IBM DATA SCIENCE CAPSTONE PROJECT

Selecting the best location for a new pizzeria in the Borough of the Bronx or Brooklyn

1. Introduction

The tools given through the course of Data Science certificate, allows to introduce a created like scenario to demonstrate a real-life problem and in search of solutions based in information scrapping and comprehension of data, as a result presenting ways to attack the problem.

2. Problem

This Project is going to involve a theoretical business expansion from an early entrepreneur, whose vision is to expand his little, but successful pizzeria to another neighborhood in the Bronx or Brooklyn, taking into consideration neighborhood's characteristics and competition in each area.

a. Problem Description

A successful pizzeria owner has the intention to further expand his little business located in Bronxdale-Bronx, into the borough of Brooklyn, since it is the nearest located borough.

Ideally our entrepreneur desires to open a new pizzeria in Brooklyn but is afraid that neighborhoods in Brooklyn will not share the same characteristics as the ones in the Bronx, and specially his neighborhood. Also, he is worried that competition for pizzeria places will crushed his little venture. In a worst-case scenario, the entrepreneur is willing to only expand withing Bronx's neighborhoods.

This problem can be generalized to all early entrepreneurs that are in search for new areas to expand their business but have to take into considerations characteristics of neighborhoods to maximize their effort to grow, and not failed by choosing a wrong location.

3. Data

The data used for this problem will be the one provided from the Lab of Battle of Neighborhoods Manhatthan in this course, containing the following data:

   I.   List of Disctricts in New York
  II.   Geo-coordinates of each neighborhood in Brooklyn and Bronx

https://cocl.us/new_york_dataset

Furthermore, using Foursquare API to search for top venues around each neighborhood in the two Boroughs (Brooklyn and Bronx), in order to find common characteristics within each neighborhood.

4. Methodology

Primary, handling the data for each neighborhood and searching for the top venues in each of the neighborhoods will provide useful insights of the characteristics for each area of Boroughs, whether is a highly commercial, residential or business sector. Then applying K-means machine learning for creating clusters of neighborhoods, without taking into consideration their respective Borough
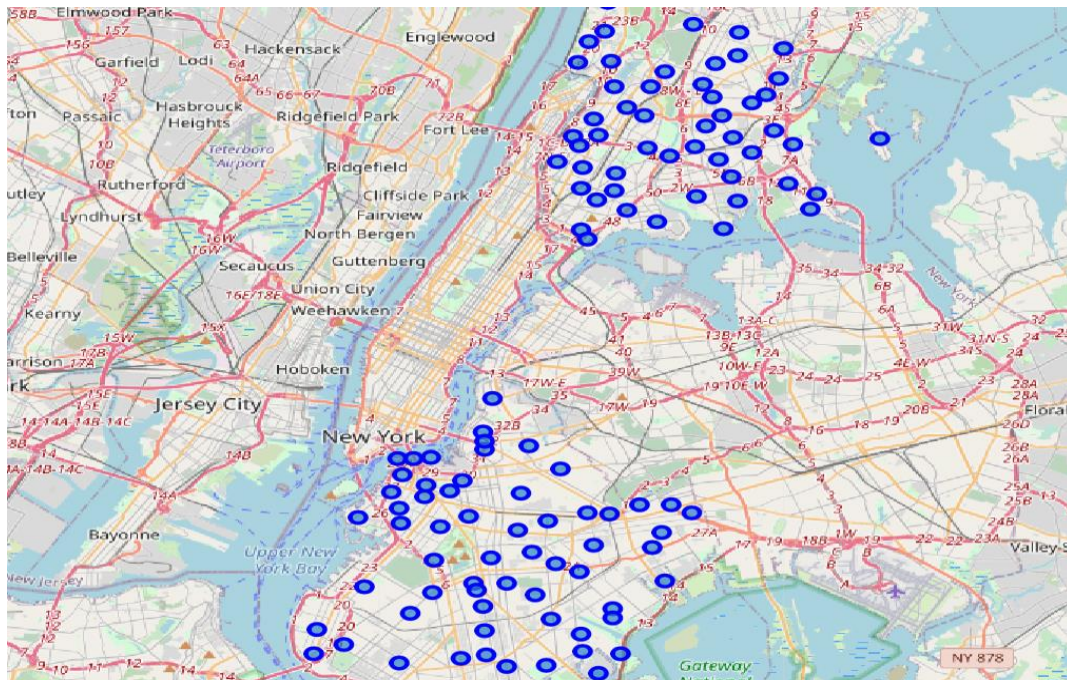
b. Data Preparation and exploration

