EVALUATION OF THE BEST NEIGHBORHOODS TO OPEN A NEW HEALTHY FOOD STORE IN THE CITY OF TORONTO, CANADA

May, 2020

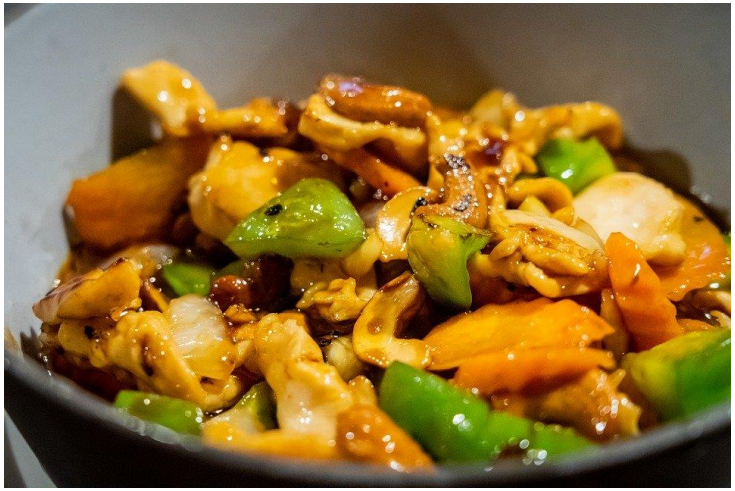
Introduction

The global health food market has been increasing the last years and it is expected to grow up to USD 250 billion for 2020

In particular the healthy food restaurants are of interest of neighbors, who want to satisfy their demand for naturally health food, functional food, organic food and food intolerance products (gluten/lactose free products, etc.).

This project will be developed to distribute information about healthy food restaurants in Toronto to investors planning to open a healthy food store in this city.

Objectives



Evaluating the neighborhood venues demands
Exploring and classifying the venues of a neighborhood through geospatial data is an efficient way to evaluate all the services that such community has available.

Additional considerations to open a new Healthy Food Store
The study will also include the number of healthy food-related restaurants /venues in order to get insights about the lifestyle of the neighbors.

Approach and methology

Busines Understanding

- Analysis of the problem to be solved. In this case, to get information about the best places to open a new Healthy Food Store

Data Understanding

- The data needed for performing the study is: demographic and geographic data of the boroughs and neighborhoods in Toronto. The postal code and the respective geospatial coordinates for each neighborhood. List of Venues for each location.

Data Preparation

- Preparare pandas dataframe with the neighborhoods, boroughs and postal code.
- Prepare pandas dataframe with the igeospatial coordinates for each neighborhood.
- Prepare the list of venues using the FoursquareAPI.

Modeling

- Create a map of Toronto's neighborhoods.
- Create a **machine learning algorithm based on a Kmean** approach to segment and cluster the list of venues in each neighborhood.

Evaluation

- Create a map of the clusters and evaluate **the accuracy of the kmean algorithm** during the clustering process.

Deployment

- According to the frequency of the Healthy Food Stores in each neighborhood, make a conclusion about which one would be the best to open a new store.

Data

- **Geographic data about Toronto's neighborhoods:** https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- **Geospatial Coordinates information:** Geocoder platform https://geocoder.readthedocs.io/index.html
- **List of Venues available at the FoursquareAPI:** (https://developer.foursquare.com/docs/api/venues/explore)
- **Other tools and library used:** Pandas library, Numpy library, Foursquare API, Folium library, Geopy library

Results

- **Creating Dataframes in pandas**

Postal Code

Geospatial Coordinates

Boroughs and Neighborhoods

Dataframe 4 of Toronto with ZIP code and geospatial coordinates (toronto_code_df)

```
[5]: toronto_code_df = toronto_code_df.drop('Postal Code2', axis =1) #Drop the duplicate columns
     toronto_code_df
```

