Coursera_Capstone Project: Battle of the Neighborhoods

Project Title: Evaluation of the best neighborhoods to open a new Healthy Food Restaurant in the city of Toronto, Canada

1. Introduction

The global health food market has been increasing the last years and it is expected to grow up to USD 250 billion for 2022. This increment is in part due to the more healthy habits adopted for people around the globe. A new lifestyle based on healthy habits as the consumption of health-food and the practice of physical exercises is now practicing for people from all ages that try to avoid diseases such as coronary heart disease, stroke, hypertension, osteoporosis, and type-2 diabetes. Mainly the change in nutrition is seemed as one opportunity to open new restaurants, online shop and physical restaurants to satisfy the demand customers with the necessities of such lifestyle.

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.). Many investors are opening new healthy food restaurants due to the growing market and demand for such places.

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. With this data, the investors can perform an analysis of the potential competitors, so they could open the store in a neighborhood with a low-frequency of such restaurants.

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. Moreover, when this information is clustered and segmented and the analysis is performed through a whole city, it can turn into a powerful tool for investors, business people and authorities interested in started new ventures. Hence, our proposal here is clustering and segmenting the neighbourhoods of Toronto according to a detailed study of the number and location of the current Healthy Food Restaurants in the neighborhoods of Toronto.

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. For example, Yoga/Pilate studios are visited for our target public, so this information is relevant in order to get information about the lifestyle of the public. This information can be also useful for marketing purpose.

2. Data

Geographic data about Toronto's neighborhoods

To perform our study, we will use the information about the boroughs and their respective neighborhoods of the city of Toronto (Ontario, Canada) available at the wikipedia Postal Code page:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M (**1**)

This list contains the information of: Postal Code, Borough and Neighborhood (see example, Figure 1).

Postal Code	Borough	Neighborhood
M1A	Not assigned	
M1B	Scarborough	Malvern, Rouge
M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
M1E	Scarborough	Guildwood, Morningside, West Hill
M1G	Scarborough	Woburn
M1H	Scarborough	Cedarbrae
M1J	Scarborough	Scarborough Village

Figure 1. Example of the data available at https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Geospatial Coordinates information

The geospatial information of each neighborhood in Toronto will be extracted from the Geocoder platform (https://geocoder.readthedocs.io/index.html) (**2**). This information is contained in a csv file that is uploaded to the code.

List of Venues available at the FoursquareAPI

It will be used the List of venues by neighborhoods in Toronto provided by the FoursquareAPI. With this information, it is possible to find all the healthy food restaurants and cluster it by neighborhood (https://developer.foursquare.com/docs/api/venues/explore) (**3**). It will also be used the same Foursquare API to explore all the healthy food-related places, which include: Yoga places, Meditation places, Vegetarian/Healthy Restaurants, Gymnasium and Fitness Center.

Figure 2. Example of the venue information provided by the FoursquareAPI tool.

It is possible to identify and get access to category, type, location and name of the venues.

Additional information about Healthy Food Restaurants in Toronto

To complement our analysis and report, we could eventually use data from the scientific article: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5674765/

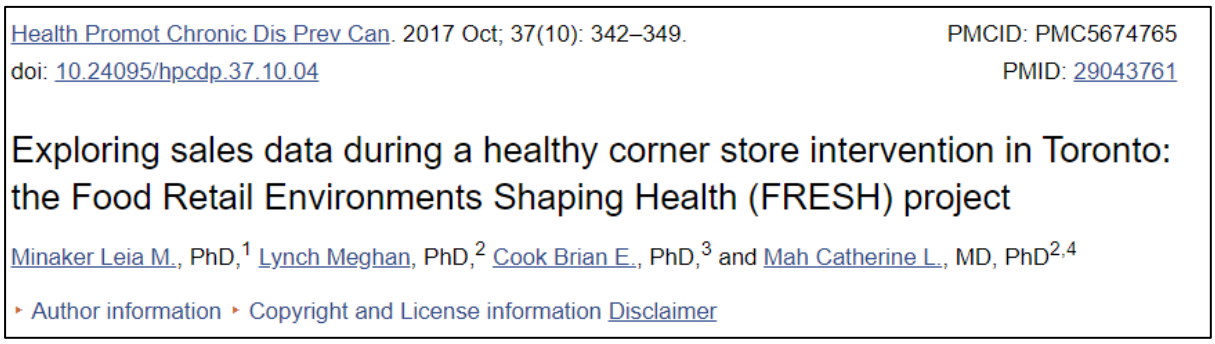

Figure 3. Scientific article about the Healthy Food markets in Toronto, CA that can be referenced in order to give some additional support to our report.

