The Neighborhoods of New York

Carlos Jimenez

April 10, 2020

1. Introduction

1.1. Background

I chose to analyze New York City. It is the largest and most influential American metropolis. New York City is in reality a collection of many neighborhoods scattered among the city's five boroughs: The Bronx, Brooklyn, Manhattan, Queens, and Staten Island, each exhibiting its own characteristics and ways of life. They say that moving from one neighborhood to another one may be like moving out to a different country. Therefore, it is advantageous to know how similar a neighborhood is to another one in each borough.

1.2. Problem

This project aims to compare by similarity each neighborhood inside each borough, making clusters of neighborhoods, in order to learn which neighborhoods are similar, and which ones are substantially different. Having this information, I will also compare the results by analyzing the clusters distribution between each borough. Finally, I will give a general analysis of the complete New York City, comparing all the neighborhoods in the city.

1.3. Interest

The following set of analyses may be useful to those moving to NYC, or moving from one neighborhood to another one within NYC. This would be especially helpful for those looking to move closer to an area with venues in their desired line of work. Real estate agents looking to improve their suggestions to clients may also find these analyses useful. In order to provide more tailored recommendations, real estate agents may use the venue information to match clients to areas that fit their desired job title and lifestyle.

2. Data acquisition and cleaning process

2.1. Data sources

The data was acquired through the city of New York Open Data team, published in the following site: https://opendata.cityofnewyork.us/. I selected the dataset named **Neighborhood Names GIS**. The raw data can be found in the following link: https://data.cityofnewyork.us/City-Government/Neighborhood-Names-GIS/99bc-9p23. The data contains the The_geom, object id, name, stacked, borough, Annoline1, Annoline2, Annoline3, AnnoAngle columns of each neighborhood. The_geom column contains the geolocation data, the borough is the name of the borough where the neighborhood belongs to, the Annolines columns are the names of the neighborhoods

word by word and the stacked column is the amount of words in the neighborhood name. I will use this data to classify each of the neighborhoods. I will also use the foursquare API to retrieve a list of venues nearby each neighborhoods to make the analysis.

2.2. Data cleaning

The data was downloaded but to work with the dataset I had to make a few changes in the dataset. First of all, I had to eliminate some columns that wouldn't contribute at all with the analysis like stacked, Annoline1, Annoline2, Annoline3, and AnnoAngle, they information that this columns give are redundant for the analysis. After dropping each of these columns there was another problem: the_geom (the geolocation column) had a format of <POINT (Longitude, Latitude)>. To use this data point properly, we needed it in two columns: one for latitude and another one for longitude. To fix this data point I:

1. Eliminated all the POINT word and the parentheses of each data point.
2. Separate the longitude and the latitude in different columns for each row.
3. After doing all this the data frame was ready.

Having cleaned the unnecessary parts of the data, I divided the data frame in six different frames, one for each borough and the final one for the whole list of neighborhoods in New York.

I had to remove some neighborhoods in the data frame: One in Staten Island and one in New York called "Chelsea". The first one didn't have any nearby venues, so it couldn't be compared to other neighborhoods. The second one was duplicated in the dataset, so I had to remove one instance.

3. Methodology

3.1. Clustering

I will use the k-means method to cluster each of the neighborhoods. First, I used the foursquare API to obtain the venues. Having the list of venues, I applied an onehot coding procedure, that consist of putting in binary information the data we are analyzing, in which I classified what type of venues each were. Following this, I now had to group each venue in the list with the neighborhood it belonged to and with this, I was able to calculate in what frequency each type of venue appeared in each neighborhood. This frequency of each type of venue can be used to classify each neighborhoods top ten most common venues.

Now that I have a data frame with the top type of venues in each neighborhood I was able to proceed to the clustering process. For this type of process the number of cluster will determine the number of different groups that each neighborhood will associate with. So for each borough I decide to create 5 cluster and for the full clustering of the neighborhoods of New York I decided to work with 8 cluster, the reason being is the amount of neighborhoods is much larger, and we can assume we will need more groups to classify the neighborhoods.

Finally plotting in a map the k-means clustering we can show which neighborhoods and their top picks are in each cluster. The analysis of each cluster and look at their properties and comment which cluster is the most common to find in each borough and in New York.

