# Capstone Project — The Battle of Neighborhoods (Week 2)


## Heru Leonardo


# Introduction

Toronto, the most populous city in Canada, is an international center of business, finance, arts, and culture. Its economy is highly diversified with strengths, such as technology, financial services, education, art, and tourism. [1] In the city of Toronto, A Recommender System for Groceries contractor is design to know the correlation between distance and venue. For someone who is looking for a groceries contractor, it is vital to choose the neighborhood and retail location. The goal of this project is to figure out where a Groceries should be set up for success with data analysis.

# Data Acquisition

1. Neighborhoods in Toronto — Wikipedia[2].
2. Using Geopy to get geological location by address name.
3. Using Foursquare API to get the most common venues of given Borough of Toronto.
4. Using Foursquare API to get the venues' record of given venues of Toronto.

# Methodology

1. Scrape the Wikipedia page [2] and transform it into a pandas dataframe. Postal codes beginning with M are located within the city of Toronto. The dataframe consists of three columns: 'Postcode', 'Borough', 'Neighborhood'. Since some cells are 'Not assigned', I drop the cells with a borough that is 'Not assigned', and change the cells' neighborhood to borough for cells having a 'Not assigned' neighborhood. I also combine Neighborhoods with the same postal code. The dataframe cleaned has three columns and 103 rows. The first five rows are shown below.

Out[13]:

| | Postalcode | Borough | Neighborhood |
|---|---|---|---|
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Regent Park, Harbourfront |
| 3 | M6A | North York | Lawrence Manor, Lawrence Heights |
| 4 | M7A | Downtown Toronto | Queen's Park, Ontario Provincial Government |

```
In [14]: df_merge.shape
```

```
Out[14]: (103, 3)
```

2. Acquire the data of latitude and the longitude coordinates in Toronto using Geopy. And combine the geological location of each neighborhood with the dataframe above. Here I have the dataset that contains Postcode, Borough, Neighborhood, Latitude, Longitude of each neighborhood. The table below shows the first rows of the new dataframe. To brief the project, I choose only to analyze the boroughs that contain the word 'Toronto'.

Out[19]:

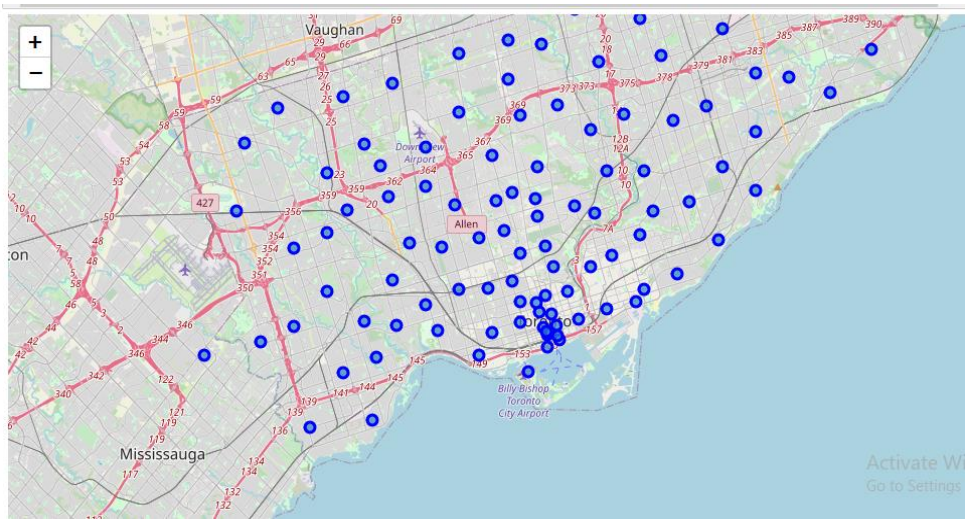
| | Postalcode | Latitude | Longitude | Borough | Neighborhood |
|---|---|---|---|---|---|
| 0 | M1B | 43.806686 | -79.194353 | Scarborough | Malvern, Rouge |
| 1 | M1C | 43.784535 | -79.160497 | Scarborough | Rouge Hill, Port Union, Highland Creek |
| 2 | M1E | 43.763573 | -79.188711 | Scarborough | Guildwood, Morningside, West Hill |
| 3 | M1G | 43.770992 | -79.216917 | Scarborough | Woburn |
| 4 | M1H | 43.773136 | -79.239476 | Scarborough | Cedarbrae |

```
In [20]: geo_data=geo_merged[['Postalcode','Borough','Neighborhood','Latitude','Longitude']]
geo_data.head()
```

Out[20]:

| | Postalcode | 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 |

3. Explore the boroughs and neighborhood in Toronto using Foursquare API. The limit is set as 100 venues, and the radius is set as 500 meters for each borough from their given latitude and longitude. As the header of the dataset shows below, the dataset contains Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Venue Latitude, Venue Longitude, Venue Category.