First, preparing the data and transforming from a json file to a pandas 'Dataframe', then filtering the table of Boroughs in New York, and setting them only to show Bronx and Brooklyn neighborhoods, as shown below:

	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
5	Bronx	Kingsbridge	40.881687	-73.902818
7	Bronx	Woodlawn	40.898273	-73.867315
8	Bronx	Norwood	40.877224	-73.879391
9	Bronx	Williamsbridge	40.881039	-73.857446
10	Bronx	Baychester	40.866858	-73.835798
11	Bronx	Pelham Parkway	40.857413	-73.854756
12	Bronx	City Island	40.847247	-73.786488
13	Bronx	Bedford Park	40.870185	-73.885512
14	Bronx	University Heights	40.855727	-73.910416
15	Bronx	Morris Heights	40.847898	-73.919672
16	Bronx	Fordham	40.860997	-73.896427
17	Bronx	East Tremont	40.842696	-73.887356
18	Bronx	West Farms	40.839475	-73.877745
19	Bronx	High Bridge	40.836623	-73.926102
20	Bronx	Melrose	40.819754	-73.909422
21	Bronx	Mott Haven	40.806239	-73.916100
22	Bronx	Port Morris	40.801664	-73.913221

*For further display of both Boroughs, please refer to the python library.

There are 52 Bronx's neighborhoods (North) and 70 Brooklyn's (South) neighborhoods that will be taking into consideration for the clustering, which are located as follows:

In second place, we call the Foursquare API to search the venues that are located within each neighborhood based on a radius of 400 meters around the block and limit the search to 100, and finally segmenting it to the top 10 venues, for further explorations and clustering, as show below:

	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	Allerton	Pizza Place	Deli / Bodega	Chinese Restaurant	Supermarket	Dessert Shop	Breakfast Spot	Bike Trail	Check Cashing Service	Martial Arts Dojo	Grocery Store
1	Bath Beach	Pharmacy	Chinese Restaurant	Fast Food Restaurant	Bubble Tea Shop	Cantonese Restaurant	Gas Station	Italian Restaurant	Donut Shop	Pizza Place	German Restaurant
2	Bay Ridge	Italian Restaurant	Spa	Pizza Place	Bar	Bagel Shop	American Restaurant	Greek Restaurant	Sandwich Place	Hookah Bar	Pharmacy
3	Baychester	Donut Shop	Sandwich Place	Bus Station	Mattress Store	Men's Store	Mexican Restaurant	Baseball Field	Boat or Ferry	Bank	Supermarket
4	Bedford Park	Diner	Mexican Restaurant	Pizza Place	Deli / Bodega	Sandwich Place	Supermarket	Chinese Restaurant	Pharmacy	Coffee Shop	Park