[5]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarbor	Woburn	43.770992	-79.216917
4	M1H	Scarbor	Cedarbrae	43.773136	-79.239476
5	M1J	Scarbor	gh Village	43.744734	-79.239476
6	M1K	Scarbor	ount Park	43.727929	-79.262029
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848

Results

- **Create a map of Toronto's neighborhoods**

Map of Toronto neighborhoods

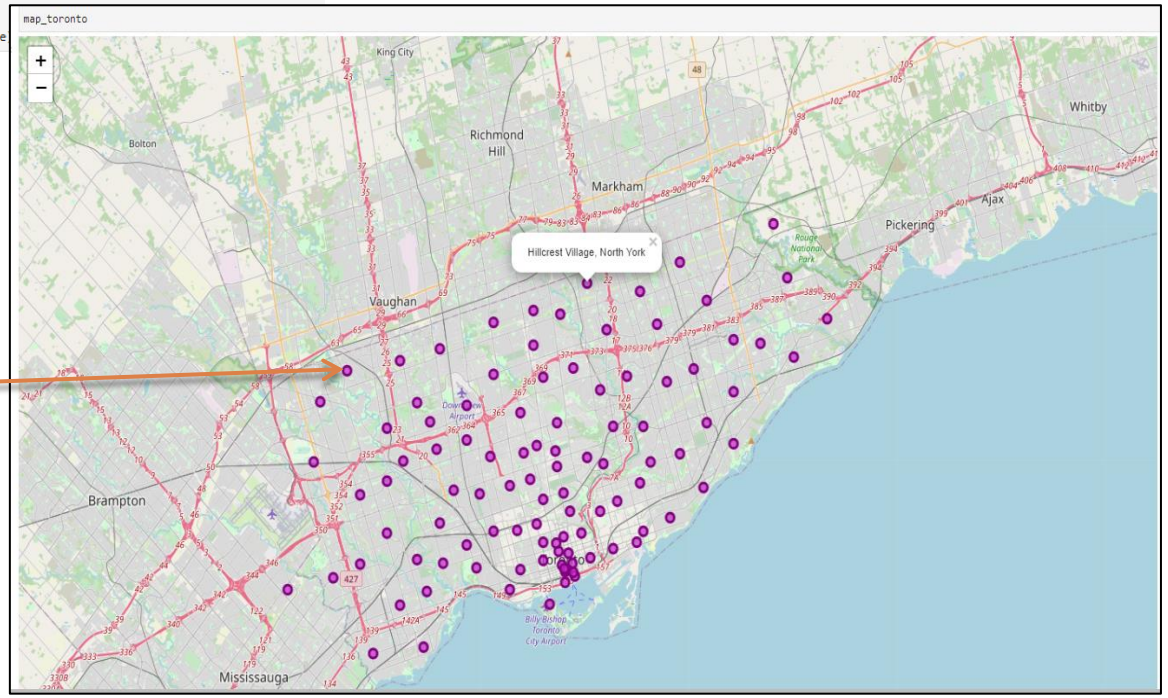
In order to define an instance of the geocoder, we need to define a user_agent. We will name our agent *toronto_explorer*, as shown below.

```
6]: address = 'Toronto, ON' #Use geopy library to get the latitude and longitude values of Toronto

geolocator = Nominatim(user_agent="toronto_explorer")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinate of Toronto are {}, {}.'.format(latitude, longitude
```

```
The geograpical coordinate of Toronto are 43.6534817, -79.3839347.
```

Purple markers indicate each neighborhood in Toronto



Results

- **Using the Explore of Foursquare query to create a list of venues**

Create a dataframe **toronto_venues**

```
[21]: # Get the venues for each neighborhood
toronto_venues = getNearbyVenues(
                                  postalCodes=toronto_code_df['Postal Code'],
                                  boroughs=toronto_code_df['Borough'],
                                  neighbourhoods=toronto_code_df['Neighborhood'],
                                  latitudes=toronto_code_df['Latitude'],
                                  longitudes=toronto_code_df['Longitude']
                                 )

toronto_venues.head()
```