4. Explore the groceries in Toronto using Foursquare API. I got the neighborhood and bookstore information, sorted by the number of groceries in a neighborhood. The first five rows are shown below.

| | Postal Code | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Summary | Venue Category | Distance |
|---|---|---|---|---|---|---|---|---|
| 0 | M1B | Malvern, Rouge | 43.806686 | -79.194353 | Images Salon & Spa | This spot is popular | Spa | 595 |
| 1 | M1B | Malvern, Rouge | 43.806686 | -79.194353 | Harvey's | This spot is popular | Restaurant | 807 |
| 2 | M1B | Malvern, Rouge | 43.806686 | -79.194353 | Staples Morningside | This spot is popular | Paper / Office Supplies Store | 735 |
| 3 | M1B | Malvern, Rouge | 43.806686 | -79.194353 | Wendy's | This spot is popular | Fast Food Restaurant | 600 |
| 4 | M1B | Malvern, Rouge | 43.806686 | -79.194353 | Wendy's | This spot is popular | Fast Food Restaurant | 387 |

5. Cluster and segment the groceries in Toronto. And use folium to visualize the distribution of these bookstores in Toronto.

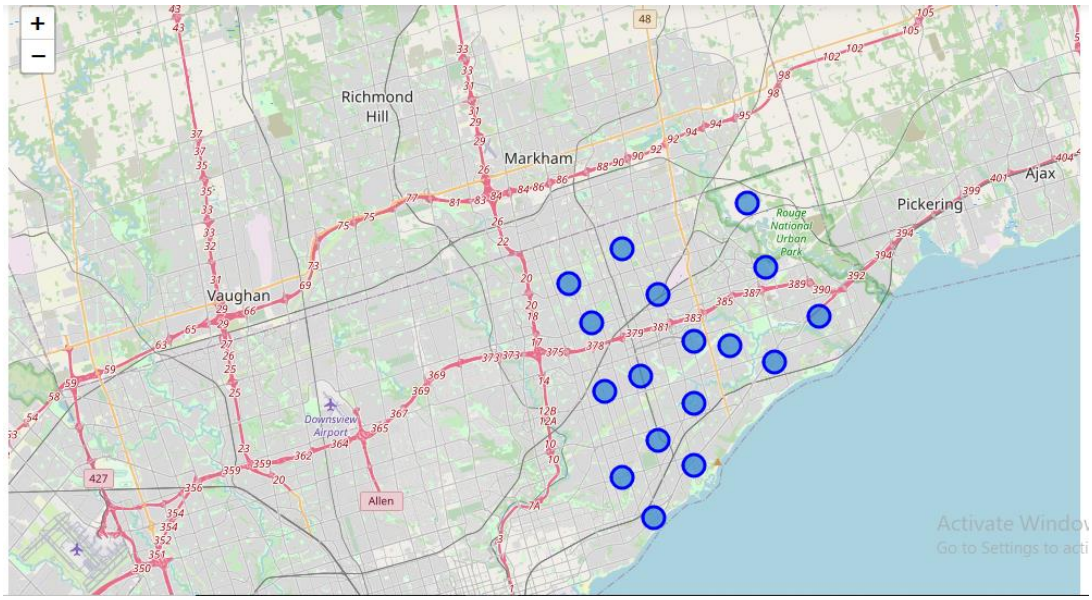
## Conclusions

A map of Toronto (mainly focus on the city center) is shown below. The city is the largest city in Canada. To open a new bookstore, location choosing is significant for success. Based on the analysis above, we try to give recommendations for a bookstore owner.

We can extract meaningful information from the dataset explored. The dataframe was sorted by the number of groceries in a neighborhood, as shown below. Here I try to find optimal neighborhoods to start up a groceries. Since there are lots of groceries in Toronto, I would not recommend neighborhoods that are already crowded with groceries. I find the neighborhoods with no groceries, which are:

| Neighborhood | Bakery | Breakfast Spot | Diner | Fish Market | Food & Drink Shop | Grocery Store | Noodle House | Pizza Place | Sandwich Place | Total Restaurants | Total Joints |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Agincourt | 2 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 19 | 0 |
| Birch Cliff, Cliffside West | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| Cedarbrae | 3 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 8 | 3 |
| Clarks Corners, Tam O'Shanter, Sullivan | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 2 | 13 | 1 |
| Cliffside, Cliffcrest, Scarborough Village West | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 |
| Dorset Park, Wexford Heights, Scarborough Town Centre | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 14 | 2 |
| Golden Mile, Clairlea, Oakridge | 2 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 0 |
| Guildwood, Morningside, West Hill | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 3 | 1 | 4 | 1 |
| Kennedy Park, Ionview, East Birchmount Park | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 2 | 1 | 6 | 1 |
| Malvern, Rouge | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 5 | 0 |
| Milliken, Agincourt North, Steeles East, L'Amoreaux East | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 2 | 0 | 12 | 2 |
| Rouge Hill, Port Union, Highland Creek | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| Scarborough Village | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 3 | 0 |

2. Showing in the map above, we get the location of the neighborhoods without a groceries. These neighborhoods will have no competition for a groceries. Other than that, this map also gives location information, for example, whether the location is close to the city center.




## Discussion

This report gives a recommendation of neighborhood and location to those who plan to open a groceries. In real world, there must be more factors to consider, such as the cost of the location. What's more, the analysis of this report cannot solve the problem of how many customers will visit every day. With the data analysis above, the report will be constructive to open a groceries in Toronto.

## Reference

[1] Toronto -Wikipedia, https://en.wikipedia.org/wiki/Toronto

[2] List of postal codes of Canada: M -Wikipedia,

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