

Final Capstone Project: The Battle of Neighborhoods

Section 1: Introduction

Vegetarians and vegans now account for nearly 10 per cent of Canada's population, and their increased presence is forcing the restaurant and meat production industries to consider new approaches.

According to research from Dalhousie University, there are 2.3 million vegetarians in Canada, up from 900,000 15 years ago. Another 850,000 people consider themselves vegan. Those two numbers add up to 9.4 per cent of the Canadian population. The interested readers could refer to the link below:

<https://www.ctvnews.ca/canada/more-than-3-million-canadians-vegetarian-or-vegan-study-1.4027606>

Since the most populated city in Canada is Toronto. It is home to 5,429,524 people. Toronto is the capital of Ontario and located in the east-central region of the country. Nearly half of the population is made up of foreign-born residents. After Miami in the US, this is the second largest percentage of foreign-born residents in the world. Interestingly, no nationality holds the dominant position, making Toronto the most diverse city in the world. Therefore we are going to explore Toronto for our business prospects. The interested readers could refer to the link below:

<https://www.worldatlas.com/articles/biggest-cities-in-canada.html>

Section 2: Business Problem and Interested Community

With increasing number of people turning vegetarian and vegan, there is a good scope of opening the Vegetarian/Vegan Restaurant. Based on this we define the following:

Problem: The aim of this project is to explore the possibility of opening Fruit & Vegetable Stores near Vegetarian/Vegan Restaurants in Canada.

Interested Community: Food Business Industry

1. The results of this project could be of interest to the investors who wants to invest in a Food Business.
2. Since the vegetarian/vegan restaurants would provide good retail opportunities for Fruits & Vegetables Store. Therefore, the results could be of interest to the investors who wants to invest in these stores.
3. The Juice Bars are also potential customers for Fruits & Vegetables Store.

Section 3: Data Sources and their description

In this section we are going to provide information about the sources from where data can be downloaded and their description.

a. Toronto neighborhood data:

Data Source:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Description:

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the Borough and their neighborhoods in Toronto.

b. Vegetarian/Vegan restaurants and Juice Bars in each Borough and their neighborhoods of Toronto.

Data source:

Fousquare API

<https://foursquare.com/>

Description:

By using this API we will get all the venues in each neighborhood. We can filter these venues to get explore the possibilities of opening Vegetarian/Vegan restaurants.

c. Geospatial Coordinates

Data source:

http://cocl.us/Geospatial_data

Description:

To utilize the Foursquare location data, we will get the latitude and the longitude coordinates of each neighborhood.

Section 4: Methodology

Before going into details, let's put down some steps:

A. Data Loading and Extraction:

Step 1: Importing all the necessary libraries

Step 2: Web Scraping using BeautifulSoup: To obtain information regarding Postal Code Borough, Neighbors

B. Data Preprocessing:

Step 3: Data Preprocessing and Cleaning: Formatting the text and remove the duplicate and Null entries

C. Exploratory Data Analysis:

Step 4: Obtaining geospatial coordinates: To obtain information regarding Latitudes and Longitudes corresponding to Postal Codes

Step 5: Using Folium: To visualize Leaflet map of Toronto via Folium

Step 6: Foursquare API: To obtain the information regarding point of interest in detail. Here we are particularly looking at Vegetarian/ Vegan restaurants and Juice Bars.

D. Machine Learning Technique:

Step 7: K-means Clustering: Here we use unsupervised machine learning techniques

known as K-means clustering as the aim of this project is to find the geo-spatial coordinates for Fruits and Vegetables Store based on cluster analysis.

Now discuss these steps in detail.

A. Data Loading and Extraction:

Step 1: Importing all the necessary libraries:

We have imported all the necessary libraries. These along with their description are shown in the table below.

Library	Description
numpy:	to handle data in a vectorized manner
BeautifulSoup:	for web scrapping
Pandas:	library for data analysis
Json:	library to handle JSON files
Geopy:	To convert an address into latitude and longitude values
requests :	library to handle requests
Matplotlib and associated plotting modules	A comprehensive library for creating static, animated, and interactive visualizations in Python
sklearn:	to use k-means from clustering stage
folium:	map rendering library

Step 2: Web Scrapping using BeautifulSoup: In this step we have scraped website and parsed HTML code using the Python package BeautifulSoup, and converted data into a pandas dataframe.