3.2. Bronx Analysis

In the map clustering we can see that two clusters are found more commonly that the others. The cluster number 3 is the one that we can see the most, this cluster contains Italian restaurants and food places. The Bronx could be a good place for family with Italian roots or that love this kind of food.
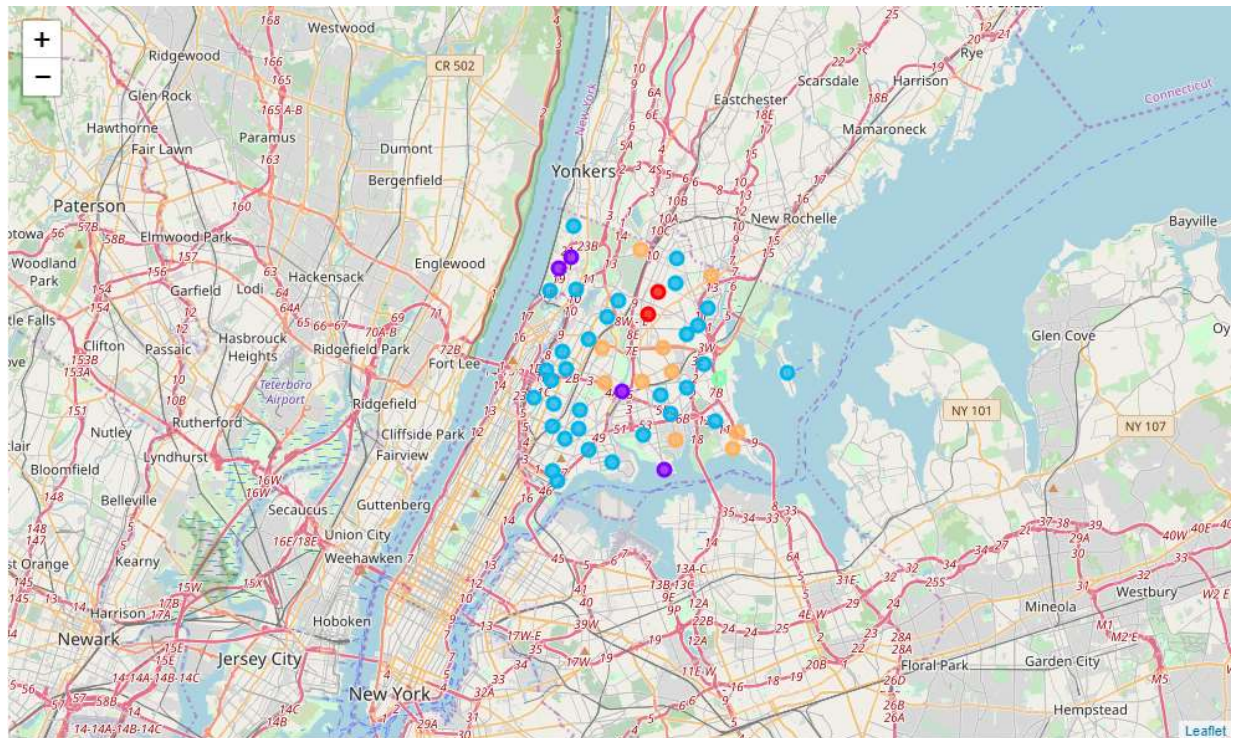

Figure 1 Bronx K-means Clustering

	Name	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
2	Throgs Neck	Deli / Bodega	Coffee Shop	Pizza Place	Italian Restaurant	Baseball Field	Bar	Liquor Store	Juice Bar	Asian Restaurant	American Restaurant
4	Parkchester	Supermarket	Pizza Place	Asian Restaurant	American Restaurant	Kids Store	Bank	Mobile Phone Shop	Discount Store	Restaurant	Plaza
5	Westchester Square	Fast Food Restaurant	Pharmacy	Donut Shop	Sandwich Place	Pizza Place	Park	Dance Studio	Health & Beauty Service	Mobile Phone Shop	Latin American Restaurant
6	Van Nest	Pizza Place	Deli / Bodega	Middle Eastern Restaurant	Furniture / Home Store	Bakery	Bus Line	Bus Station	Bus Stop	Caribbean Restaurant	Food Truck
7	Morris Park	Pizza Place	Deli / Bodega	Burger Joint	Bakery	Supermarket	Jewelry Store	Coffee Shop	Juice Bar	Bar	Buffet
8	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Dessert Shop	Mexican Restaurant	Donut Shop	Grocery Store	Bank	Liquor Store
10	North Riverdale	Pizza Place	Bank	Italian Restaurant	Ice Cream Shop	Pharmacy	Coffee Shop	Donut Shop	Chinese Restaurant	Pool	Sandwich Place
12	Pelham Bay	Italian Restaurant	Bank	Fast Food Restaurant	Gym / Fitness Center	Convenience Store	Cosmetics Shop	Diner	Mexican Restaurant	Sandwich Place	Donut Shop
13	Edgewater Park	Italian Restaurant	Pizza Place	Ice Cream Shop	Japanese Restaurant	Park	Liquor Store	Sports Bar	Chinese Restaurant	Bar	Pub
14	Olinville	Supermarket	Liquor Store	Caribbean Restaurant	Deli / Bodega	Chinese Restaurant	Fried Chicken Joint	Basketball Court	Furniture / Home Store	Food	Donut Shop

