Final Capstone Project: The Battle of Neighborhoods

Sukhpreet Kaur Khangura

Introduction

- Vegetarians and vegans now account for nearly 10 per cent of Canada’s population.
- According to research from Dalhousie University, there are 2.3 million vegetarians in Canada, up from 900,000 15 years ago.
- The most populated city in Canada is Toronto.
- 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

Business Problem and Interested Community

- **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.

Data Sources and their description

- **Toronto neighborhood data:**

Data Source:

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

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

Data source:

Fousquare API

https://foursquare.com/

- **Geospatial Coordinates**

Data source:

http://cocl.us/Geospatial_data

Methodology

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

Data Preprocessing:

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

Methodology

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:*** Foursqaure API: To obtain the information regarding point of interest in detail. Here we are particulary looking at Vegetarian/ Vegan restaurants and Juice Bars.

Methodology

- **Machine Learning Technique:**
- ***Step 7:*** 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.

Data Loading and Extraction:

Step 1: Importing all the necessary libraries

Library	Description
numpy:	to handle data in a vectorized manner
BeautifulSoup:	for web scrapping
Pandas:	library for data analsysis
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

Data Loading and Extraction:

- ***Step 2:*** Web Scraping using Beautifulsoup

	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

Data Preprocessing:

- 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

	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

Exploratory Data Analysis

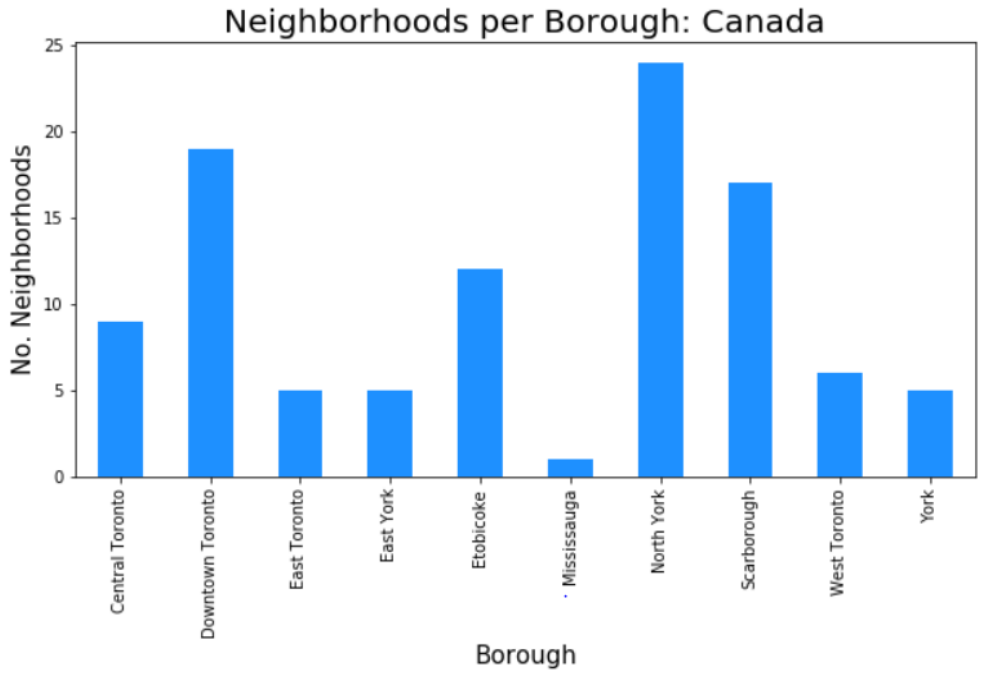
- ***Step 4:*** Geospatial Coordinates

	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

Exploratory Data Analysis



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.