Web scraping, web harvesting, or web data extraction is data scraping used for extracting data from websites. Web scraping software may access the World Wide Web directly using the Hypertext Transfer Protocol, or through a web browser.

Since we chose to explore Toronto as a potential place to open Fruits and Vegetable Store. For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We scrape the Wikipedia page and wrangle the data, clean it, and then read it into a *pandas* dataframe so that it is in a structured format. The output of this step is as shown below:

	Postcode	Borough	Neighbourhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor
7	M7A	Downtown Toronto	Queen's Park
8	M8A	Not assigned	Not assigned
9	M9A	Etobicoke	Islington Avenue

Our dataframe consist of three columns: PostalCode, Borough, and Neighborhood and it is structured.

But it contains “ Not assigned” entries and then certain post codes which are occurring more than once. e.g., M6A. It means our data will require further processing and cleaning which is done in Section below:

B. Data Preprocessing:

We have cleaned the dataframe obtained from the above step:

1. Removing the entries of Borough which is not assigned.
2. Assigning the neighborhood as a borough, where neighborhood is Not assigned
3. Grouping by Postcode to remove duplicate entries

Following the above steps we get the data frame which is clean as shown below :

	Postcode	Borough	Neighbourhood
0	M1B	Scarborough	Rouge, Malvern
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

As can be seen now M1C and M1E are combined into one row with the neighborhoods separated with a comma. Now we have unique Postal codes.

C. Exploratory Data Analysis:

In this section we will do analysis if it make sense to choose Toronto as a place for opening Fruits and Vegetable Store. It is always better to do this kind of analysis as without having the insight of Problem in hand , we may end up with a model which may not produce desired results.

Step 4: Geospatial Coordinates:

Here's we obtain latitude and longitude:

Latitude lines run east-west and are parallel to each other.

Longitude lines run north-south.

These values are obtained from the following link for Toronto:

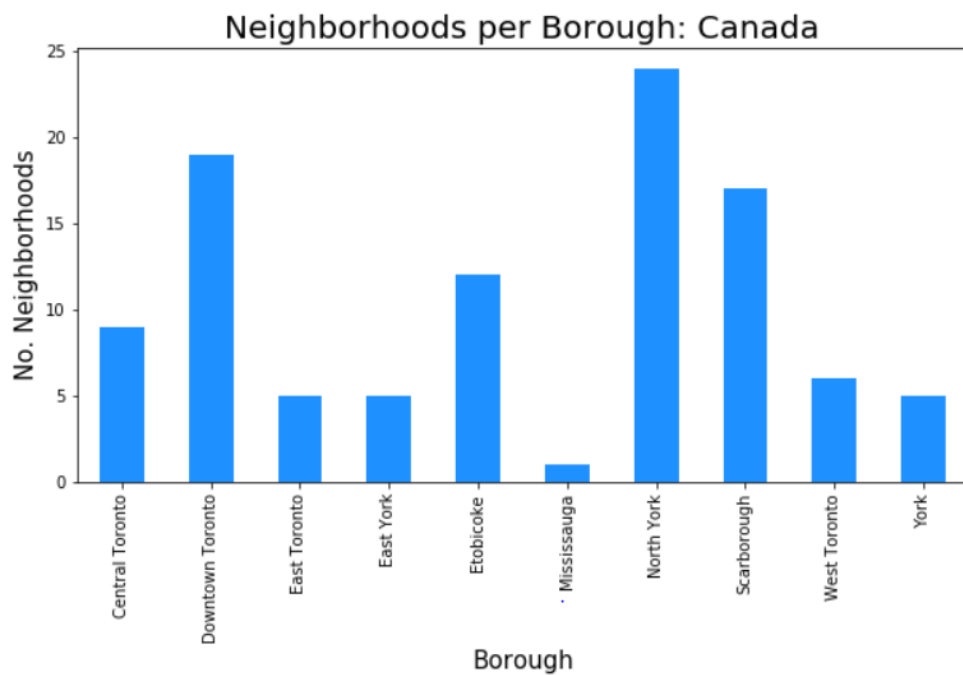
http://cocl.us/Geospatial_data

	Postal Code	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

After merging the data frame of geospatial coordinates with that of our cleaned dataframe, we obtained the following:

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	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

Then we tried to find out the neighbors per Borough.