Figure 2 Bronx Most Common Cluster (Number 3) (Not full list of neighborhoods)

3.3. Brooklyn Analysis

In the map we can see two cluster again like the Brooklyn clustering, this time the cluster 4 is the most common. With restaurants, banks and grocery stores are common urban area venues. People looking to live close to convenience venues will have a lot of options in Brooklyn.
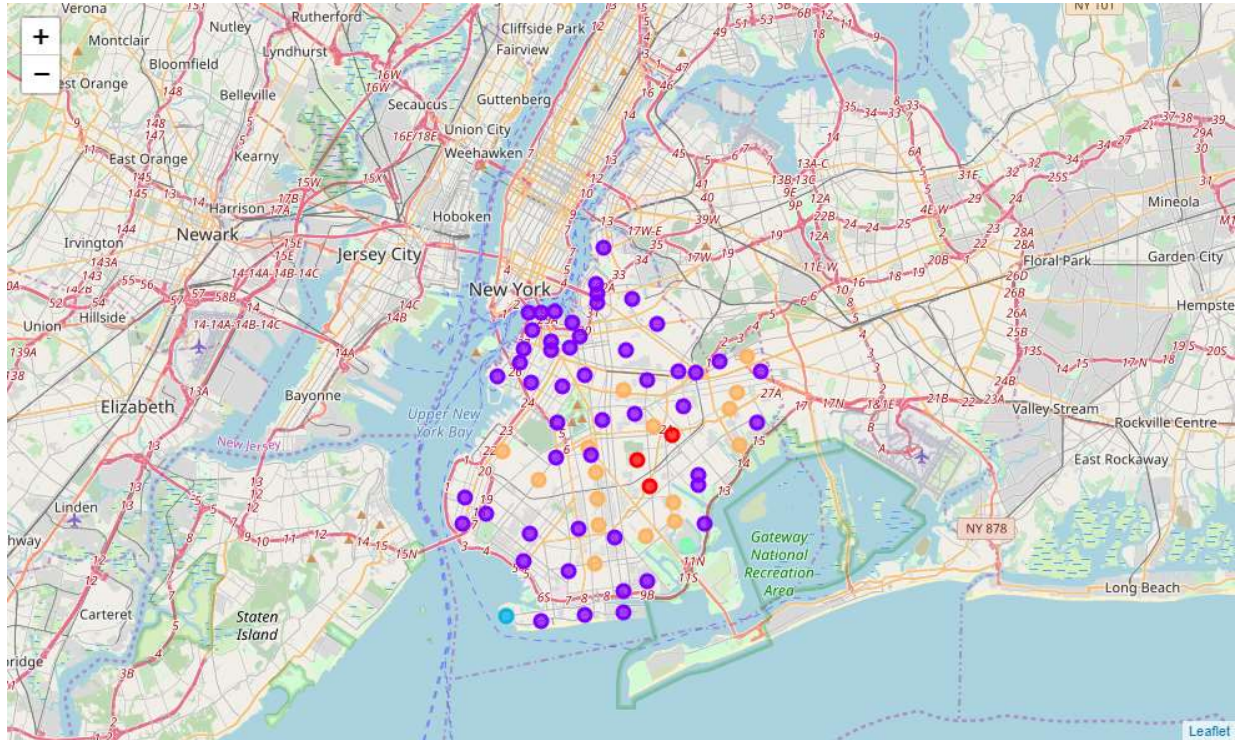

Figure 3 Brooklyn K-means Clustering

	Name	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	East Williamsburg	Bar	Deli / Bodega	Cocktail Bar	Coffee Shop	Bakery	Music Venue	Concert Hall	Mexican Restaurant	Donut Shop	Vegetarian / Vegan Restaurant
1	Spring Creek	Discount Store	Women's Store	Kids Store	Hardware Store	Shopping Mall	Shoe Store	Burger Joint	Pizza Place	Mobile Phone Shop	Caribbean Restaurant
2	Georgetown	Bank	Pharmacy	Breakfast Spot	Donut Shop	Mexican Restaurant	Shipping Store	Supermarket	Frozen Yogurt Shop	Miscellaneous Shop	Shopping Mall
3	Prospect Park South	Caribbean Restaurant	Pizza Place	Grocery Store	Fast Food Restaurant	Mobile Phone Shop	Latin American Restaurant	Fried Chicken Joint	Donut Shop	Mexican Restaurant	Women's Store
4	Midwood	Pizza Place	Ice Cream Shop	Video Game Store	Field	Candy Store	Electronics Store	Pharmacy	Bakery	Moving Target	Convenience Store
5	North Side	Coffee Shop	Pizza Place	American Restaurant	Wine Bar	Bakery	Bar	Jewelry Store	Yoga Studio	Burger Joint	Food Truck
7	City Line	Donut Shop	Fried Chicken Joint	Grocery Store	South American Restaurant	Mobile Phone Shop	Food	Flower Shop	Shoe Store	Fast Food Restaurant	Bus Stop
8	Ocean Hill	Deli / Bodega	Grocery Store	Chinese Restaurant	Southern / Soul Food Restaurant	Convenience Store	Bus Stop	Playground	Fried Chicken Joint	Bus Line	Supermarket
9	Prospect Lefferts Gardens	Bakery	Café	Caribbean Restaurant	Pizza Place	Ice Cream Shop	Indian Restaurant	Sushi Restaurant	Wine Shop	Deli / Bodega	Smoke Shop
10	Boerum Hill	Coffee Shop	Bar	Furniture / Home Store	Spa	Grocery Store	Sandwich Place	French Restaurant	Japanese Restaurant	Gift Shop	Dance Studio