*For further display of both all neighborhoods, please refer to the python library.
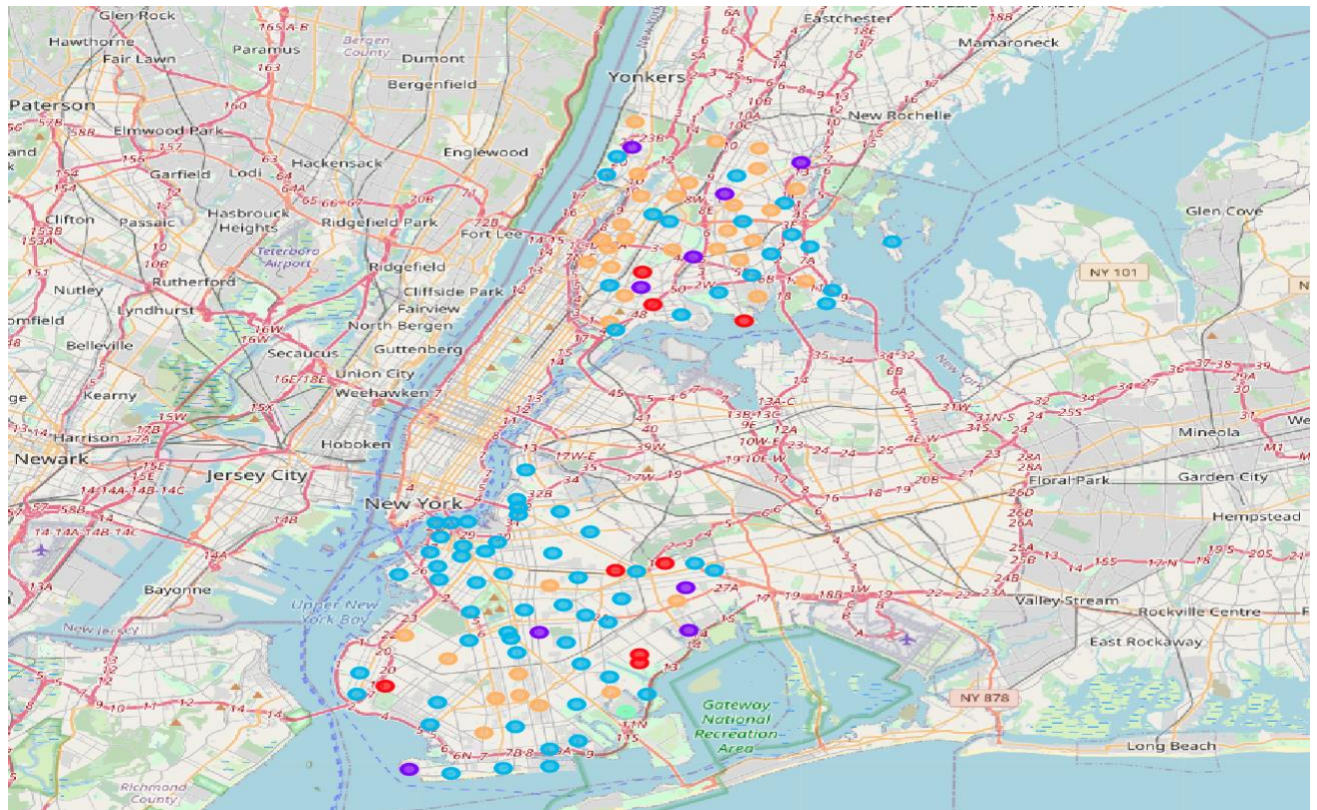
c. Clustering

The data set are prepared for clustering, and using the unsupervised machine learning based on K-clustering, we define that 6 cluster will be used for the segmentation on neighborhoods, as shown next:

	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
0	Bronx	Wakefield	40.894705	-73.847201	1	Pharmacy	Dessert Shop	Deli / Bodega	Ice Cream Shop	Gas Station
1	Bronx	Co-op City	40.874294	-73.829939	1	Fast Food Restaurant	Pharmacy	Accessories Store	Post Office	Discount Store
2	Bronx	Eastchester	40.887556	-73.827806	1	Caribbean Restaurant	Bus Station	Diner	Deli / Bodega	Pizza Place
3	Bronx	Fieldston	40.895437	-73.905643	2	Plaza	Music Venue	River	Cosmetics Shop	Bus Station
4	Bronx	Riverdale	40.890834	-73.912585	2	Bus Station	Park	Bank	Gym	Plaza
5	Bronx	Kingsbridge	40.881687	-73.902818	3	Pizza Place	Bar	Latin American Restaurant	Bakery	Supermarket
7	Bronx	Woodlawn	40.898273	-73.867315	3	Deli / Bodega	Pizza Place	Pub	Playground	Train Station
8	Bronx	Norwood	40.877224	-73.879391	0	Pizza Place	Bank	Park	Pharmacy	Chinese Restaurant
9	Bronx	Williamsbridge	40.881039	-73.857446	3	Nightclub	Soup Place	Caribbean Restaurant	Bar	Food Truck
10	Bronx	Baychester	40.866858	-73.835798	1	Donut Shop	Sandwich Place	Bus Station	Mattress Store	Men's Store

*For further display of both all neighborhoods and clusters, please refer to the python library.

The cluster are distributed as follows:

5. Results

Understanding the clusters

d. CLUSTER #1

This group of neighborhoods are defined by red circles in the map. This cluster englobes venues thar located, surrounded by different types of venues related to diners or Deli/Bodegas, meaning these neighborhoods are more likely located in residencial area preferring restaurants over fast-food, so this cluster is discarded.