Findings: As can be seen in the above Graph, if we combine different Borough of 'Toronto', the number of neighbors are going to be highest. Therefore, we explored and clustered the neighborhood of Toronto further.

We obtained the following data frame after filtering the boroughs that contain the word Toronto

	Postcode	Borough	Neighbourhood	Latitude	Longitude
0	M4E	East Toronto	The Beaches	43.676357	-79.293031
1	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
2	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
3	M4M	East Toronto	Studio District	43.659526	-79.340923
4	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Step 5: Using Folium:

We used Folium that builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library.

In this step we manipulated our data in Python, then visualized it in on a Leaflet map via Folium. In other words, we created a map of Toronto with neighborhoods superimposed on top.



Step 6: Foursquare API:

By now we have built a dataframe of the postal code of each neighborhood along with the borough name and neighborhood name. In order to utilize the Foursquare location data, we have also got the latitude and the longitude coordinates of each neighbourhood for Toronto.

In this step we utilized the Foursquare API to explore the neighborhoods and segment them based on vegetarian/vegan restaurants and Juice Bars. We obtained the information about Venue, its latitude, longitude and category as shown below:

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West, Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

We have find out the unique categories of venues that can be curated from all the returned venues. The categories which are of interest to us are:

1. Vegetarian / Vegan Restaurants
2. Juice Bars

We have explored these categories one by one.

Vegetarian / Vegan Restaurants : We have started with the Vegetarian / Vegan Restaurants.

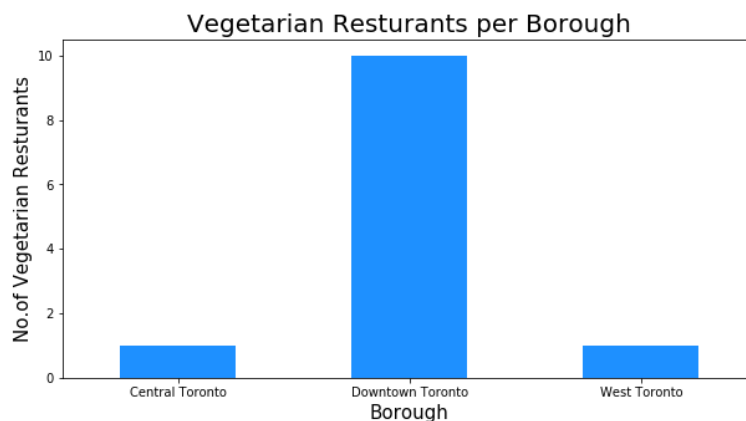
We have used FourSquare API to obtain the Postcode, Borough, Neighborhood, ID, Name, Latitude and Longitude of these Restaurants as shown below:

	Postcode	Borough	Neighbourhood	ID	Name	Latitude	Longitude
0	M5E	Downtown Toronto	Berczy Park	5b5bca904aa3f8002c97f85d	Fresh On Front	43.644771	-79.373306
1	M5H	Downtown Toronto	Adelaide, King, Richmond	5aff06ca6e4650002cc6286b	Rosalinda	43.650571	-79.384568
2	M5H	Downtown Toronto	Adelaide, King, Richmond	5c9d602f25fb7b002c79b669	Planta Queen	43.650571	-79.384568
3	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	5481b06c498ee191fa045a00	Kupfert & Kim	43.640816	-79.381752
4	M5R	Central Toronto	The Annex, North Midtown, Yorkville	4ad4c061f964a52099f720e3	Live Organic Food Bar	43.672710	-79.405678