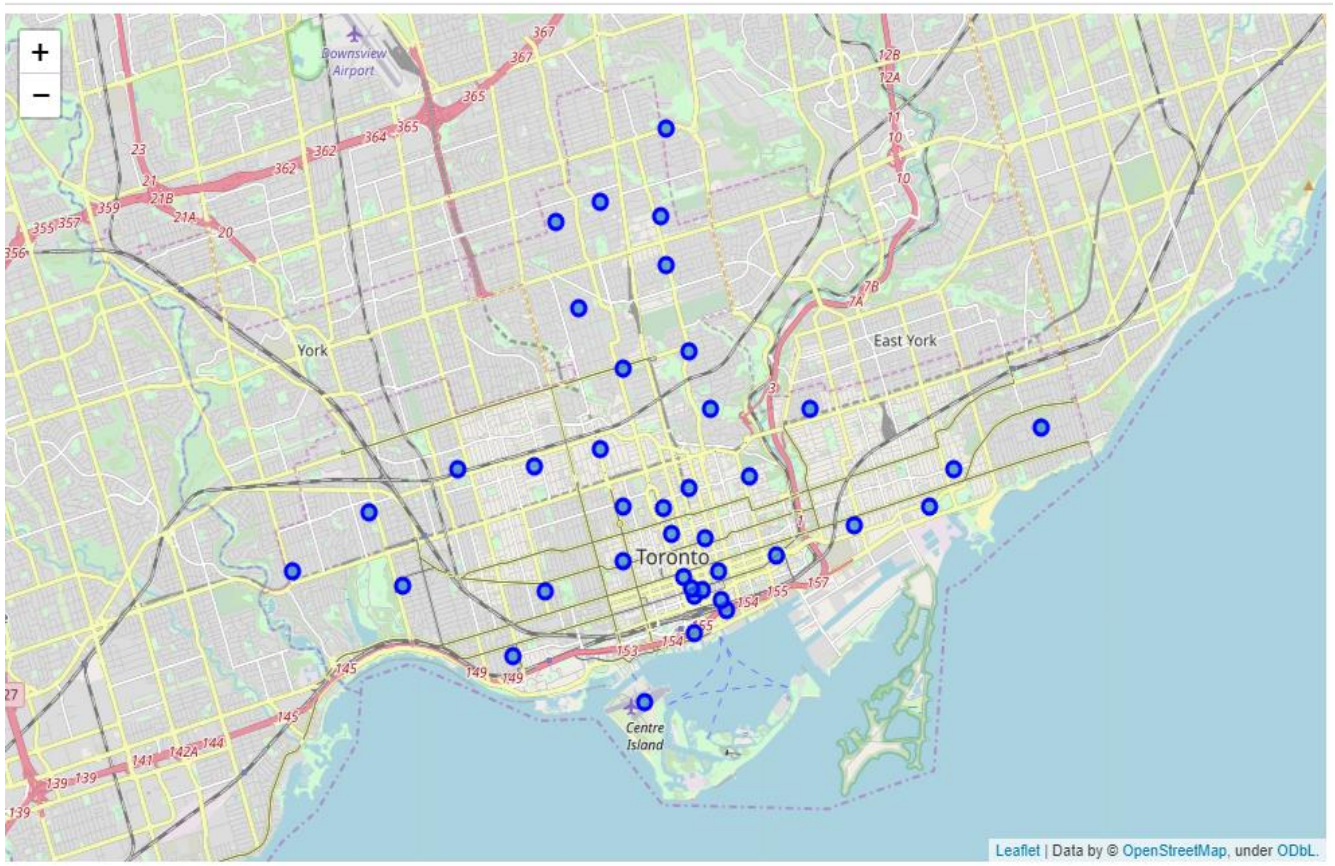
Exploratory Data Analysis

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

Exploratory Data Analysis

- ***Step 5:*** Using Folium:



Exploratory Data Analysis

- **Step 6:** Foursqaure API:

	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

The categories which are of interest to us are:

- Vegetarian / Vegan Restaurants
- Juice Bars

Exploratory Data Analysis

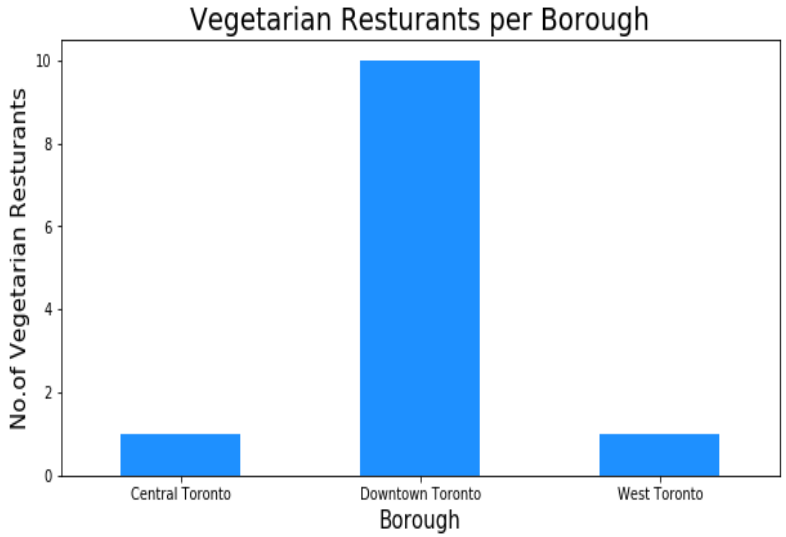
- **Vegetarian / Vegan Restaurants :**

	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.