Other tools and library used

- Pandas library
- Numpy library
- Foursquare API
- Folium library
- Geopy library

3. Methodology

The methodology applied for this study is the CRISP-DM. The follow scheme shows the steps followed in our specific study case.

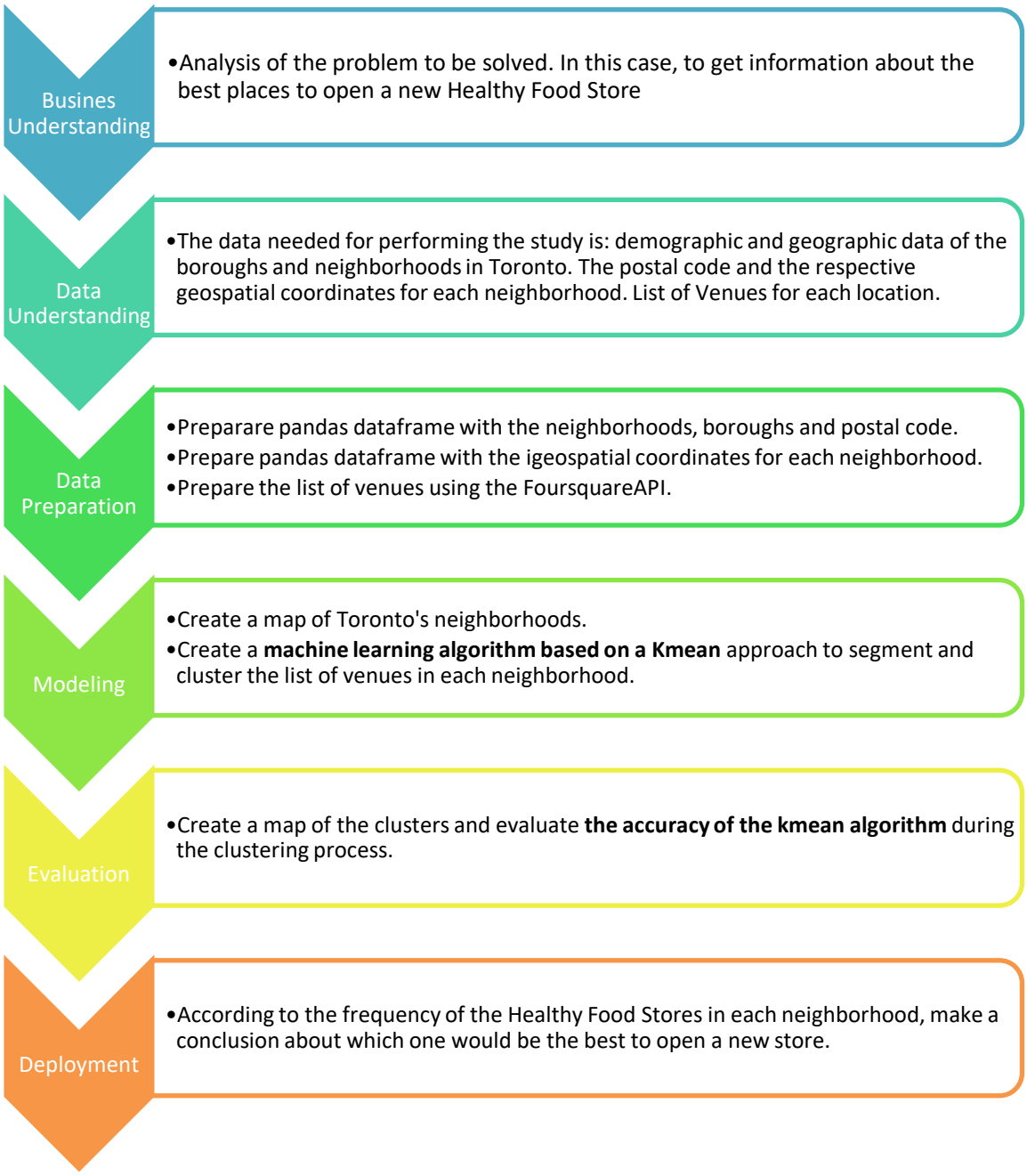

Figure 4. Methodology followed for performing the data analysis study case.