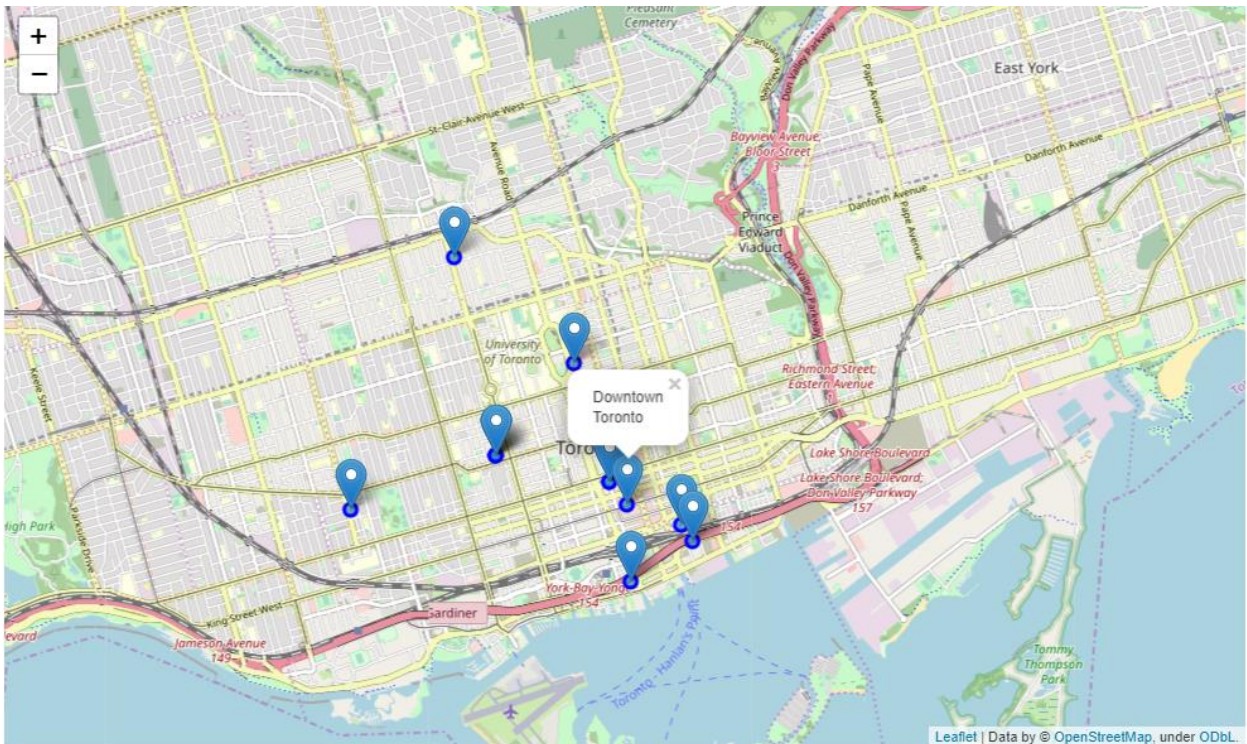
Findings: We have found out that out of 39 Neighbor counts of Toronto, only 12 neighbor counts having Vegetarian/Vegan restaurants.

To answer the question which Borough has the highest count of vegetarian restaurant, we explored the number of Vegetarian/Vegan restaurants per Borough as shown below:

	Borough	ID
0	Central Toronto	1
1	Downtown Toronto	10
2	West Toronto	1



Findings: From the bar graph it makes sense if we open the Fruits and Vegetables store in Downtown Toronto as it has the highest number of Vegetarian Restaurants. This can be also visualized on the Map as shown below:



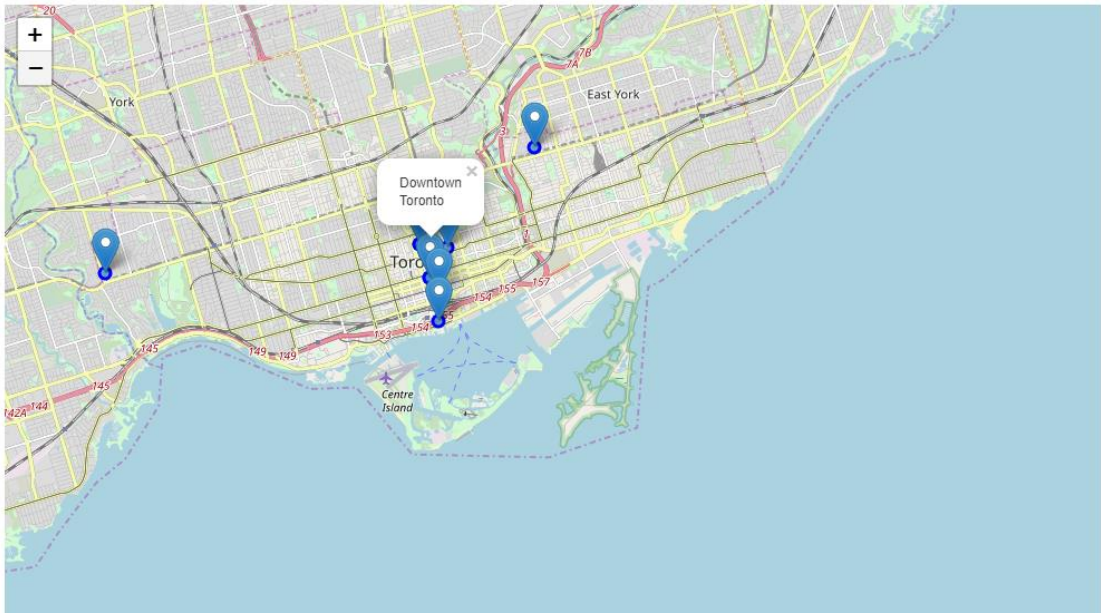
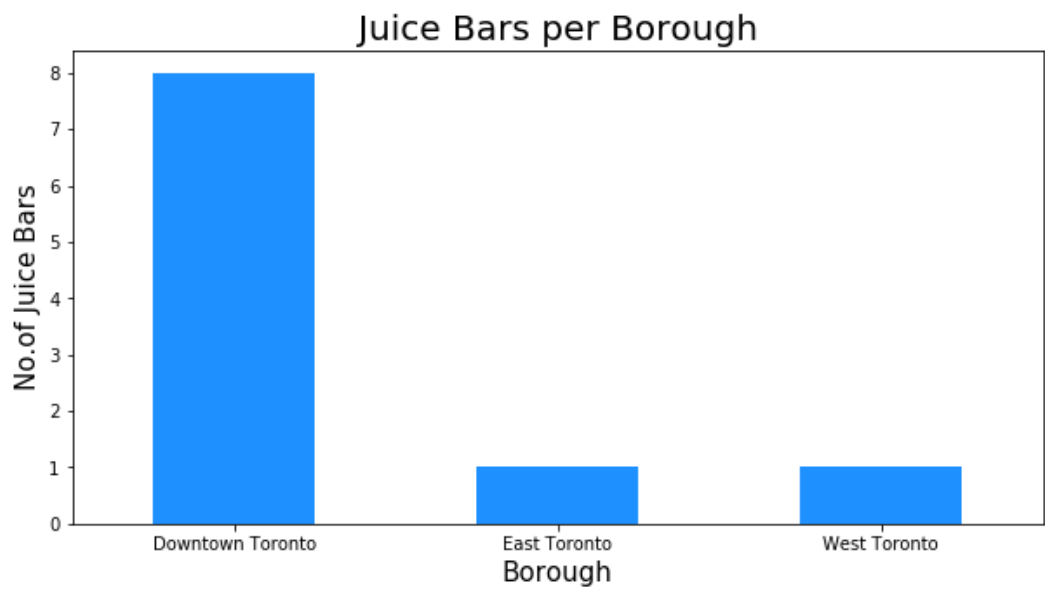
Juice Bars: We have explored the neighborhood of Downtown further to look for the more opportunities for business, i.e., Juice Bars.

Next we moved towards getting details about the Juice Bar in Toronto

	Postcode	Borough	Neighbourhood	ID	Name	Latitude	Longitude
0	M4K	East Toronto	The Danforth West, Riverdale	4bd1db4e046076b0f1a57271	The Big Carrot Organic Juice Bar	43.679557	-79.352188
1	M4Y	Downtown Toronto	Church and Wellesley	50c20a89e4b02803a189384b	FUEL+	43.665860	-79.383160
2	M5B	Downtown Toronto	Ryerson, Garden District	4bdf1c4bffdec92826c0eaa1	Booster Juice	43.657162	-79.378937
3	M5G	Downtown Toronto	Central Bay Street	4b9a73ecf964a5209db635e3	Booster Juice	43.657952	-79.387383
4	M5G	Downtown Toronto	Central Bay Street	5becb2b274a002c3ef499	Booster Juice	43.657952	-79.387383

Findings: Out of 39 Neighbour counts of Toronto, we are left only with 10 neighbor counts having Juice Bars

	Borough	ID
0	Downtown Toronto	8
1	East Toronto	1
2	West Toronto	1



Findings: As can be seen from the results, the number of Juice Bars are also highest in Downtown Toronto.