	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
23	Longwood	Diner	Deli / Bodega	Park	Donut Shop	Grocery Store	Latin American Restaurant	Fast Food Restaurant	Sandwich Place	Flower Shop	Filipino Restaurant
27	Clason Point	Park	Recording Studio	South American Restaurant	Grocery Store	Convenience Store	Boat or Ferry	Pool	Bus Stop	Food Court	Food Stand
74	Canarsie	Thai Restaurant	Deli / Bodega	Caribbean Restaurant	Gym	Grocery Store	Asian Restaurant	Women's Store	Flower Shop	Fish & Chips Shop	Fish Market
81	Dyker Heights	Burger Joint	Dance Studio	Bagel Shop	Park	Grocery Store	Cosmetics Shop	Playground	Golf Course	Women's Store	Flea Market
89	Ocean Hill	Deli / Bodega	Bus Stop	Fried Chicken Joint	Food	Southern / Soul Food Restaurant	Donut Shop	Ice Cream Shop	Grocery Store	Chinese Restaurant	Seafood Restaurant
261	Paerdegat Basin	Food	Deli / Bodega	Harbor / Marina	Auto Garage	Gym	Asian Restaurant	Flower Shop	Filipino Restaurant	Fish & Chips Shop	Fish Market
267	Claremont Village	Grocery Store	Deli / Bodega	Bus Station	Bakery	Chinese Restaurant	Pizza Place	Discount Store	Park	Gym	Food
295	Highland Park	Deli / Bodega	Grocery Store	Garden	Food	Spanish Restaurant	Tennis Court	Big Box Store	Park	Pizza Place	Metro Station

e. **CUSTER #2**

This group of neighborhoods are defined by purple circles in the map. This cluster englobes venues thar located in highly cultural area with few related food venues, mainly being restaurants. As a result, this cluster is discarded.

	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
2	Eastchester	Caribbean Restaurant	Bus Station	Diner	Deli / Bodega	Bus Stop	Fast Food Restaurant	Chinese Restaurant	Seafood Restaurant	Food & Drink Shop	Automotive Shop
3	Fieldston	River	Music Venue	Bus Station	Plaza	Women's Store	Food & Drink Shop	Filipino Restaurant	Fish & Chips Shop	Fish Market	Flea Market
18	West Farms	Bus Station	Donut Shop	Chinese Restaurant	Park	Scenic Lookout	Lounge	Outdoors & Recreation	Metro Station	Basketball Court	Sandwich Place
25	Morrisania	Bus Station	Discount Store	Donut Shop	Pizza Place	Fast Food Restaurant	Grocery Store	Metro Station	Liquor Store	Sandwich Place	Seafood Restaurant
41	Olinville	Caribbean Restaurant	Food	Chinese Restaurant	Fried Chicken Joint	Deli / Bodega	Supermarket	Convenience Store	Metro Station	Pizza Place	Liquor Store
72	East New York	Deli / Bodega	Fried Chicken Joint	Plaza	Fast Food Restaurant	Chinese Restaurant	Caribbean Restaurant	Pharmacy	Event Service	Pizza Place	Convenience Store
73	Starrett City	Pharmacy	Intersection	American Restaurant	Shopping Mall	Donut Shop	Bus Station	Bus Stop	River	Caribbean Restaurant	Moving Target
85	Sea Gate	Sports Club	Beach	American Restaurant	Spa	Bus Station	Food Court	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop
300	Erasmus	Caribbean Restaurant	Yoga Studio	Pizza Place	Furniture / Home Store	Music Venue	Mobile Phone Shop	Chinese Restaurant	Grocery Store	Gym / Fitness Center	Pharmacy

f. CLUSTER #3

This group of neighborhoods are defined by turquoise circles in the map. This cluster englobes venues thar located in highly commercial area for all types of venues, including food related, this would be a good option for opening a pizzeria place, but the neighborhood for our entrepreneurs is not grouped within this cluster, this could be a 2nd option.