4. Results

Creating Dataframes in pandas

They were created different pandas dataframe in this project. Dataframe 1 (toronto_df, using data from **1**) and 2 (toronto_gsp_df, using data from **2**) contained the information about the

neighborhoods, potal code and boroughs of Toronto and the geospatial coordinates of these places. We merged into Dataframe 4 (toronto_code_df) to have all the data needed for the study.

Dataframe 1 (toronto_df)

```
[2]: toronto_df = pd.read_csv('Toronto_2.csv') #Create Dataframe 1 **(toronto_df)** with the zip codes of Toronto
toronto_df
```

[2]:

	Postal Code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
5	M1J	Scarborough	Scarborough Village
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount Park
7	M1L	Scarborough	Golden Mile, Clairlea, Oakridge
8	M1M	Scarborough	Cliffside, Cliffcrest, Scarborough Village West
9	M1N	Scarborough	Birch Cliff, Cliffside West
10	M1P	Scarborough	Dorset Park, Wexford Heights, Scarborough Town...

Dataframe 2 (toronto_gsp_df)

```
3]: toronto_gsp_df = pd.read_csv('Geospatial_Coordinates.csv') #Create Dataframe 2 **(toronto_gsp_df)** with the geospatial coordinates
toronto_gsp_df
```

3]:

	Postal Code2	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476
5	M1J	43.744734	-79.239476
6	M1K	43.727929	-79.262029

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	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	Kennedy Park, Ionview, East Birchmount 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

Figure 5. Head of the dataframe 1, 2 and 4 used for this study.

Create a map of Toronto's neighborhoods

After we had available all the information into a dataframe, we created a map of the neighborhoods of Toronto, in order to check the accuracy of the data. First, we needed to get the location of the city of Toronto using the geopy library and finally, using this information, the dataframe 4 and the Folium package we got the map.

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.
```

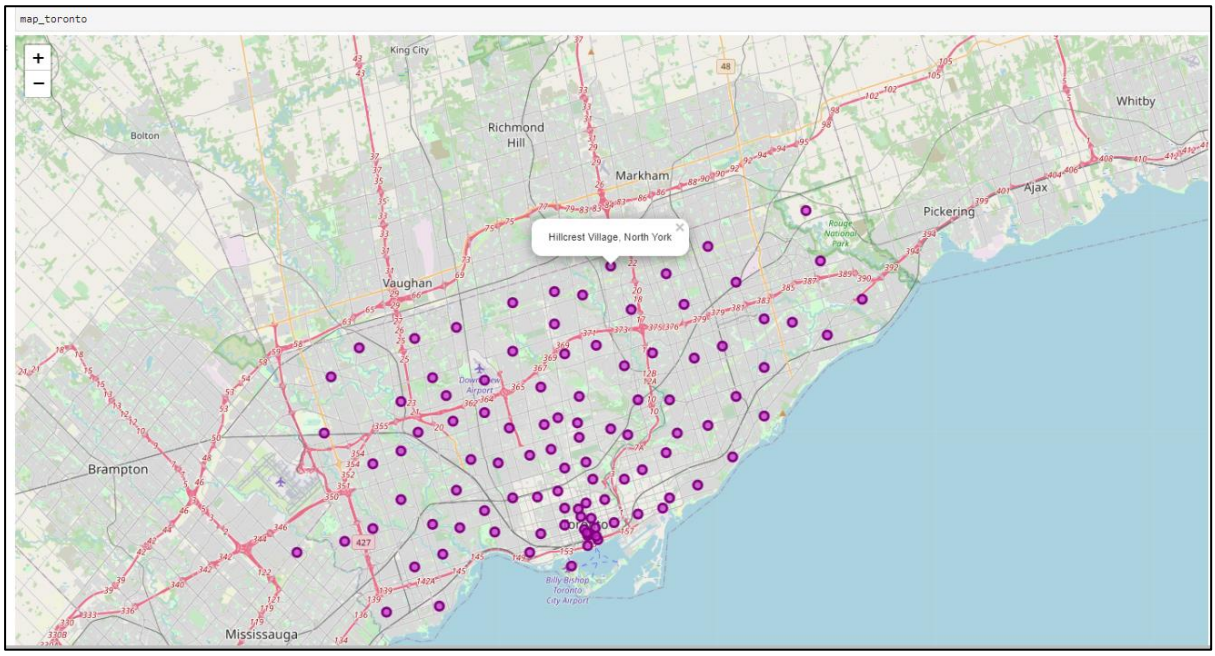

Figure 6. Creating a map of the neighborhoods in Toronto. First, we found the coordinates of Toronto using the geopy library and then, with the Folium library we draw the map.

As we can see the map created shows in purple circle markers the neighborhoods in Toronto. We selected one of them to see the type of label shown. A total of 102 markers are spread around the downtown, representing the 102 neighborhoods of the city. The density of neighborhoods is bigger around the center and decreases while we walked to the outer of it.

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

After the map is displayed, we used the Foursquare API to find the list of all the venues in Toronto, having as return a dataframe with the venues of every neighborhood.

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)
```

Figure 7. Dataframe resulting after using the exploring query of the Foursquare API.