In the beginning of the Problem, we defined that these two could be our potential customers. And for both the potential customers, Downtown Toronto is the good choice. This completes our analysis approach. Now after having an idea we tried to find out what comes from the results of machine learning.

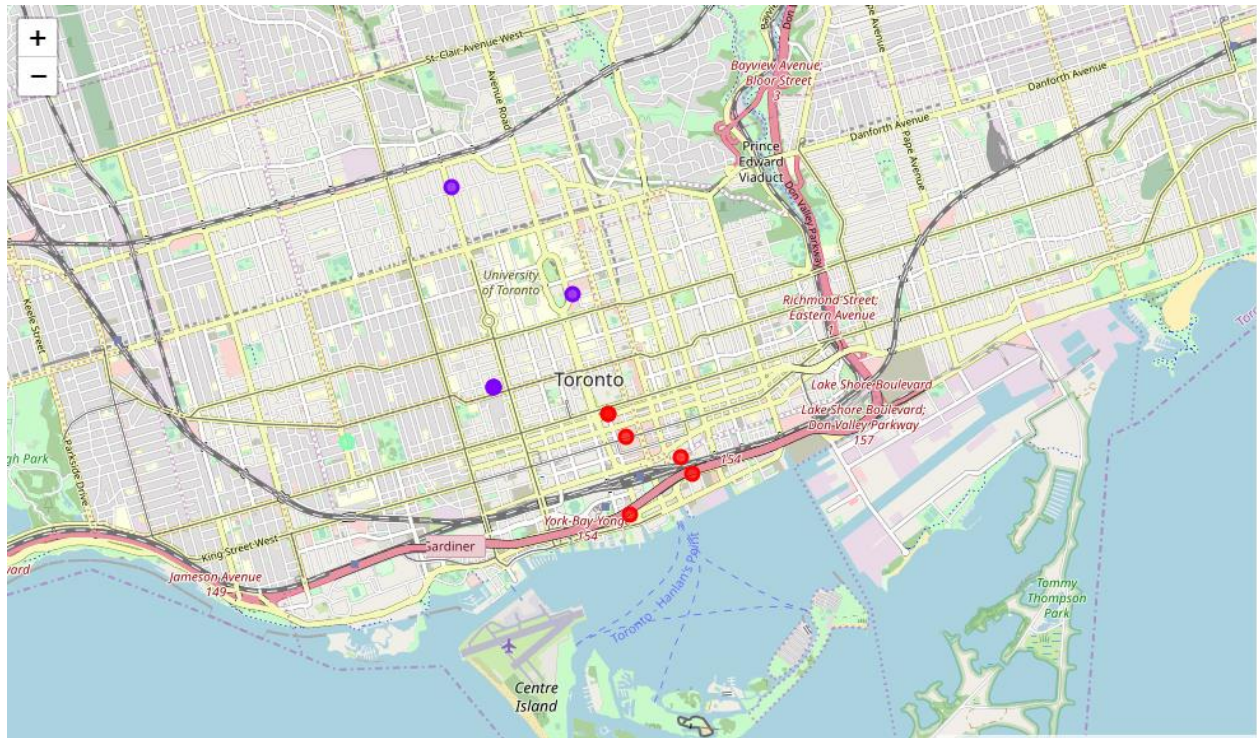
D. Machine Learning Technique

- **K-means clustering:** We choose K-means clustering because our problem is related to cluster analysis.
- This can be seen from the results of Exploratory data Analysis also, that in our problem we have clusters corresponding to different Boroughs.