Figure 4 Brooklyn Most Common Cluster (Number 4) (Not full list of neighborhoods)

3.4. Manhattan Analysis

In this clustering we can see a dominance of neighborhood cluster again the number 2 consists of restaurants, coffee shops, hotels and gyms. This type of neighborhood can be attractive for young families and teenagers looking for places with a lot of entertainment and different persons thanks to the hotels and bars.
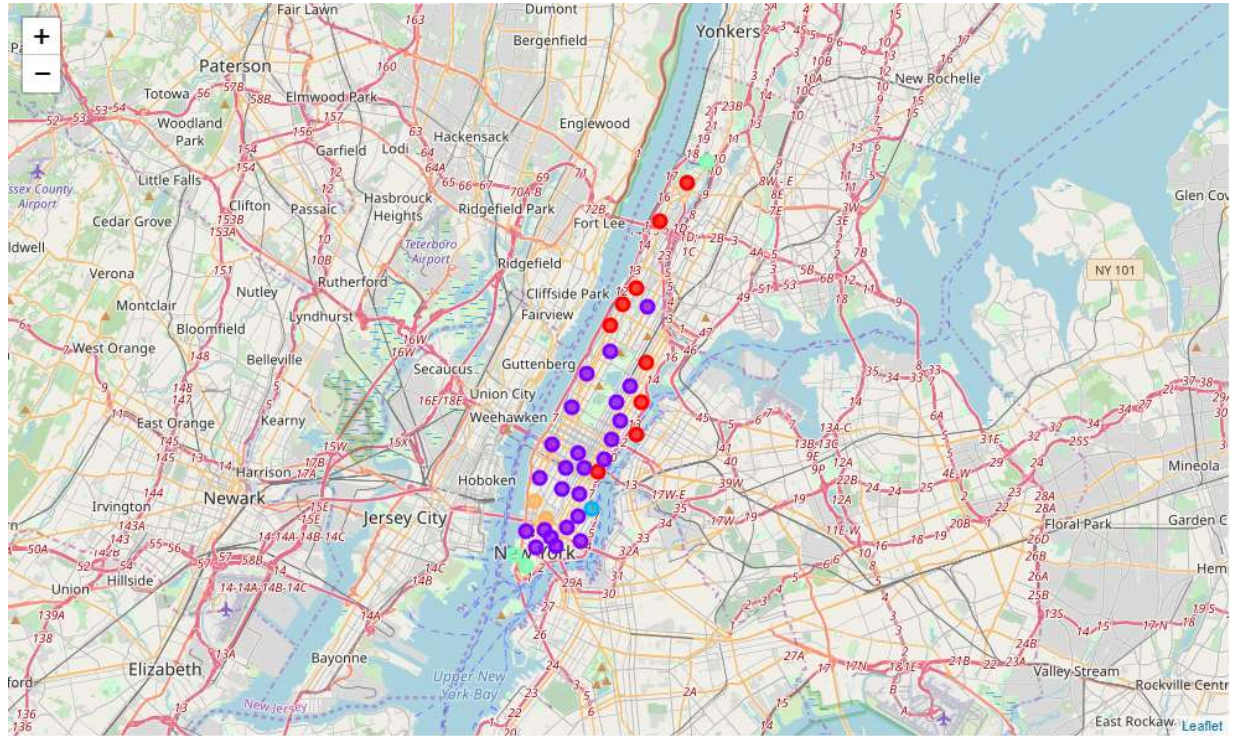