Exploratory Data Analysis

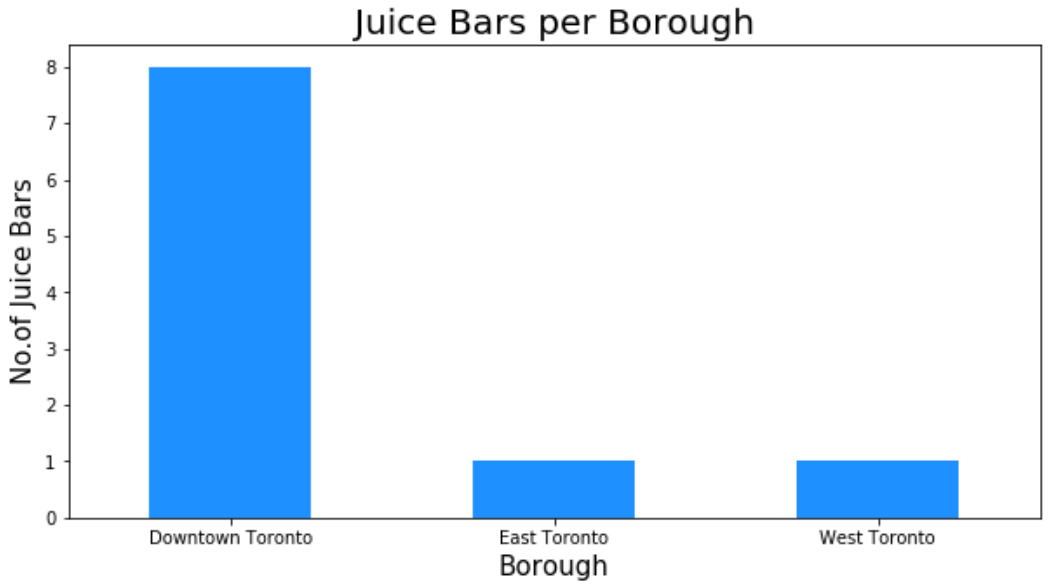
- **Vegetarian / Vegan Restaurants :**



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

Exploratory Data Analysis

- Juice Bars



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

Exploratory Data Analysis

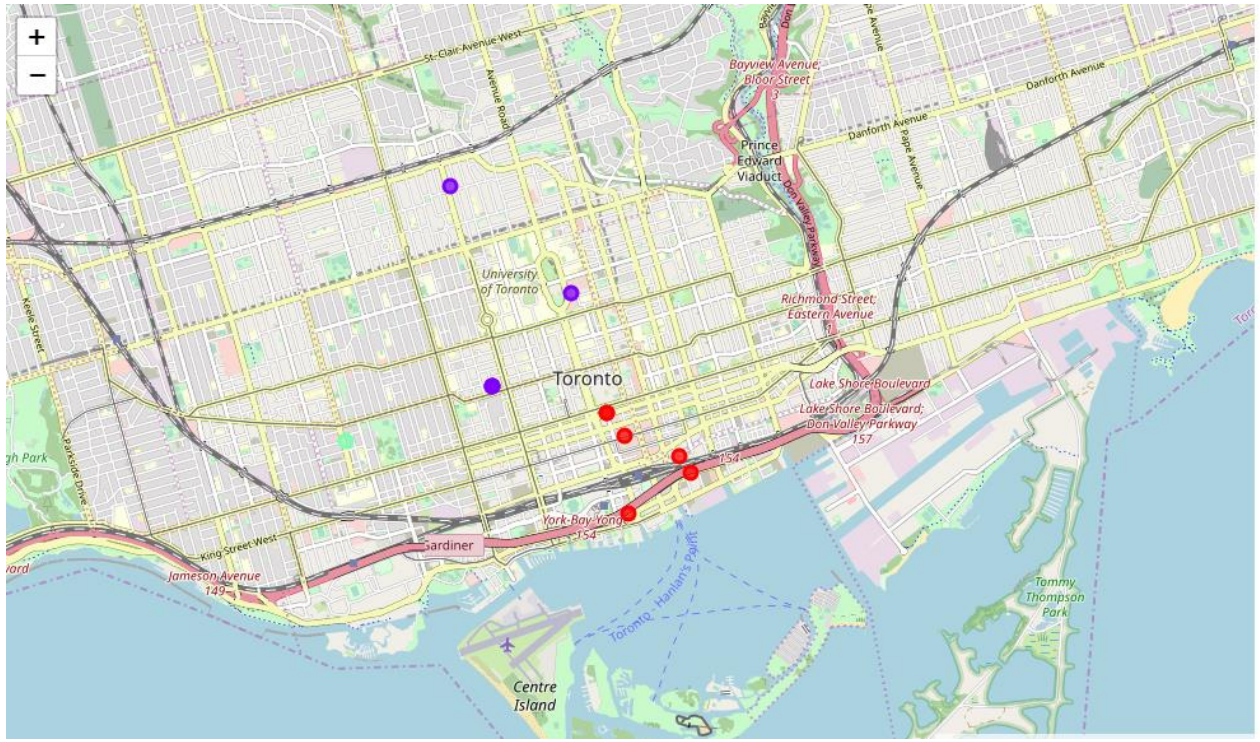
- In the beginning of the Problem, we defined that these two could be our potential customers.
- And for both the potential customers, Downtown Toronto is the good choice.
- This completes our analysis approach. Now after having an idea we tried to find out what comes from the results of machine learning.

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.

K-means clustering

- **Cluster Vegetarian/Vegan Restaurants**



K-means clustering

- **Cluster Juice Bars**

Results

- **Vegetarian/ vegan Restaurants**

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

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

Cluster Labels	Latitude	Longitude
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 Vegeatable Stores is Downtown Toronto with latitude and longitude: (43.646932,-79.380220)

Results

- **Juice Bars:**

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

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

Cluster Labels	Latitude	Longitude
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)

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.

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.