Capstone Project - The Battle of the Neighborhoods (week 2)

use case: The Bronx (in New York) segmentation and clustering
to find a location for a "fruit and vegetable store"

Table of content

I/ Description of the problem

M Rob Zuntfy the General Manager of a "fruits and vegetable store" approached me as a data scientist and asked me to solve his problem.

He wants to launch a store in the Bronx, in New York City but doesn't know very well how to proceed especially with the location of his future store.

But he has some constraints and preferences: First he surely wants to have a store in Bedford Park, a neighborhood of the Bronx, because he loves this place and he has already notice that there is some stores with the same activity there. So that he thinks that, having a fruit and vegetable business there will give him business opportunities, as some stores already exist there. And according to him, this fact shows that this location is a very interesting market place.

Second, in another hand, he has some doubt because he is afraid of the competition he will face with the existing stores already established in this place.

The last point raised by M Rob Zuntfy is that, he has a preference for a location which is near a bus station because for him, his target clients are people who usually uses bus to move. And with this last option, he'd like to launch his store in another place, but he doesn't know where!

II/ Description of the data and how it will be used to solve the problem

In [ ]:

I/ Data source description

To solve the problem described in the introduction section, I used the dataset proposed in this coursera Data Science capstone course as shown here: newyork_data.json' https://cocl.us/new_york_dataset.

II/ How the dataset had been used

Using this dataset, I have my focus on the Bronx and its neighborhoods.
So to successfully complete this task I step by step imported the dataset, cleaned it, extracted the useful dataset of it, used the geospatial coordinate of the Bronx and my Foursquare credential to look at the different stores around the Bronx.

III/ Methodology (working with the dataset): overview

After downloading, loading, exploring and looking at in the dataset, I transformed the data into a dataframe with the usefull columns I needed (['Borough', 'Neighborhood', 'Latitude', 'Longitude']).(see section a here)

Then I got the geographical coordinate of the Bronx. I created The Bronx's map to have a idea of this map, by using the folium.Map method and put in it, the Bronx geographical coordinate.

I've extracted the Bronx city of the previous New York City dataframe. (see section b below)

I continued by using my foursquare client and secret ID to use the foursquare method to look at around the first Bronx neighborhood.

And with this first neighborhood geographical coordinate, I explored the 50 top venues in a 300 meter radius.(section c)

I managed different dataframes by creating unique category in Bronx venues and group them.

And finally, I returned the most common venues and finished by creating the clusters I needed.(section e)

Now let's look at this analysis showing the flow of these tasks using the dataset set.

IV/ Analysis: the flow of the described methodology using the dataset
Section a): Importing libraries, Downloading, loading, exploring and looking at the dataset

In this section, I start with the preliminaries actions to have the dataset and be ready to work with it.
Please see in my github book to view the different codes I used to complete this section)

Section b): plotting The Bronx Map
In this section, I used The Bronx geographical coordinate and plot the needed map
On this map, you can see in blue points, the neighborhoods of The Bronx

Section c) Have a look of the 50 top venues in 300m of radius

Using Foursquare method and some python codes , the output table below summarize the request.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
	Neighborhood						
1	**Allerton**	64	64	64	64	64	64
2	**Baychester**	40	40	40	40	40	40
3	**Bedford Park**	16	16	16	16	16	16
4	**Belmont**	148	148	148	148	148	148
5	**Castle Hill**	16	16	16	16	16	16
6	**City Island**	72	72	72	72	72	72
7	**Claremont Village**	20	20	20	20	20	20
8	**Clason Point**	20	20	20	20	20	20
9	**Co-op City**	44	44	44	44	44	44
10	**Concourse**	44	44	44	44	44	44
11	**Concourse Village**	52	52	52	52	52	52
12	**Country Club**	8	8	8	8	8	8
13	**East Tremont**	48	48	48	48	48	48
14	**Eastchester**	36	36	36	36	36	36
15	**Edenwald**	24	24	24	24	24	24
16	**Edgewater Park**	32	32	32	32	32	32
17	**Fieldston**	12	12	12	12	12	12
18	**Fordham**	180	180	180	180	180	180
19	**High Bridge**	48	48	48	48	48	48
20	**Hunts Point**	28	28	28	28	28	28
21	**Kingsbridge**	84	84	84	84	84	84
22	**Kingsbridge Heights**	64	64	64	64	64	64
23	**Longwood**	20	20	20	20	20	20
24	**Melrose**	20	20	20	20	20	20
25	**Morris Heights**	16	16	16	16	16	16
26	**Morris Park**	8	8	8	8	8	8

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
27	Morrisania	28	28	28	28	28	28
28	Mott Haven	40	40	40	40	40	40
29	Mount Eden	32	32	32	32	32	32
30	Mount Hope	16	16	16	16	16	16
31	North Riverdale	40	40	40	40	40	40
32	Norwood	44	44	44	44	44	44
33	Olinville	32	32	32	32	32	32
34	Parkchester	40	40	40	40	40	40
35	Pelham Bay	44	44	44	44	44	44
36	Pelham Gardens	64	64	64	64	64	64
37	Pelham Parkway	52	52	52	52	52	52
38	Port Morris	32	32	32	32	32	32
39	Riverdale	16	16	16	16	16	16
40	Schuylerville	16	16	16	16	16	16
41	Soundview	28	28	28	28	28	28
42	Spuyten Duyvil	20	20	20	20	20	20
43	Unionport	32	32	32	32	32	32
44	University Heights	48	48	48	48	48	48
45	Van Nest	64	64	64	64	64	64
46	Wakefield	20	20	20	20	20	20
47	West Farms	44	44	44	44	44	44
48	Westchester Square	64	64	64	64	64	64
49	Woodlawn	48	48	48	48	48	48