Figure 5 Manhattan K-means Clustering

	Name	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	Flatiron	Yoga Studio	Japanese Restaurant	Gym / Fitness Center	Spa	American Restaurant	Mediterranean Restaurant	Vegetarian / Vegan Restaurant	New American Restaurant	Café	Cycle Studio
1	Civic Center	Coffee Shop	Hotel	Gym / Fitness Center	Spa	Cocktail Bar	French Restaurant	American Restaurant	Sandwich Place	Yoga Studio	Park
3	Turtle Bay	Italian Restaurant	Sushi Restaurant	Coffee Shop	Wine Bar	Steakhouse	French Restaurant	Japanese Restaurant	Park	Café	Hotel
4	Sutton Place	Italian Restaurant	Gym / Fitness Center	Gym	Coffee Shop	American Restaurant	Indian Restaurant	Furniture / Home Store	Park	Beer Garden	Chinese Restaurant
5	Midtown South	Korean Restaurant	Japanese Restaurant	Hotel	Dessert Shop	Coffee Shop	Hotel Bar	Salad Place	Restaurant	Burger Joint	Gym / Fitness Center
6	Noho	Italian Restaurant	French Restaurant	Hotel	Cocktail Bar	Coffee Shop	Pizza Place	Sushi Restaurant	American Restaurant	Yoga Studio	Grocery Store
7	Carnegie Hill	Coffee Shop	Café	Pizza Place	Japanese Restaurant	Bookstore	Cosmetics Shop	Bakery	French Restaurant	Gym	Wine Shop
10	West Village	Italian Restaurant	New American Restaurant	American Restaurant	Jazz Club	Park	Wine Bar	Cocktail Bar	Coffee Shop	Theater	Bakery
11	Clinton	Theater	Gym / Fitness Center	Italian Restaurant	Wine Shop	American Restaurant	Hotel	Gym	Coffee Shop	Cocktail Bar	Spa
12	Upper East Side	Coffee Shop	Italian Restaurant	Exhibit	Bakery	Gym / Fitness Center	Juice Bar	Yoga Studio	Art Gallery	French Restaurant	Hotel

Figure 6 Manhattan Most Common Cluster (Number 2) (Not full list of neighborhoods)

3.5. Queens Analysis

This time we can see a clear dominance of the cluster number 7 that consist of donut shops, Chinese restaurants and delis and bodegas, but this cluster doesn't have a clear pattern we can look at. We can see a bit of Asian influence and can be attractive to people looking for this kind of venues, or by the sea side looking for site near to beaches.
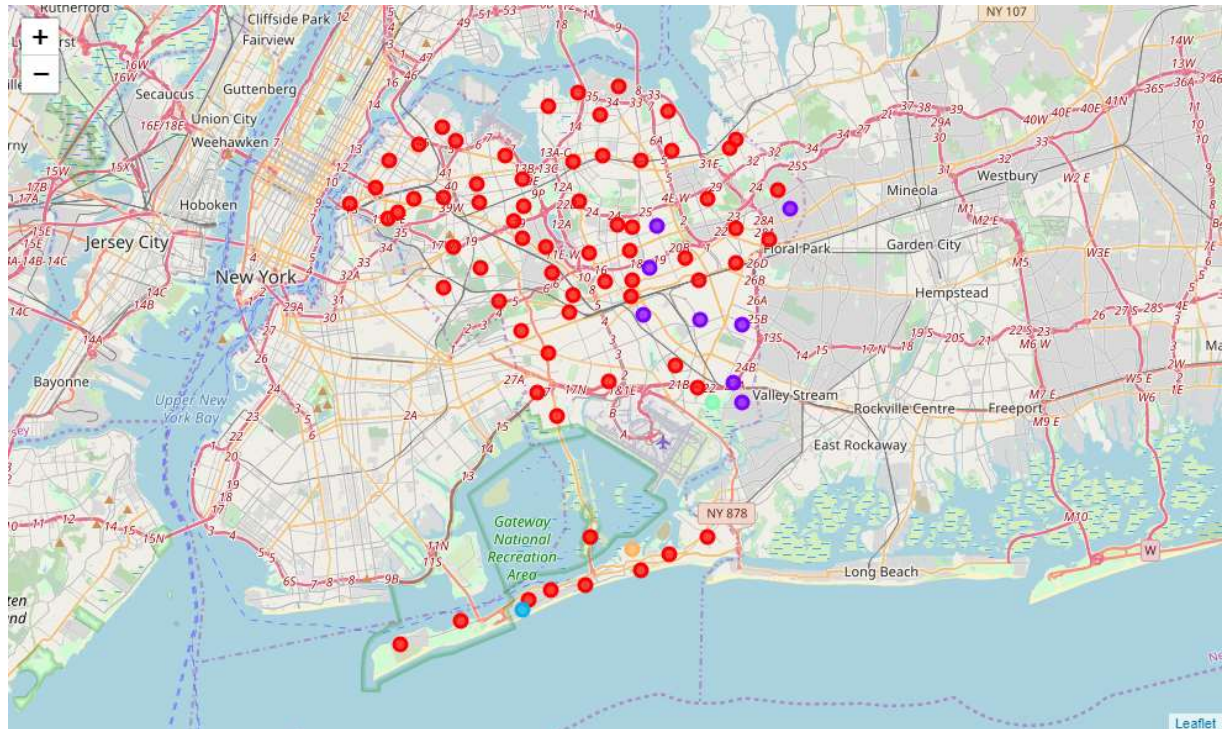

