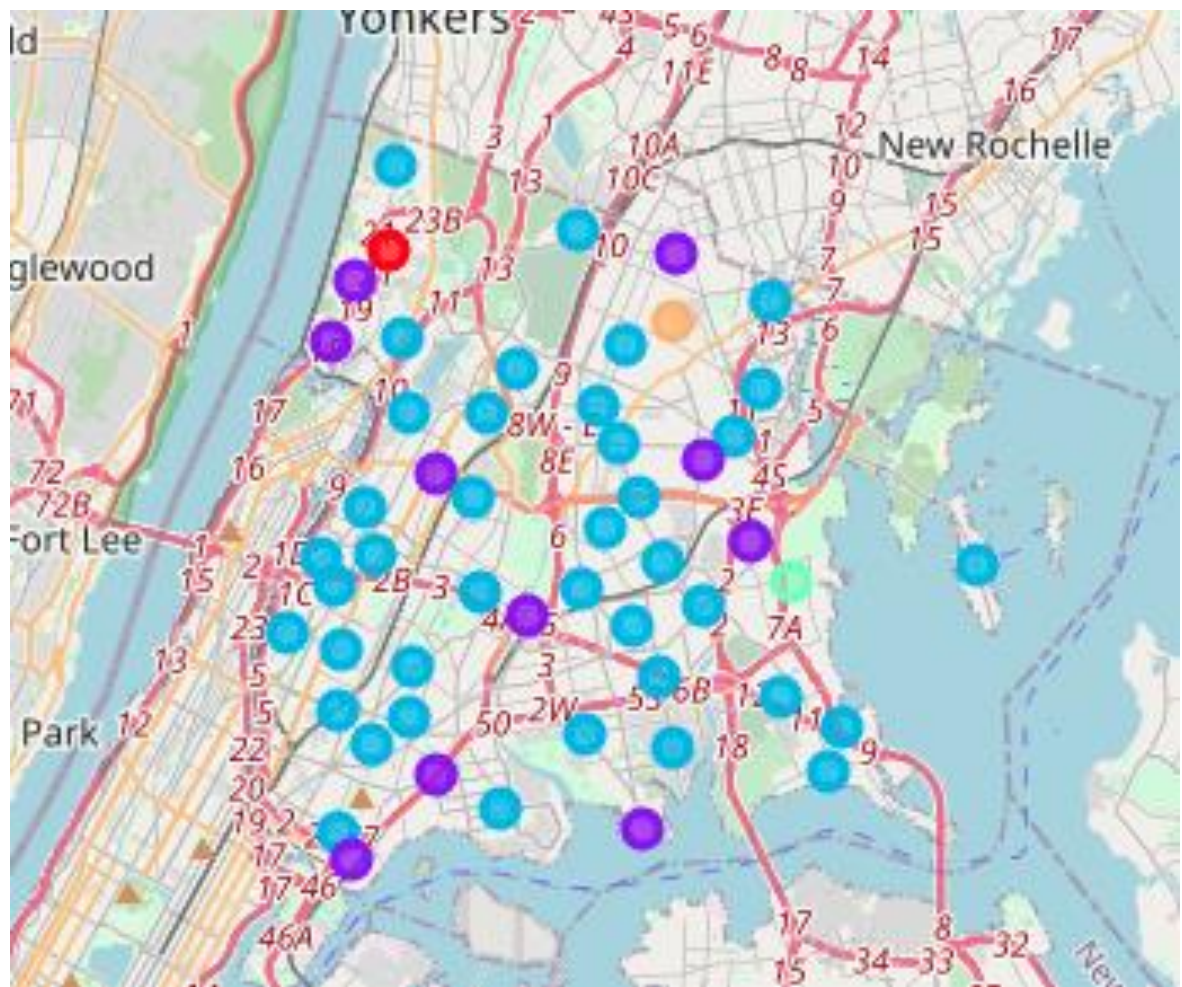*Image 1: Staten Island Neighborhoods cluster map.*