[21]:

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	Affordable Toronto Movers	43.787919	-79.162977	Moving Target
3	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
4	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store

Checking the size of the dataframe

```
[22]: toronto_venues.shape
```

```
[22]: (2142, 9)
```

Venues by category

Ordering neighborhoods by number of venues

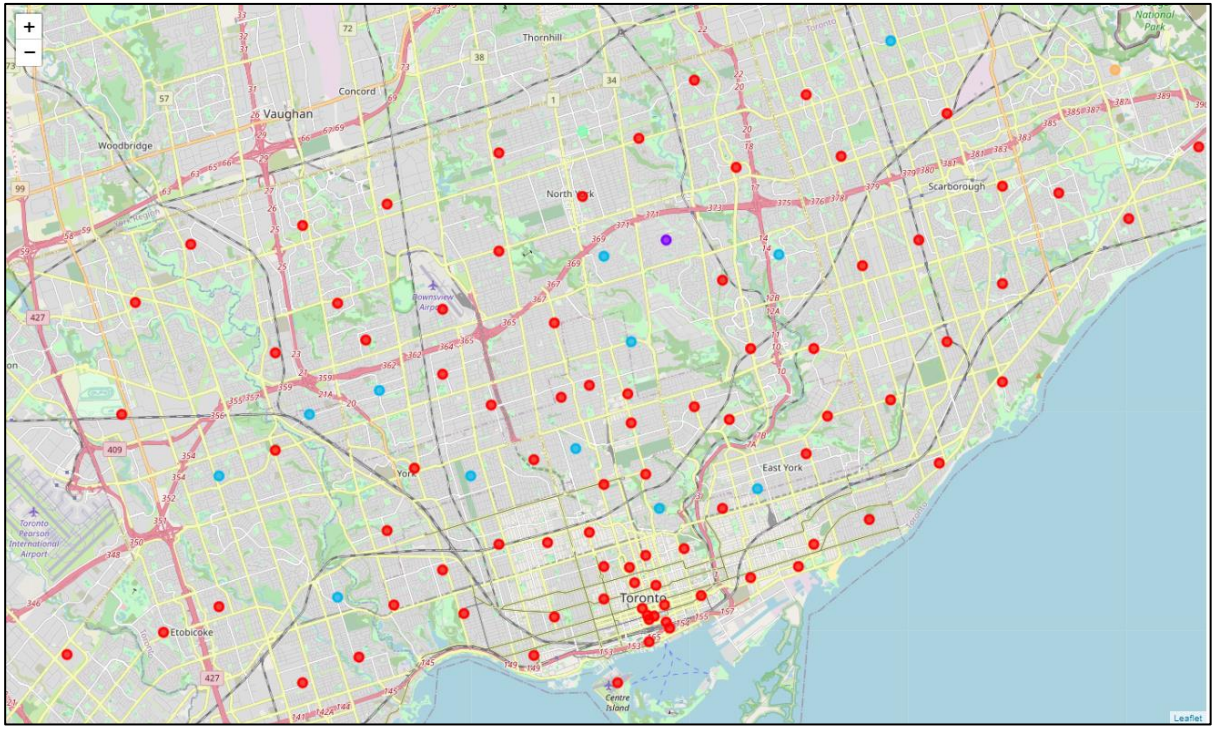
Checking how many venues were returned for each neighborhood

```
toronto_venues.groupby('Neighborhood').count().sort_values(by=['Venue'], ascending=False)
```

Neighborhood	Postal Code	Borough	Latitude	Longitude	Venue
Garden District, Ryerson	100	100	100	100	100
First Canadian Place, Underground city	100	100	100	100	100
Commerce Court, Victoria Hotel	100	100	100	100	100
Toronto Dominion Centre, Design Exchange	100	100	100	100	100
Harbourfront East, Union Station, Toronto Islands	100	100	100	100	100
Richmond, Adelaide, King	92	92	92	92	92

Results

- **Machine learning approach to cluster the neighborhoods**



Map of the neighborhoods of Toronto clustered through the K-mean algorithm.