Figure 7 Queens K-means Clustering

	Name	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	Queensboro Hill	Chinese Restaurant	Bank	Bakery	Bus Station	Intersection	Dumpling Restaurant	Frozen Yogurt Shop	Bar	Bagel Shop	Park
2	Rockaway Park	Beach	Donut Shop	Steakhouse	Pizza Place	Bank	Bar	Bagel Shop	Japanese Restaurant	Seafood Restaurant	Sandwich Place
3	Belle Harbor	Beach	Deli / Bodega	Pub	Spa	Bagel Shop	Chinese Restaurant	Mexican Restaurant	Boutique	Italian Restaurant	Bakery
4	Lefrak City	Department Store	Cosmetics Shop	Bakery	Pharmacy	Restaurant	Shopping Mall	Furniture / Home Store	Fruit & Vegetable Store	Fried Chicken Joint	Steakhouse
5	Lindenwood	Pizza Place	Deli / Bodega	Fruit & Vegetable Store	Liquor Store	Donut Shop	Bakery	Japanese Restaurant	Gym	Bank	Pharmacy
6	Ravenswood	Chinese Restaurant	Grocery Store	Breakfast Spot	Bagel Shop	Food & Drink Shop	Food Truck	Fried Chicken Joint	Brazilian Restaurant	Market	Lounge
7	Hillcrest	Donut Shop	Grocery Store	College Academic Building	Food Court	Market	Shipping Store	Coffee Shop	College Basketball Court	College Stadium	College Cafeteria
9	Bellaire	Deli / Bodega	Convenience Store	Italian Restaurant	Chinese Restaurant	Bus Station	Moving Target	Greek Restaurant	Gym	Breakfast Spot	Health & Beauty Service
10	Middle Village	Sandwich Place	Sports Bar	Bakery	Discount Store	Diner	Thrift / Vintage Store	Italian Restaurant	Dessert Shop	Spanish Restaurant	Bank
11	Forest Hills Gardens	Bakery	Grocery Store	Sandwich Place	Pharmacy	Park	Sushi Restaurant	Chinese Restaurant	Playground	Pizza Place	Plaza

Figure 8 Queens Most Common Cluster (Number 5) (Not full list of neighborhoods)

3.6. Staten Island Analysis

In Figure 9, we can see that there is clearly a cluster that is more common that the others. In this specific cluster that is the number 1, we can see that most of this venues are coffee shops, some restaurants, mostly Chinese and American ones. Common people that like family restaurants and Chinese food can look forward to living in Staten Island that most of the neighborhood have this characteristics.
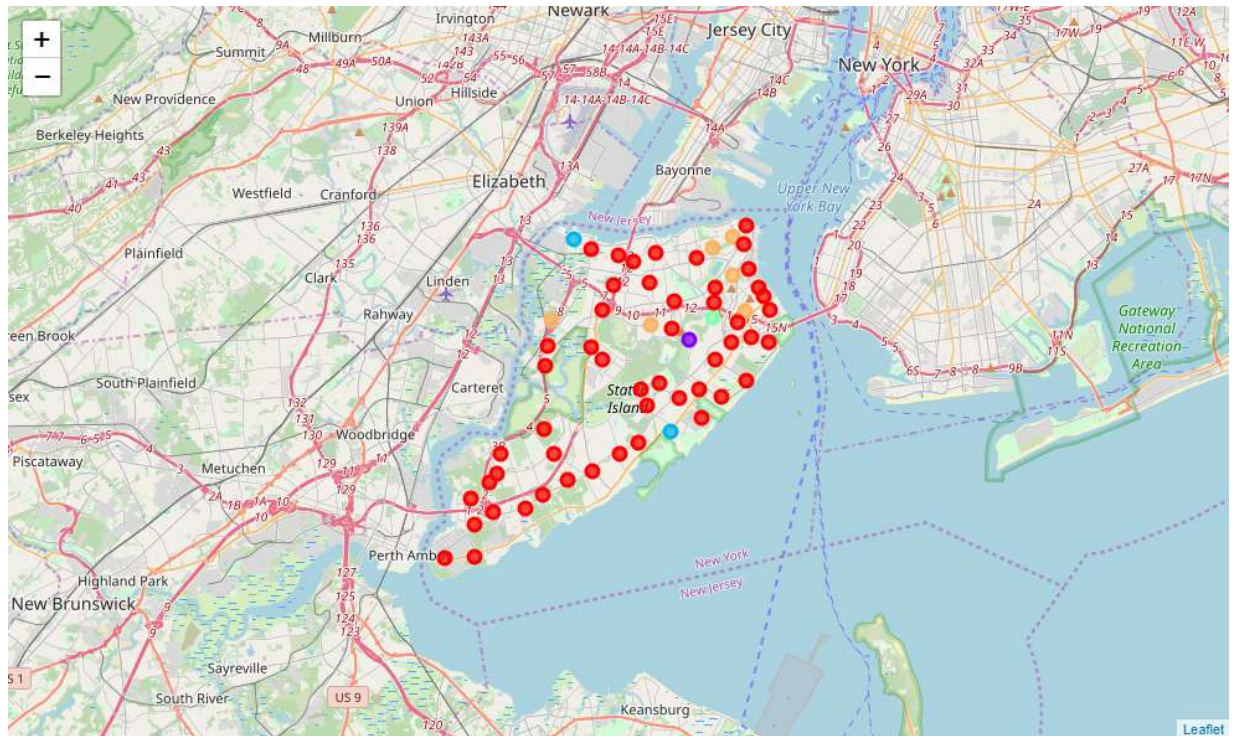