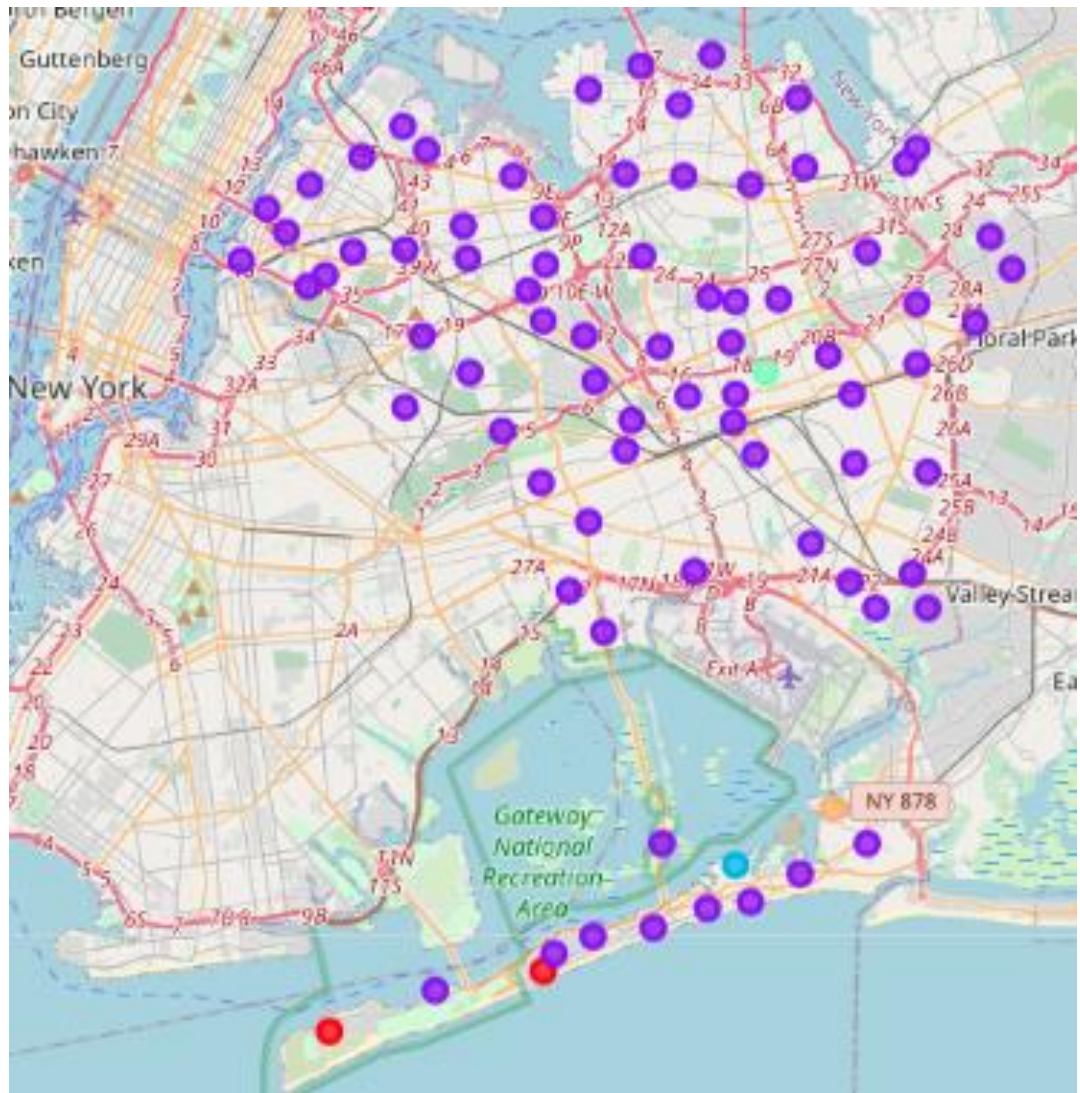
Image 2: Bronx Neighborhoods cluster map.