Section e) the most common venues

Now I am going to show the most common venues to try to know where the suitable and best places are, regarding the type of activity we can find there. So these categorizations will give opportunity to advise M Rob Z in choosing the right place/best location to establish his fruit and vegetable store

N°	Neighbor hood	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	Allerton	Pizza Place	Chinese Restaurant	Deli / Bodega	Electronics Store	Breakfast Spot	Fast Food Restaurant	Discount Store	Department Store	Martial Arts Dojo	Donut Shop
2	Baychester	Discount Store	Mattress Store	Bus Station	Donut Shop	Baseball Field	Spanish Restaurant	Bank	Sandwich Place	Pizza Place	American Restaurant
3	Bedford Park	Deli / Bodega	Fruit & Vegetable Store	Chinese Restaurant	Donut Shop	Wings Joint	Fast Food Restaurant	Cupcake Shop	Dance Studio	Department Store	Dessert Shop
4	Belmont	Italian Restaurant	Deli / Bodega	Pizza Place	Dessert Shop	Liquor Store	Food & Drink Shop	Grocery Store	Bar	Piano Bar	Diner
5	Castle Hill	Pharmacy	Diner	Spanish Restaurant	Market	Wings Joint	Event Space	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega

IV/ Results: different locations groupped by activity in 5 big areas (or 5 clusters)

2nd area (2nd cluster): This cluster shows "Fieldson" neighborhood and its most common venues

	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
3	Fieldston	Plaza	Bus Station	River	Wings Joint	Construction & Landscaping	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store
55	Fieldston	Plaza	Bus Station	River	Wings Joint	Construction & Landscaping	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store
107	Fieldston	Plaza	Bus Station	River	Wings Joint	Construction & Landscaping	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store
159	Fieldston	Plaza	Bus Station	River	Wings Joint	Construction & Landscaping	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store

3rd area (3rd cluster): This cluster shows "Fruit and Vegetable stores" and other most common places of activity in each neighborhood

	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
12	Bedford Park	Deli / Bodega	Fruit & Vegetable Store	Chinese Restaurant	Donut Shop	Wings Joint	Fast Food Restaurant	Cupcake Shop	Dance Studio	Department Store	Dessert Shop
13	University Heights	Deli / Bodega	Wings Joint	Food Truck	Chinese Restaurant	Donut Shop	Seafood Restaurant	Bank	Bakery	Fish Market	Supermarket
32	Morris Park	Deli / Bodega	Donut Shop	Wings Joint	Gym	Cosmetics Shop	Cupcake Shop	Dance Studio	Department Store	Dessert Shop	Diner
38	Edgewater Park	Deli / Bodega	Park	Bus Station	Cupcake Shop	Wings Joint	Fast Food Restaurant	Dance Studio	Department Store	Dessert Shop	Diner
64	Bedford Park	Deli / Bodega	Fruit & Vegetable Store	Chinese Restaurant	Donut Shop	Wings Joint	Fast Food Restaurant	Cupcake Shop	Dance Studio	Department Store	Dessert Shop
65	University Heights	Deli / Bodega	Wings Joint	Food Truck	Chinese Restaurant	Donut Shop	Seafood Restaurant	Bank	Bakery	Fish Market	Supermarket
84	Morris Park	Deli / Bodega	Donut Shop	Wings Joint	Gym	Cosmetics Shop	Cupcake Shop	Dance Studio	Department Store	Dessert Shop	Diner
90	Edgewater Park	Deli / Bodega	Park	Bus Station	Cupcake Shop	Wings Joint	Fast Food Restaurant	Dance Studio	Department Store	Dessert Shop	Diner
116	Bedford Park	Deli / Bodega	Fruit & Vegetable Store	Chinese Restaurant	Donut Shop	Wings Joint	Fast Food Restaurant	Cupcake Shop	Dance Studio	Department Store	Dessert Shop

4th area (4th cluster): This cluster shows "Country Club" neighborhood and its most common venues

	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
28	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store	Dessert Shop
80	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store	Dessert Shop
132	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store	Dessert Shop
184	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint	Fast Food Restaurant	Cosmetics Shop	Cupcake Shop	Dance Studio	Deli / Bodega	Department Store	Dessert Shop

For more details on all the five clusters, please see here.

V/ Results and comments

The different clusters put forward these main points I can use to advise M Rob Z:

1. As M Rob first preference is to locate his Fruit and Vegetable store in Bedford Park, this option looks fine, because the 3rd cluster tells us that Bedford Park is effectively among the most 2nd common places where we find fruit and vegetable stores.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
12	Bedford Park	Deli / Bodega	Fruit & Vegetable Store	Chinese Restaurant

2. Respecting the constraints of a Bus station, I will advise M Rob Z to locate his future fruit and store in Fieldson neighborhood, as it shows by the 2nd cluster for example. And there, he will also have the chance to create the first and only one fruit and vegetable store

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
3	Fieldston	Plaza	Bus Station	River
55	Fieldston	Plaza	Bus Station	River
107	Fieldston	Plaza	Bus Station	River
159	Fieldston	Plaza	Bus Station	River

3. Another good option for M Rob Z is to create his fruit and vegetable store in Country Club neighborhood which is a place where we find the most common Vegetarian/Vietnam Restaurant; as shown by the 4th cluster. These restaurants can also be M Rob Z's store clients!

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
28	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint
80	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint
132	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint
184	Country Club	Vegetarian / Vegan Restaurant	Sandwich Place	Wings Joint