Figure 9 Staten Island K-means Clustering

	Name	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
1	Todt Hill	Park	Yoga Studio	Gas Station	French Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
3	Port Richmond	Deli / Bodega	Pizza Place	Donut Shop	Rental Car Location	Ice Cream Shop	Food Truck	Food	Flower Shop	Filipino Restaurant	Fast Food Restaurant
6	Castleton Corners	Pizza Place	Japanese Restaurant	Bagel Shop	Grocery Store	Mini Golf	Tattoo Parlor	Hardware Store	Bar	Bank	Sandwich Place
7	New Springville	Mobile Phone Shop	Coffee Shop	Pizza Place	Health & Beauty Service	Bagel Shop	Optical Shop	Grocery Store	Chinese Restaurant	Sandwich Place	Mexican Restaurant
8	West Brighton	Coffee Shop	Bus Stop	Ice Cream Shop	Bank	Bar	Pharmacy	Breakfast Spot	Music Store	Italian Restaurant	Diner
9	Rosebank	Italian Restaurant	Grocery Store	American Restaurant	Pizza Place	Breakfast Spot	Cajun / Creole Restaurant	Sandwich Place	Restaurant	Ice Cream Shop	Deli / Bodega
10	Silver Lake	American Restaurant	Golf Course	Burger Joint	Dog Run	Farmers Market	French Restaurant	Food Truck	Food & Drink Shop	Food	Flower Shop
12	St. George	Clothing Store	Sporting Goods Shop	Bar	Italian Restaurant	Scenic Lookout	Pharmacy	Snack Place	Burger Joint	Monument / Landmark	Farmers Market
13	Travis	Hotel	Deli / Bodega	Bowling Alley	Comedy Club	Spanish Restaurant	Park	Gym	Gym / Fitness Center	Baseball Field	Sports Club
14	Manor Heights	Deli / Bodega	Bagel Shop	Liquor Store	Pharmacy	Donut Shop	Chinese Restaurant	Campground	American Restaurant	Athletics & Sports	German Restaurant

Figure 10 Staten Island Most Common Cluster (Number 1) (Not full list of neighborhoods)

3.7. New York Analysis

In the map of New York we can see that there is clearly a cluster that is the most common one, the number 3 in our k-means procedure, we can see that most of them have as common place restaurants be it American or Italian and pizza places. In the first ten we can't see a clear pattern and we can't expect to see any clear pattern around this type of cluster.
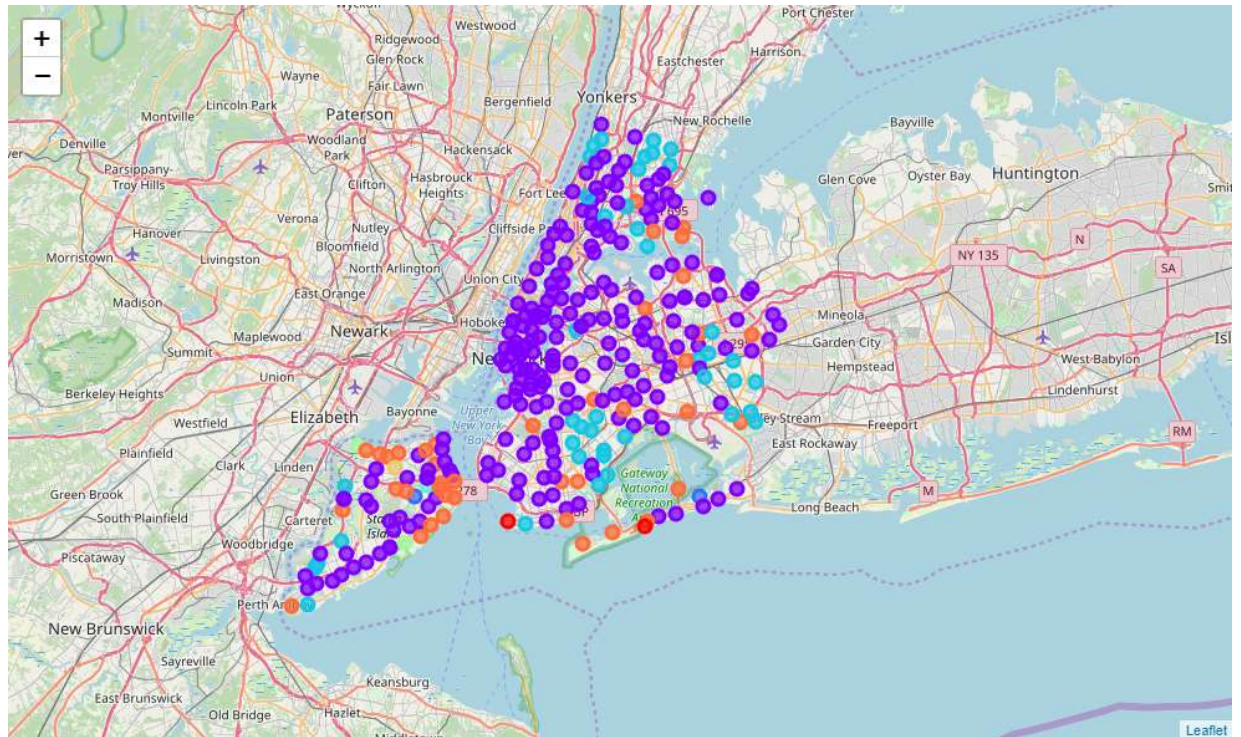