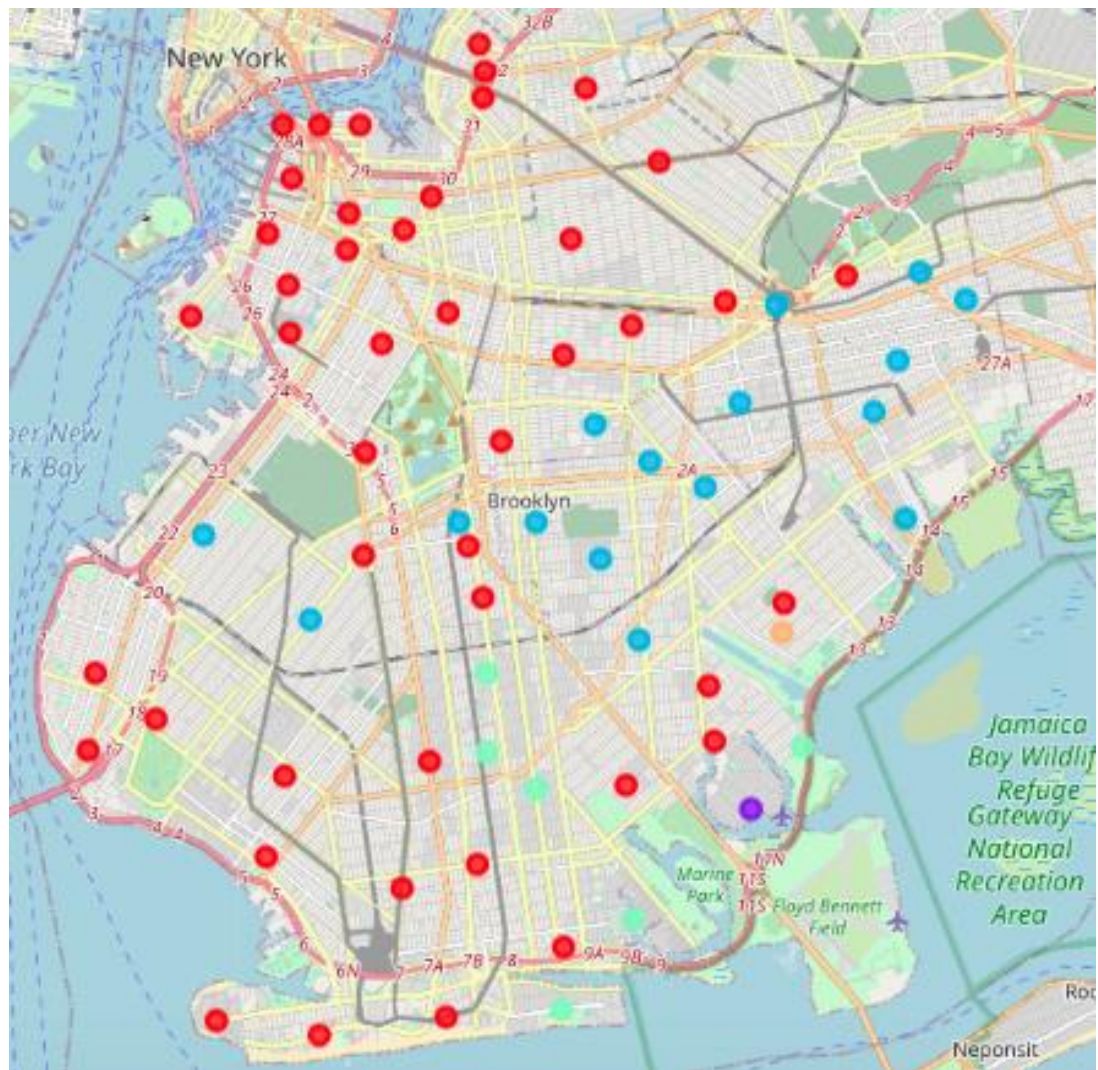
Image 3: Queens Neighborhoods cluster map.