	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
4	Riverdale	Park	Plaza	Home Service	Baseball Field	Bank	Gym	Medical Supply Store	Bus Station	Food Truck	Food Stand
9	Williamsbridge	Bar	Nightclub	Caribbean Restaurant	Soup Place	Convenience Store	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Frozen Yogurt Shop
10	Baychester	Donut Shop	Bank	Fried Chicken Joint	Convenience Store	Mexican Restaurant	Mattress Store	Fast Food Restaurant	Pet Store	Sandwich Place	Supermarket
11	Pelham Parkway	Italian Restaurant	Pizza Place	Chinese Restaurant	Deli / Bodega	Metro Station	Coffee Shop	Bank	Eye Doctor	Performing Arts Venue	Mexican Restaurant
12	City Island	Thrift / Vintage Store	Seafood Restaurant	Jewelry Store	Diner	Baseball Field	Boat or Ferry	Spanish Restaurant	Smoke Shop	Deli / Bodega	Pizza Place
16	Fordham	Mobile Phone Shop	Spanish Restaurant	Clothing Store	Shoe Store	Bank	Donut Shop	Pharmacy	Fast Food Restaurant	Pizza Place	Supplement Shop
22	Port Morris	Furniture / Home Store	Brewery	Storage Facility	Food Truck	Latin American Restaurant	Donut Shop	Peruvian Restaurant	Distillery	Spanish Restaurant	Restaurant
24	Hunts Point	Waste Facility	Spanish Restaurant	Bank	Food	Pizza Place	BBQ Joint	Café	Grocery Store	Bakery	Restaurant
26	Soundview	Chinese Restaurant	Grocery Store	Discount Store	Pharmacy	Liquor Store	Home Service	Cosmetics Shop	Playground	Latin American Restaurant	Video Store
28	Throgs Neck	Italian Restaurant	Asian Restaurant	Deli / Bodega	Pizza Place	Coffee Shop	Baseball Field	Bar	Sports Bar	American Restaurant	Juice Bar
29	Country Club	Sandwich Place	Italian Restaurant	Trail	Athletics & Sports	Playground	Flea Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop
31	Westchester Square	Fast Food Restaurant	Pizza Place	Sandwich Place	Donut Shop	Pharmacy	Baseball Field	Mexican Restaurant	Mobile Phone Shop	Metro Station	Park
34	Belmont	Italian Restaurant	Pizza Place	Deli / Bodega	Bakery	Grocery Store	Dessert Shop	Donut Shop	Spanish Restaurant	Bank	Sandwich Place
35	Spuyten Duyvil	Intersection	Grocery Store	Tennis Stadium	Bank	Pharmacy	Tennis Court	Scenic Lookout	Park	Thai Restaurant	Bus Line
37	Pelham Bay	Fast Food Restaurant	Italian Restaurant	Bank	Sandwich Place	Donut Shop	Diner	Convenience Store	Dive Bar	Cosmetics Shop	Thai Restaurant
39	Edgewater Park	Italian Restaurant	Pizza Place	Coffee Shop	Asian Restaurant	Donut Shop	Deli / Bodega	Chinese Restaurant	Park	Bookstore	Bar
44	Unionport	Diner	Ice Cream Shop	Donut Shop	Latin American Restaurant	Plaza	Discount Store	Caribbean Restaurant	Fast Food Restaurant	Mobile Phone Shop	Seafood Restaurant

g. CLUSTER #4

	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
76	Mill Island	Locksmith	Pool	Women's Store	Fish Market	Farm	Farmers Market	Fast Food Restaurant	Field	Filipino Restaurant	Fish & Chips Shop

This neighborhood is defined by light green circle in the map. Since this neighborhood is located at the harbor, the food related venues are limited to seafood, this probably would not be a good environment, not to mention it is not a commercial area for restaurants. This cluster is discarded.