Discussion - Analysis of Clusters

- **Cluster 1 (cluster label = 0) (Red color)**

The **Cluster 1** is formed by neighborhood with a high density of restaurants of a diverse range of options. There are 100 neighborhoods in this cluster

```
toronto_merged3.loc[toronto_merged3['Cluster Labels'] == 0]
```

	Postal Code	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
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0	Bar	Moving Target	Women's Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Deli / Bodega
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0	Rental Car Location	Mexican Restaurant	Medical Center	Bank	Intersection	Moving Target	Breakfast Spot	Electronics Store	Doner Restaurant	Dog Run
3	M1G	Scarborough	Woburn	43.770992	-79.216917	0	Coffee Shop	Indian Restaurant	Korean Restaurant	Women's Store	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	0	Fried Chicken Joint	Gas Station	Bakery	Hakka Restaurant	Bank	Athletics & Sports	Caribbean Restaurant	Thai Restaurant	Dessert Shop	Dim Sum Restaurant
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476	0	Grocery Store	Playground	Jewelry Store	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park	43.727929	-79.262029	0	Department Store	Bus Station	Coffee Shop	Discount Store	Convenience Store	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	Distribution Center
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge	43.711112	-79.284577	0	Bakery	Park	Metro Station	Bus Station	Intersection	Bus Line	Ice Cream Shop	Soccer Field	Dog Run	Distribution Center
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West	43.716316	-79.239476	0	American Restaurant	Motel	Movie Theater	Intersection	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848	0	College Stadium	Skating Rink	Café	General Entertainment	Women's Store	Distribution Center	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...	43.757410	-79.273304	0	Indian Restaurant	Chinese Restaurant	Pet Store	Vietnamese Restaurant	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
11	M1R	Scarborough	Wexford, Maryvale	43.750072	-79.295849	0	Sandwich Place	Vietnamese Restaurant	Breakfast Spot	Bakery	Auto Garage	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store
12	M1S	Scarborough	Agincourt	43.794200	-79.262029	0	Latin American Restaurant	Skating Rink	Lounge	Breakfast Spot	Women's Store	Dog Run	Dim Sum Restaurant	Diner	Discount Store	Distribution Center
13	M1T	Scarborough	Clarks Corners, Tam O'Shanter, Sullivan	43.781638	-79.304302	0	Pizza Place	Shopping Mall	Bank	Fried Chicken Joint	Italian Restaurant	Thai Restaurant	Chinese Restaurant	Gas Station	Fast Food Restaurant	Pharmacy
15	M1W	Scarborough	Steeles West, L'Amoreaux West	43.799525	-79.318389	0	Grocery Store	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Noodle House	Sandwich Place	Pharmacy	Pizza Place	Bank	Breakfast Spot
17	M2H	North York	Hillcrest Village	43.803762	-79.363452	0	Pool	Golf Course	Fast Food Restaurant	Dog Run	Mediterranean Restaurant	Women's Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner

Discussion

- **Cluster 2: (cluster label = 1) (Purple color)**

The **Cluster 2** is formed for 1 neighborhood and interesting, although some restaurants are present, it is mostly formed for stores as department and women's store.

```
toronto_merged3.loc[toronto_merged3['Cluster Labels'] == 1]
```

	Postal Code	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
20	M2L	North York	York Mills, Silver Hills	43.75749	-79.374714	1	Cafeteria	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Women's Store	Deli / Bodega

Discussion

- **Cluster 3: (cluster label = 2) (Blue color)**

The **Cluster 3** is formed by neighborhoods that are near to parks and other leisure areas aas playgrounds and swimming pool. There are 12 neighborhoods in this cluster, and also can be found some restaurants, although in less proportion compared to cluster 1.