Figure 11 New York K-means Clustering

	Name	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	Wakefield	Pharmacy	Gas Station	Ice Cream Shop	Donut Shop	Dessert Shop	Sandwich Place	Deli / Bodega	Laundromat	Caribbean Restaurant	Factory
3	Country Club	Sandwich Place	Spa	Playground	Yoga Studio	Falafel Restaurant	Electronics Store	Empanada Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant
4	Parkchester	Supermarket	Pizza Place	Mobile Phone Shop	American Restaurant	Bank	Kids Store	Asian Restaurant	Fried Chicken Joint	Shipping Store	Mexican Restaurant
5	Westchester Square	Fast Food Restaurant	Sandwich Place	Pharmacy	Pizza Place	Donut Shop	Indie Theater	Latin American Restaurant	Park	Bank	Bar
7	Morris Park	Pizza Place	Burger Joint	Deli / Bodega	Bakery	Juice Bar	Sandwich Place	Bar	Bank	Coffee Shop	Pharmacy
9	Spuyten Duyvil	Park	Bus Line	Thai Restaurant	Pharmacy	Tennis Stadium	Bank	Intersection	Film Studio	Event Space	Empanada Restaurant
10	North Riverdale	Pizza Place	Italian Restaurant	Bank	Burger Joint	Sushi Restaurant	Grocery Store	Mexican Restaurant	Building	Sandwich Place	Bagel Shop
11	Soundview	Chinese Restaurant	Grocery Store	Playground	Latin American Restaurant	Liquor Store	Discount Store	Video Store	Bus Stop	Pharmacy	Bus Station
12	Pelham Bay	Bank	Italian Restaurant	Donut Shop	Cosmetics Shop	Diner	Fast Food Restaurant	Convenience Store	Gym / Fitness Center	Sandwich Place	Mexican Restaurant

Figure 12 New York Most Common Cluster (Number 3) (Not full list of neighborhoods)

4. Results

 With the k-means clustering it was a success in classifying the different neighborhoods in the boroughs. The clusters in each of the borough most of the neighborhoods fall into one of the labels and the others are left like outlier neighborhoods, so if you are looking for a similar neighborhood in the same borough you will likely be in the common cluster and find a similar neighborhood without a problem, for the outliers it will be a tough search.

 Now analyzing the New York City clustering, the analysis wasn't as effective as the other clustering. This kind of process won't be as effective with data that is so broad. Starting with New York having a lot of stores and restaurants repeated in a lot of neighborhoods, it can be hard to look in an analytic way which neighborhoods are different.

5. Discussion

 Finding a clean and organized dataset is of vital importance to have a smooth work, so it's important to remark the great work of the open data of New York. Having the right tools it's what makes this work possible. Even if the Foursquare API worked this time, it has some minor problems and it would be best to avoid this kind of situations when doing a work, so it's better to study which tools you are going to use before working in the code and use another type of venues API if you have access or knowledge of one.

6. Conclusion

 I was able to achieve the aim of the project in doing a successful clustering of the boroughs. The city of New York have a lot of different features, this study didn't include things like urban areas, atmospheres and lifestyle of the people living in it. So even if the neighborhood similitude is of vital importance for this analysis, to improve this study it might be helpful to add more variables that could be helpful to generate more insights.