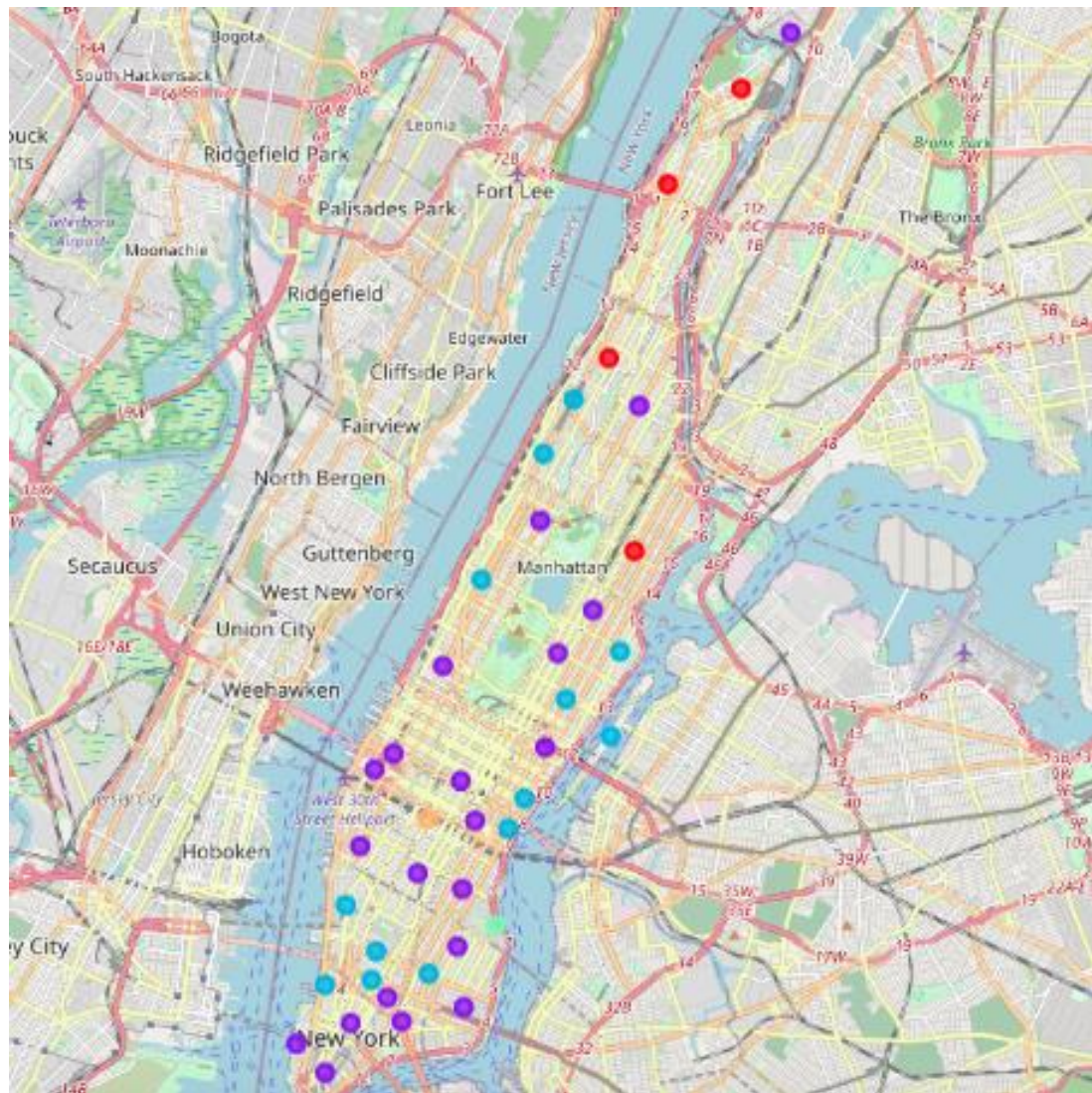
Image 4: Brooklyn Neighborhoods cluster map.

Image 5: Manhattan Neighborhoods cluster map.

Also as mentioned before, we separate for each Borough the top 10 venues considering the costumers preferences. The results were represented in bar chart graphics that are displayed below.
