

The Battle of Neighborhoods

Toronto Coffee Shop

*A Brand-New Coffee Shop for friends, family, business meetings and more
Toronto, CA*

INTRODUCTION / BUSINESS PROBLEM

With the speed growth of startups in the city of Toronto and the developing casual work cultures, it'll be fantastic to be able to host a meeting over coffee. Coffee shops have become a great deal of business as well, within the food industry today and there are numerous coffee shops within the city that have been set up in the recent years. However, as an entrepreneur, I'd like to find out the best possible locality to open yet another coffee shop that will withstand the competition and sustain longer in business that will provide high profitability. This project will provide all the details in establishing one.

DATA

Below is a description of the data and how it'll be used in the Methodology to solve the problem.

1. Data Scraping

Reading a Wiki Page and scraping it to retrieve all the zip codes of CAN that are in a table format
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

2. Data Cleaning and Processing

We further clean the data to remove any 'Not Assigned' or unavailable data to ensure no redundancy and then group neighborhoods together

	Postal Code	Borough	Neighborhood
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

3. Geospatial Data

Download and read the geospatial data coordinates of each postal code: 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

4. Data Merge

Merge Neighborhood data with geospatial coordinates.

	Postal Code	Borough	Neighborhood	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

Once all the cleaned and processed data is stored in a data frame for further processing, a frequency analysis for Coffee Shops in Toronto will be done utilizing this data and come up with the best choices of neighborhood for the customer/investor, which will be explained in the Methodology section below.

METHODOLOGY

Once the data is gathered, cleansed and stored in the data frame, the following step-by-step procedure is followed throughout this analysis.

1. Draw a map and show the neighborhoods
 - a) Create a map of Toronto with all neighborhoods superimposed on top
 - b) Identify neighborhoods that contain the word Toronto
2. Explore the neighborhoods
 - a) Explore all the neighborhoods in the city of Toronto, using the latitude & longitude data
 - b) Using Foursquare API, obtain a list of all the venues available in Toronto
 - c) Check how many venues and unique categories were returned for each neighborhood
 - d) Filter the venues' details for all possible 'Coffee' shops
 - e) Create a map between all the neighborhoods and all possible types of coffee shops
 - f) Group the same map by neighborhood and by taking the mean of the frequency of occurrence of each category
 - g) Display each neighborhood along with the top 5 most common venues
 - h) Identify the top 10 venues for each neighborhood. This data will be further used for clustering
3. Cluster neighborhoods
 - a) With an assumption of 5 clusters, use K-Cluster algorithm to come up with 5 different clusters in Toronto with similar set of venues
 - b) Map all the defined clusters in the city of Toronto.
4. Examine clusters and predict
 - a) Examine each cluster and determine the discriminating venue categories that distinguish each cluster.
 - b) Identify the clusters and neighborhoods with the maximum number of coffee shops

Assumptions

- The location with the maximum number of coffee shops will have need for more coffee shops – so probable location will be arrived at based on lower frequency of coffee shops in a neighborhood.
- From the list of neighborhoods in Canada, the neighborhoods with words “Toronto” is selected as that is our area of interest.
- K-Means clustering mechanism is used to cluster the city of Toronto.
- Foursquare API is used to get details on all possible venues. User must have an account with Foursquare to execute the API.

RESULTS

Results below are based on the steps defined in the Data and Methodology sections above.

Neighborhood data cleansed, processed and merged with geospatial coordinates:

There are 103 neighborhoods as per the information available, out of which 4 belong to Toronto. The following image shows this information.

	Postal Code	Borough	Neighborhood	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
5	M1J	Scarborough	Scarborough Village	43.744734	-79.239476
6	M1K	Scarborough	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029
7	M1L	Scarborough	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577
8	M1M	Scarborough	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476
9	M1N	Scarborough	Birch Cliff, Cliffside West	43.692657	-79.264848
10	M1P	Scarborough	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304
11	M1R	Scarborough	Maryvale, Wexford	43.750072	-79.295849
12	M1S	Scarborough	Agincourt	43.794200	-79.262029
13	M1T	Scarborough	Clarks Corners, Sullivan, Tam O'Shanter	43.781638	-79.304302
14	M1V	Scarborough	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577
15	M1W	Scarborough	L'Amoreaux West	43.799525	-79.318389
16	M1X	Scarborough	Upper Rouge	43.836125	-79.205636
17	M2H	North York	Hillcrest Village	43.803762	-79.363452
18	M2J	North York	Fairview, Henry Farm, Oriole	43.778517	-79.346556
19	M2K	North York	Bayview Village	43.786947	-79.385975
20	M2L	North York	Silver Hills, York Mills	43.757490	-79.374714
...					
...					
...					
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
101	M9V	Etobicoke	Albion Gardens, Beaumont Heights, Humburgate, ...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

Explore the neighborhoods

- Using Foursquare API, Total 1719 venues have been identified in Toronto (all 4 boroughs, out of which there are 237 unique categories of the venues.
- Once the venues are filtered for all possible 'coffee', we found there are 145 coffee shops available that will make it easier for us to invest into one.
- Based on the mean frequency calculating the following by each neighborhood for a postal code along with the top 5 most common venues