h. CLUSTER #5

	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	Wakefield	Pharmacy	Dessert Shop	Gas Station	Ice Cream Shop	Donut Shop	Laundromat	Sandwich Place	Flower Shop	Fast Food Restaurant	Field
1	Co-op City	Bus Station	Baseball Field	Discount Store	Pizza Place	Grocery Store	Park	Fast Food Restaurant	Bagel Shop	Trail	Pharmacy
5	Kingsbridge	Pizza Place	Bar	Bakery	Mexican Restaurant	Latin American Restaurant	Supermarket	Fried Chicken Joint	Spanish Restaurant	Pharmacy	Liquor Store
7	Woodlawn	Deli / Bodega	Pizza Place	Pub	Food & Drink Shop	Playground	Bar	Cosmetics Shop	Pharmacy	Food Truck	Bakery
8	Norwood	Pizza Place	Park	Bank	American Restaurant	Pharmacy	Coffee Shop	Spanish Restaurant	Bus Station	Caribbean Restaurant	Chinese Restaurant
13	Bedford Park	Chinese Restaurant	Diner	Pizza Place	Mexican Restaurant	Deli / Bodega	Sandwich Place	Train Station	Spanish Restaurant	Pub	Smoke Shop
14	University Heights	Fried Chicken Joint	Pizza Place	Bakery	Bus Station	Supermarket	Shoe Store	Grocery Store	Chinese Restaurant	Fast Food Restaurant	Sandwich Place
15	Morris Heights	Recreation Center	Pizza Place	Spanish Restaurant	Bus Station	Moving Target	Grocery Store	Bank	Pharmacy	Deli / Bodega	Dry Cleaner
17	East Tremont	Pizza Place	Discount Store	Lounge	Breakfast Spot	Supermarket	Smoke Shop	Café	Mobile Phone Shop	Fish & Chips Shop	Shoe Store
19	High Bridge	Pharmacy	Chinese Restaurant	Bus Station	Supermarket	Pizza Place	Deli / Bodega	Latin American Restaurant	Seafood Restaurant	Spanish Restaurant	Park
20	Melrose	Pizza Place	Pharmacy	Discount Store	Gym / Fitness Center	Sandwich Place	Grocery Store	Supermarket	Gym	Mexican Restaurant	Supplement Shop
21	Mott Haven	Spanish Restaurant	Donut Shop	Gym	Pizza Place	Fish & Chips Shop	Storage Facility	Mobile Phone Shop	Bookstore	Peruvian Restaurant	Latin American Restaurant
30	Parkchester	Supermarket	Pizza Place	Chinese Restaurant	American Restaurant	Kids Store	Restaurant	Mattress Store	Gym	Sandwich Place	Mobile Phone Shop
92	Midwood	Pizza Place	Convenience Store	Ice Cream Shop	Field	Bakery	Candy Store	Video Game Store	Pharmacy	Food Court	Food & Drink Shop
98	Ocean Parkway	Sushi Restaurant	Pizza Place	Liquor Store	Dessert Shop	Sake Bar	Bus Station	Food & Drink Shop	Food	Steakhouse	Supermarket
260	New Lots	Pharmacy	Fried Chicken Joint	Grocery Store	Discount Store	Art Gallery	Park	Bank	Pizza Place	Plaza	Asian Restaurant
262	Mill Basin	Japanese Restaurant	Chinese Restaurant	Pizza Place	Bank	Italian Restaurant	Bagel Shop	Pet Store	Sandwich Place	Liquor Store	Pharmacy
269	Mount Eden	Pharmacy	Pizza Place	Grocery Store	Fast Food Restaurant	Dog Run	Park	Clothing Store	Bus Station	Sandwich Place	Gym
270	Mount Hope	Grocery Store	Intersection	Donut Shop	Supermarket	Metro Station	Spanish Restaurant	Sandwich Place	Pizza Place	Video Game Store	Pharmacy
296	Madison	Bagel Shop	Deli / Bodega	Italian Restaurant	Restaurant	Pizza Place	Pilates Studio	Candy Store	Spa	Dessert Shop	Hotel
297	Bronxdale	Italian Restaurant	Pizza Place	Breakfast Spot	Supermarket	Mexican Restaurant	Gym	Paper / Office Supplies Store	Performing Arts Venue	Spanish Restaurant	Bank
298	Allerton	Pizza Place	Deli / Bodega	Supermarket	Grocery Store	Gas Station	Fried Chicken Joint	Fast Food Restaurant	Pharmacy	Martial Arts Dojo	Bakery
299	Kingsbridge Heights	Pizza Place	Spanish Restaurant	Bus Station	Grocery Store	Chinese Restaurant	Coffee Shop	Intersection	Pharmacy	Bus Stop	Café

This neighborhood is defined by orange circle in the map. This cluster contains Bronxdale, we can evidence that's grouped within neighborhoods with highly commercial areas, with a lot of food related venues, and there a lot of neighborhoods that are from both borough (Bronx and Brooklyn), this cluster would be perfect to select a good neighborhood taking into consideration the les competition for pizzerias, like New Lotus, Brooklyn.

6. Discussion

Based on the result in our cluster we can advise to our entrepreneur to consider selecting a neighborhood from the cluster number 5, which was the areas mostly related to his neighborhood, and that both Brooklyn's and Bronx's neighborhoods are included in the cluster selected.

7. Conclusion

There similarities between Bronx and Brooklyn neighborhoods, and our client can rest assured that common characteristics within these neighborhoods make a perfect place to set for a new pizzeria in a place like New Lotus, which is in the Borough he wanted and can think into creating fast food chain along various neighborhoods located in both Boroughs, thinking in the long term.