Cluster Vegetarian/Vegan Restaurants

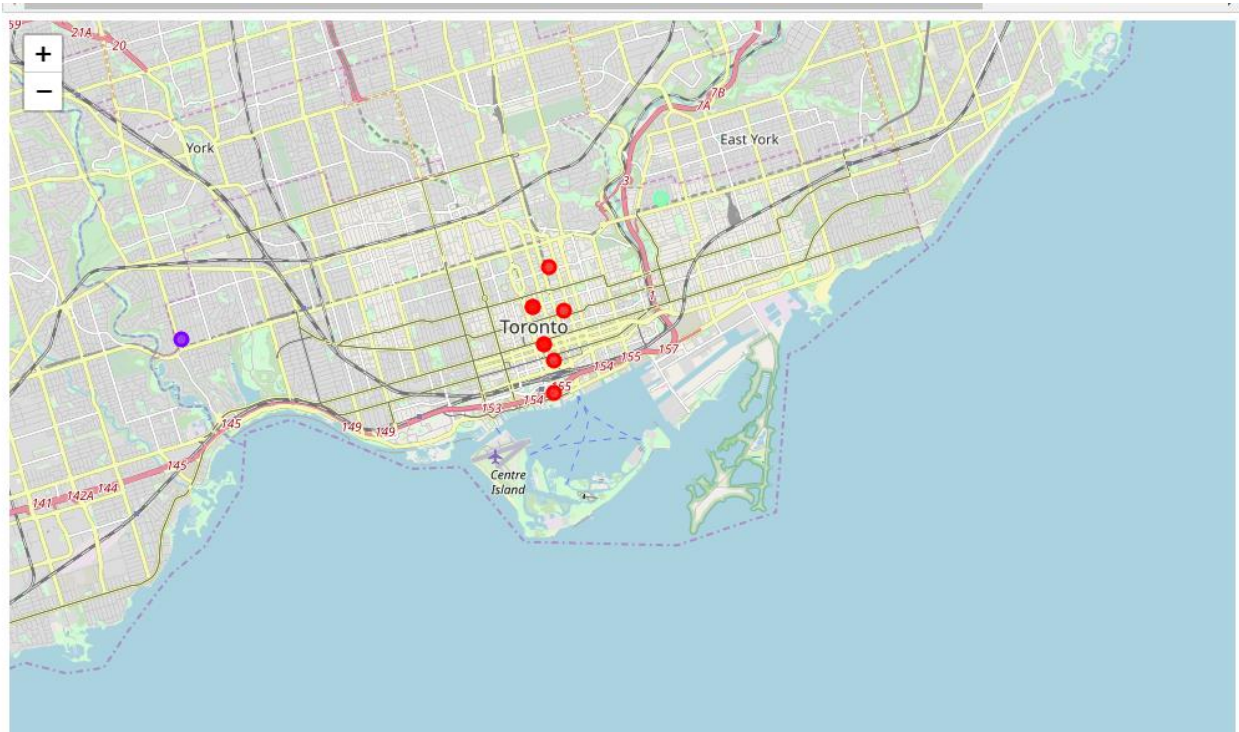
As in the Exploratory Data Analysis section we started with the Vegetarian Restaurants and then moved to Juice Bars. By choosing k=3 we get the results shown below:

	Cluster Labels	Postcode	Borough	Neighbourhood	ID	Name	Latitude	Longitude
0	0	M5E	Downtown Toronto	Berczy Park	5b5bca904aa3f8002c97f85d	Fresh On Front	43.644771	-79.373306
1	0	M5H	Downtown Toronto	Adelaide, King, Richmond	5aff06ca6e4650002cc6286b	Rosalinda	43.650571	-79.384568
2	0	M5H	Downtown Toronto	Adelaide, King, Richmond	5c9d602f25fb7b002c79b669	Planta Queen	43.650571	-79.384568
3	0	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	5481b06c498ee191fa045a00	Kupfert & Kim	43.640816	-79.381752
4	1	M5R	Central Toronto	The Annex, North Midtown, Yorkville	4ad4c061f964a52099f720e3	Live Organic Food Bar	43.672710	-79.405678
5	1	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	4b2d2ab2f964a52007d024e3	Hibiscus	43.653206	-79.400049
6	1	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	4ae4e982f964a5205d9f21e3	Buddha's Vegetarian	43.653206	-79.400049
7	1	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	4ad7e7acf964a520f40f21e3	Urban Herbivore	43.653206	-79.400049
8	0	M5W	Downtown Toronto	Stn A PO Boxes 25 The Esplanade	5b5bca904aa3f8002c97f85d	Fresh On Front	43.646435	-79.374846
9	0	M5X	Downtown Toronto	First Canadian Place, Underground city	5aff06ca6e4650002cc6286b	Rosalinda	43.648429	-79.382280
10	2	M6J	West Toronto	Little Portugal, Trinity	52f652a5498ec43f09d514c4	The Goods	43.647927	-79.419750
11	1	M7A	Downtown Toronto	Queen's Park	540e1fa1498e9b339bc79212	The Green Beet Cafe	43.662301	-79.389494