Clustering the neighborhoods by venues

The dataframe gotten after using the explore query was re-organized to show the total of venues for by neighborhood and so, we could see the neighborhoods **Toronto Dominion Centre**, **Design Exchange**, **First Canadian Place**, **Underground city**, **Garden District**, **Ryerson, Harbourfront East, Union Station, Toronto Islands** and **Commerce Court, Victoria Hotel** showed the most venues with a total of 100 each one.

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

Figure 8. Neighborhoods with the most venues.

Machine learning approach to cluster the neighborhoods

The clustering of the neighborhoods according to the venues, to evaluate the best candidates to open a new Healthy Food store was made using the K-means algorithm. This algorithm is unsupervised, fast and easy to apply, so it is largely used in application where clustering and segmentation of people, locations and data in general is required. The study was performed with K = 5, which means the data was segmented in 5 different clusters

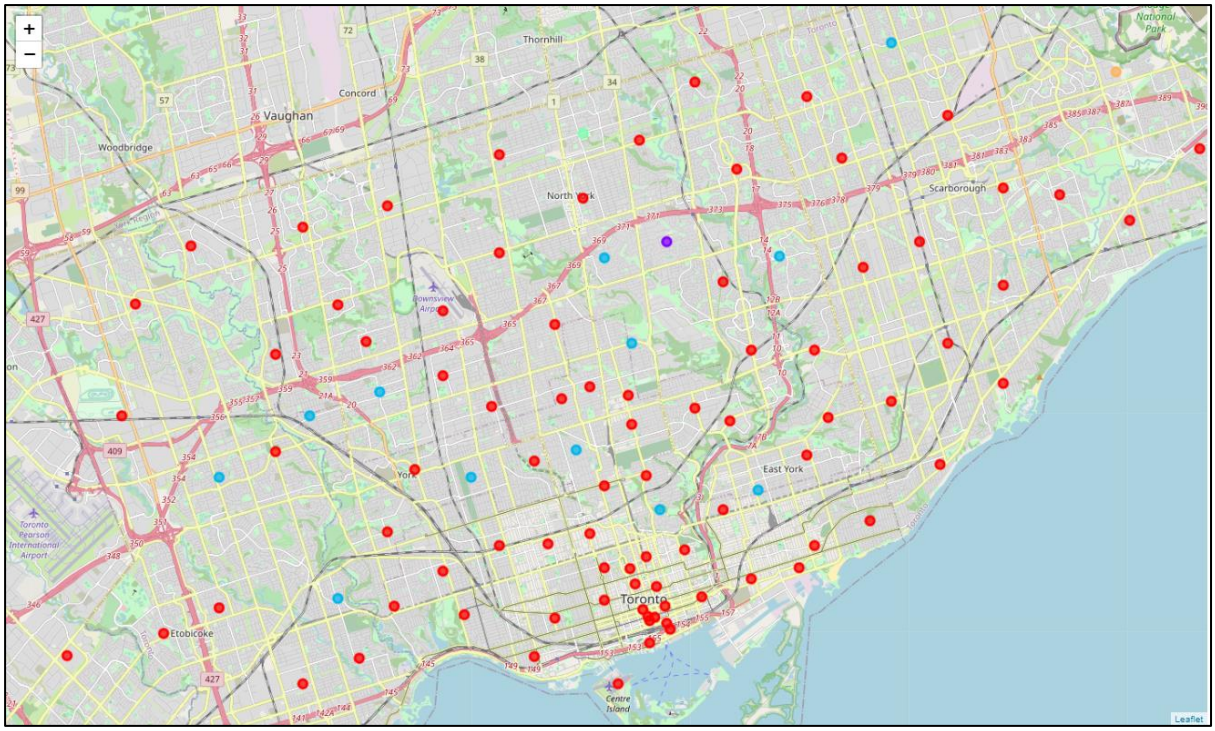

Figure 9. Map of the neighborhoods of Toronto clustered through the K-mean algorithm.

The visualization of the clusters in the map was made by the Folium package and it is possible to observe the 5 different color markers, although one of them, the red marker is largely representative.