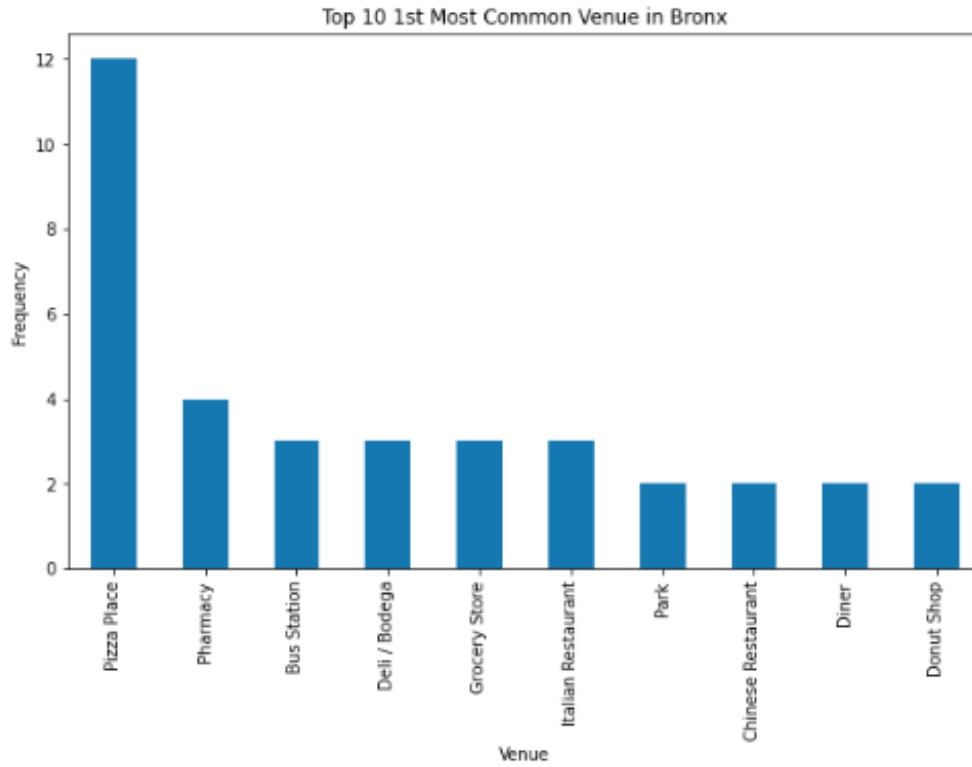
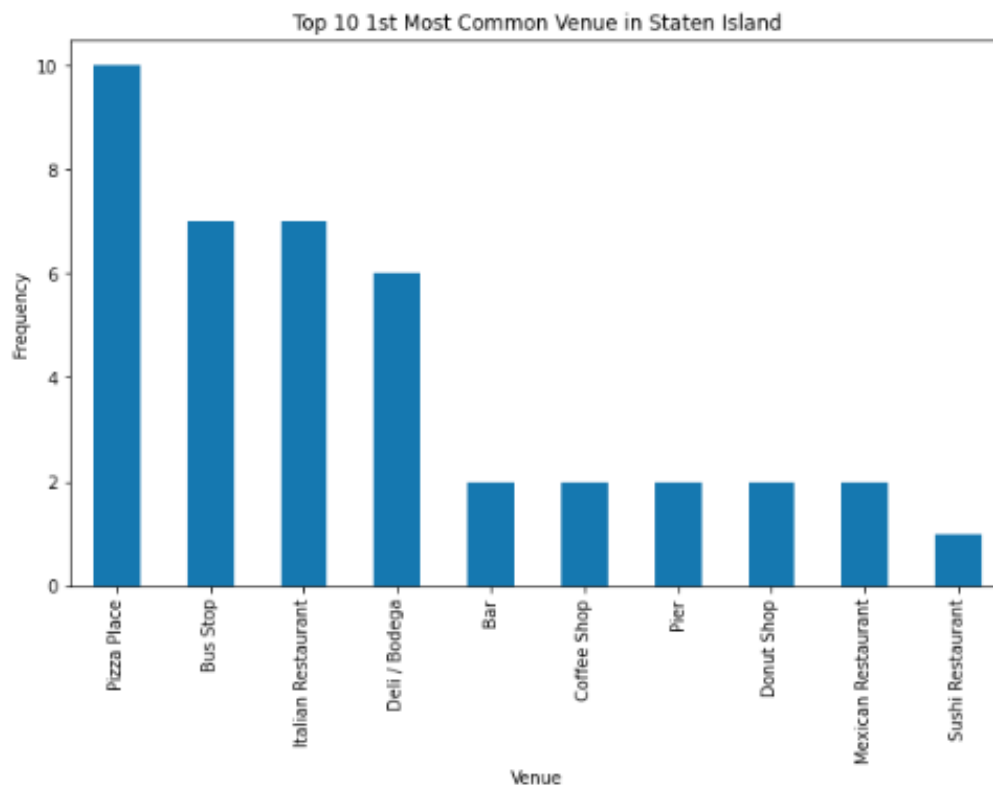*Image 6: top 10 most common venues in Bronx borough.*

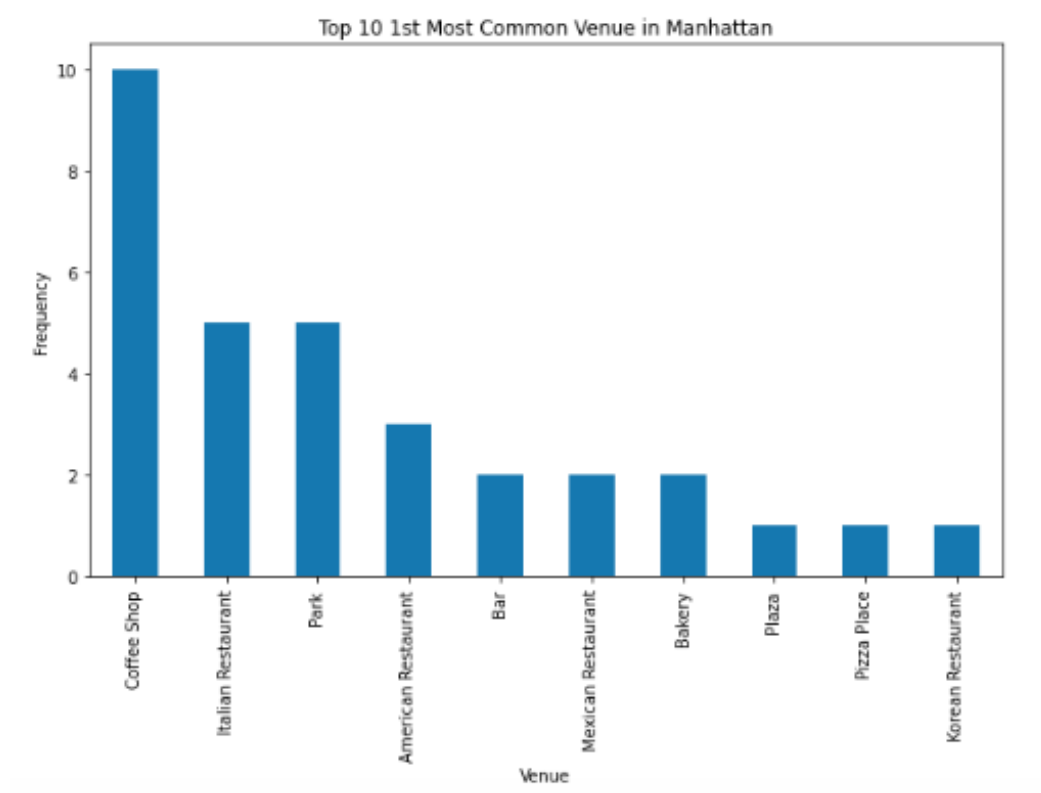*Image 7: top 10 most common venues in Staten Island borough.*

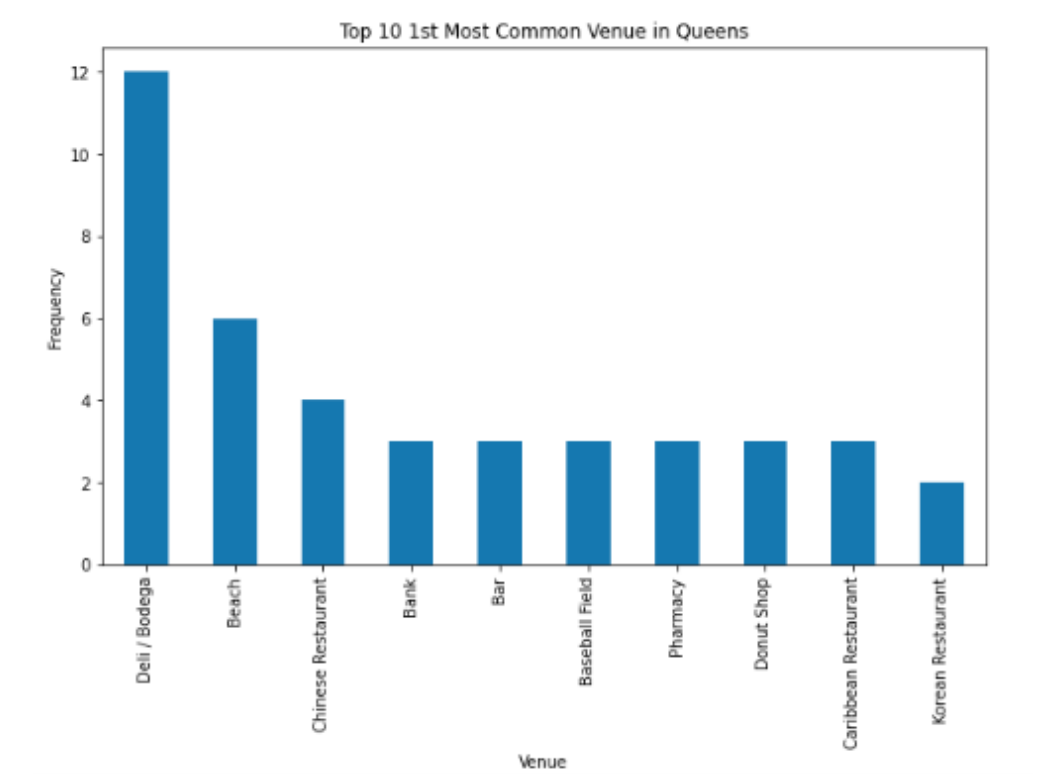*Image 8: top 10most common venues in Manhattan borough.*

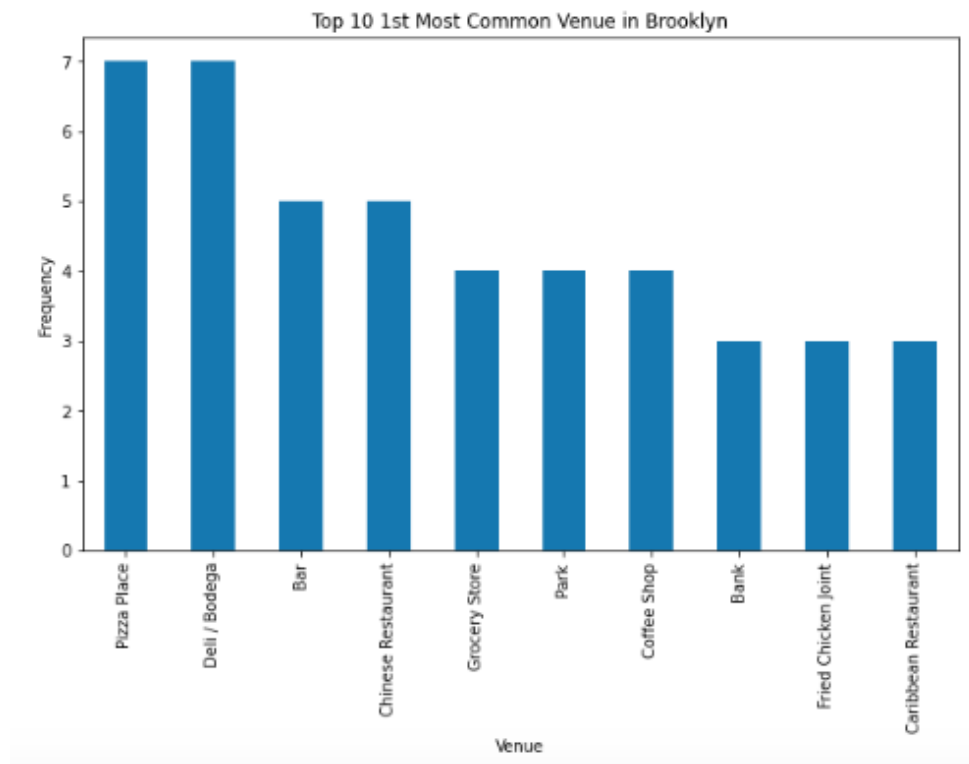*Image 9: top 10 most common venues in Queens borough.*

Image 10: top 10 most common venues in Brooklyn borough.

5. Conclusions and Recommendations

Considering all the work presented on this report. We can conclude that data science tools can help us to make relevant, fundamental and critical decisions in different industries that can affect in one way or another how our company, job, etc. will be affected in the future. Considering the application made on this report, the selection of the place to open a new restaurant is fundamental for the business development. Using the data science tools we can realize the places in which our type of cuisine or the type of restaurant would be more valuable and in which places it will not be considered as fundamental.

One of the recommendations I would give to any entrepreneur looking to start a new job is considering the data of the industry, market, place, etc. before he gets involved. Decisions could be taken much more efficiently when you realize an study of all the factors involved in the area.