Cluster Juice Bars

	Cluster Labels	Postcode	Borough	Neighbourhood	ID	Name	Latitude	Longitude
0	2	M4K	East Toronto	The Danforth West, Riverdale	4bd1db4e046076b0f1a57271	The Big Carrot Organic Juice Bar	43.679557	-79.352188
1	0	M4Y	Downtown Toronto	Church and Wellesley	50c20a89e4b02803a189384b	FUEL+	43.665860	-79.383160
2	0	M5B	Downtown Toronto	Ryerson, Garden District	4bdf1c4bfdec92826c0eaa1	Booster Juice	43.657162	-79.378937
3	0	M5G	Downtown Toronto	Central Bay Street	4b9a73ecf964a5209db635e3	Booster Juice	43.657952	-79.387383
4	0	M5G	Downtown Toronto	Central Bay Street	5beccbec2b274a002c3ef499	Booster Juice	43.657952	-79.387383
5	0	M5H	Downtown Toronto	Adelaide, King, Richmond	5548cfb6498ec3e4bdfde59a	Daily Press Juicery	43.650571	-79.384568
6	0	M5H	Downtown Toronto	Adelaide, King, Richmond	4dcd355452b1f8915b7710e8	Booster Juice	43.650571	-79.384568
7	0	M5J	Downtown Toronto	Harbourfront East, Toronto Islands, Union Station	4bbcd229593fef3b77ab0256	Booster Juice	43.640816	-79.381752
8	0	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	4dcd355452b1f8915b7710e8	Booster Juice	43.647177	-79.381576
9	1	M6S	West Toronto	Runnymede, Swansea	4b3a379bf964a520696225e3	Booster Juice	43.651571	-79.484450



Section 5: Results

Vegetarian/ vegan Restaurants : We analyzed the results obtained from Vegetarian/ vegan Restaurants and found the number of Borough corresponding to different cluster Labels.

Cluster Labels		Borough
0	0	6
1	1	5
2	2	1

Furthermore, we found that Label 0 and 1 corresponds to Downtown Toronto.

```
Cluster Labels
0    Downtown Toronto
1    Downtown Toronto
2         West Toronto
Name: Borough, dtype: object
```

We found the mean coordinates corresponding to cluster Labels to find the centroid where we can open our Food Store.

	Latitude	Longitude
Cluster Labels		
0	43.646932	-79.380220
1	43.658926	-79.399064
2	43.647927	-79.419750

Findings

Based on the vegetarian restaurants, the best location to open Fruits and Vegetable Stores is Downtown Toronto with latitude and longitude: (43.646932,-79.380220)

Juice Bars: Similar results are obtained for the Juice Bars, where Label 0, corresponds to the maximum number of Juice Bar.

Cluster Labels	ID
0	0 8
1	1 1
2	2 1

Furthermore, we found that Label 0 corresponds to Downtown Toronto as shown below:

```
Cluster Labels
0    Downtown Toronto
1         West Toronto
2         East Toronto
Name: Borough, dtype: object
```


We found the mean coordinates corresponding to cluster Labels to find the centroid where we can open our Food Store

	Latitude	Longitude
Cluster Labels		
0	43.651571	-79.484450
1	43.653508	-79.383666
2	43.679557	-79.352188

Findings: Based on the Juice Bars, the best location to open Fruits and Vegetable Stores is Downtown Toronto with latitude and longitude: (43.651571, -79.484450)

Section 6: Discussion

Based on our exploratory data analysis results and k-clustering algorithm: We have seen that in both the cases, i.e., for vegetarian/ vegan restaurants and Juice Bars: Downtown Toronto is the best option to open Fruits and Vegetable Stores.

We can say that the best place to open the Store can be chosen based on the centroid location of Vegetarian /Vegan Restaurant or on the basis of Juice Bar centroid.

Section 7: Conclusion

We have seen, that in both the cases, i.e., for vegetarian/ vegan restaurants and Juice Bars:

Downtown Toronto is the best option to open Fruits and Vegetable Stores.

Here, we have analyzed two prospects namely Vegetarian/ Vegan Restaurants but it can be explored further based on Logistics such as Airport and Boat and Ferry venues.