	Postal Code	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
14	M1V	Scarborough	Milliken, Agincourt North, Steeles East, L'Amo...	43.815252	-79.284577	2	Park	Playground	Distribution Center	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Dog Run
23	M2P	North York	York Mills West	43.752758	-79.400049	2	Park	Bank	Convenience Store	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Doner Restaurant
25	M3A	North York	Parkwoods	43.753259	-79.329656	2	Park	Food & Drink Shop	Discount Store	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Distribution Center	Ethiopian Restaurant
40	M4J	East York	East Toronto	43.685347	-79.338106	2	Park	Coffee Shop	Intersection	Convenience Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Donut Shop	Doner Restaurant	Deli / Bodega
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	2	Park	Bus Line	Swim School	Dumpling Restaurant	Drugstore	Donut Shop	Doner Restaurant	Dog Run	Dance Studio	Discount Store
50	M4W	Downtown Toronto	Rosedale	43.679563	-79.377529	2	Park	Playground	Trail	Drugstore	Donut Shop	Doner Restaurant	Dog Run	Distribution Center	Dumpling Restaurant	Curling Ice
64	M5P	Central Toronto	Forest Hill North & West	43.696948	-79.411307	2	Park	Jewelry Store	Bus Line	Trail	Sushi Restaurant	Drugstore	Dumpling Restaurant	Donut Shop	Doner Restaurant	Dog Run
74	M6E	York	Caledonia-Fairbanks	43.689026	-79.453512	2	Park	Pool	Women's Store	College Stadium	Colombian Restaurant	Electronics Store	Eastern European Restaurant	Dumpling Restaurant	Drugstore	Donut Shop
79	M6L	North York	North Park, Maple Leaf Park, Upwood Park	43.713756	-79.490074	2	Park	Bakery	Construction & Landscaping	Dog Run	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Doner Restaurant
90	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	2	Park	River	Pool	Curling Ice	Dance Studio	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
98	M9N	York	Weston	43.706876	-79.518188	2	Park	Convenience Store	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Dog Run	Dance Studio
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724	2	Park	Bus Line	Pizza Place	Sandwich Place	Diner	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Discount Store

Discussion

- **Cluster 4: (cluster label = 3) (Light green color)**

The **Cluster 3** is also formed by 1 neighborhood and it is characterized by having some places to eat and stores at the same proportion.

	Postal Code	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
21	M2M	North York	Willowdale, Newtonbrook	43.789053	-79.408493	3	Piano Bar	Women's Store	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Doner Restaurant

Discussion

- **Cluster 5: (cluster label = 4) (Orange color)**

The **Cluster 5** is also formed by one neighborhood in which there are few restaurants and also few options of leisure.

	Postal Code	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	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	4	Fast Food Restaurant	Women's Store	Dog Run	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Doner Restaurant

Analyzing the best neighborhoods

- The recommendation for the best neighborhood to open a new Healthy food restaurant is based on the neighborhoods with less/lack of option of heathy food.
- Considering it, I will recommend the neighborhoods **Willowdale, Newtonbrook** (cluster 4) and **York Mills, Silver Hills** (cluster 2) because in both neighborhoods the algorithm found that there are few restaurants.
- As second and strong option, I will add the neighborhoods **Golden Mile, Clairlea, Oakridge, Don Mills, Downsview, Leaside Davisville North** from cluster 1, because even when it is a cluster with a lot of competitors, there is an advantage because it is also present Yoga Studio, Gym, Fitness center and other that are mostly visited for people very interested in Healthy food, which are our target.

Conclusion

- After preparing, modelling and analyzing the data, it can be concluded the algorithm **k-mean** of the machine learning approach is very useful to find segmentation in real life application that involves studying the market, competitors, target, etc.
- We found there are 2 great potential neighborhoods to open a healthy food restaurant and three more neighborhoods in which, even with a lot of competitors, the target customers can be found.