5. 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. It includes food from different places of the world, being the Asian food restaurants very popular. 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
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0	Clothing Store	Coffee Shop	Fast Food Restaurant	Restaurant	Shoe Store	Bank	Japanese Restaurant	Sporting Goods Shop	Bus Station	Convenience Store
19	M2K	North York	Bayview Village	43.786947	-79.385975	0	Japanese Restaurant	Café	Bank	Chinese Restaurant	Department Store	Dim Sum Restaurant	Diner	Discount Store	Distribution Center	Women's Store
22	M2N	North York	Willowdale	43.770120	-79.408493	0	Ramen Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Café	Grocery Store	Restaurant	Sandwich Place	Bubble Tea Shop	Hotel
24	M2R	North York	Willowdale	43.782736	-79.442259	0	Ramen Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Café	Grocery Store	Restaurant	Sandwich Place	Bubble Tea Shop	Hotel
26	M3B	North York	Don Mills	43.745906	-79.352188	0	Gym	Asian Restaurant	Japanese Restaurant	Restaurant	Beer Store	Coffee Shop	Caribbean Restaurant	Athletics & Sports	Café	Sporting Goods Shop
27	M3C	North York	Don Mills	43.725900	-79.340923	0	Gym	Asian Restaurant	Japanese Restaurant	Restaurant	Beer Store	Coffee Shop	Caribbean Restaurant	Athletics & Sports	Café	Sporting Goods Shop
28	M3H	North York	Bathurst Manor, Wilson Heights, Downsview North	43.754328	-79.442259	0	Bank	Coffee Shop	Diner	Supermarket	Deli / Bodega	Sushi Restaurant	Middle Eastern Restaurant	Restaurant	Fried Chicken Joint	Pizza Place
29	M3J	North York	Northwood Park, York University	43.767980	-79.487262	0	Massage Studio	Falafel Restaurant	Coffee Shop	Furniture / Home Store	Caribbean Restaurant	Miscellaneous Shop	Bar	Discount Store	Dessert Shop	Dim Sum Restaurant
30	M3K	North York	Downsview	43.737473	-79.464763	0	Park	Grocery Store	Athletics & Sports	Liquor Store	Snack Place	Bus Stop	Baseball Field	Food Truck	Bank	Hotel
31	M3L	North York	Downsview	43.739015	-79.506944	0	Park	Grocery Store	Athletics & Sports	Liquor Store	Snack Place	Bus Stop	Baseball Field	Food Truck	Bank	Hotel
32	M3M	North York	Downsview	43.728496	-79.495697	0	Park	Grocery Store	Athletics & Sports	Liquor Store	Snack Place	Bus Stop	Baseball Field	Food Truck	Bank	Hotel
33	M3N	North York	Downsview	43.761631	-79.520999	0	Park	Grocery Store	Athletics & Sports	Liquor Store	Snack Place	Bus Stop	Baseball Field	Food Truck	Bank	Hotel
34	M4A	North York	Victoria Village	43.725882	-79.315572	0	French Restaurant	Hockey Arena	Coffee Shop	Intersection	Portuguese Restaurant	Deli / Bodega	Department Store	Dessert Shop	Dim Sum Restaurant	Diner
35	M4B	East York	Parkview Hill, Woodbine Gardens	43.706397	-79.309937	0	Pizza Place	Fast Food Restaurant	Pharmacy	Bank	Intersection	Athletics & Sports	Gastropub	Café	Pet Store	Gym / Fitness Center
36	M4C	East York	Woodbine Heights	43.695344	-79.318389	0	Park	Pharmacy	Beer Store	Skating Rink	Cosmetics Shop	Curling Ice	Video Store	Comfort Food Restaurant	Eastern European Restaurant	College Rec Center
37	M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Trail	Pub	Health Food	Distribution	Deli / Bodega	Department	Dessert Shop	Dim Sum	Diner	Discount Store

Figure 10. Part of the neighborhoods and venues of Cluster 1.

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

Figure 11. Part of the neighborhoods and venues of Cluster 2.

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

Figure 12. Part of the neighborhoods and venues of Cluster 3.

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

The **Cluster 4** 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

Figure 13. Part of the neighborhoods and venues of Cluster 4.

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

Figure 14. Part of the neighborhoods and venues of Cluster 5.

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. The cluster 1 is full of neighbourhood with many restaurants and although it was found any of these restaurants were of Healthy food, we can assume they have indeed healthy options in their menus, so they would be strong competitors for the new restaurant. 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.

6. 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.