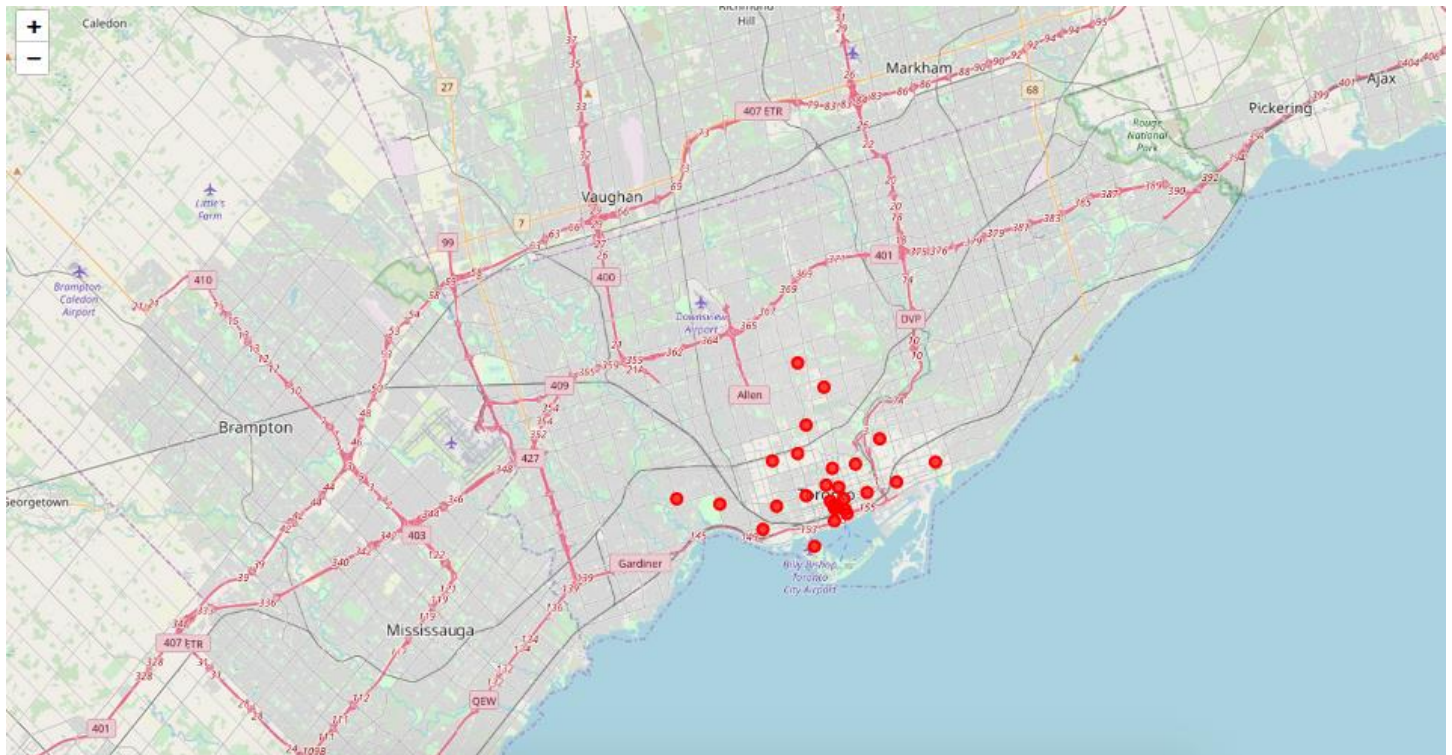
```
----Adelaide, King, Richmond----
    venue freq
0 Coffee Shop 1.0
----Berczy Park----
    venue freq
0 Coffee Shop 1.0
----Brockton, Exhibition Place, Parkdale Village----
    venue freq
0 Coffee Shop 1.0
----CN Tower, Bathurst Quay, Island airport, Harbourfront West, King and Spadina, Railway Lands, South Niagara----
    venue freq
0 Coffee Shop 1.0
----Cabbagetown, St. James Town----
    venue freq
0 Coffee Shop 1.0
----Central Bay Street----
    venue freq
0 Coffee Shop 1.0
----Chinatown, Grange Park, Kensington Market----
    venue freq
0 Coffee Shop 1.0
----Christie----
    venue freq
0 Coffee Shop 1.0
----Church and Wellesley----
    venue freq
0 Coffee Shop 1.0
----Commerce Court, Victoria Hotel----
    venue freq
0 Coffee Shop 1.0
----Davisville----
    venue freq
0 Coffee Shop 1.0
----Deer Park, Forest Hill SE, Rathnelly, South Hill, Summerhill West----
    venue freq
0 Coffee Shop 1.0
----Design Exchange, Toronto Dominion Centre----
    venue freq
0 Coffee Shop 1.0
----First Canadian Place, Underground city----
    venue freq
0 Coffee Shop 1.0
----Harbourfront East, Toronto Islands, Union Station----
    venue freq
0 Coffee Shop 1.0
----Harbourfront, Regent Park----
    venue freq
0 Coffee Shop 1.0
----Little Portugal, Trinity----
    venue freq
0 Coffee Shop 1.0
```


- The following neighborhoods have at least 1 Coffee Shop in the Top 10 venues –

0	Adelaide, King, Richmond
1	Harbourfront East, Toronto Islands, Union Station
2	The Beaches West, India Bazaar
3	The Annex, North Midtown, Yorkville
4	Studio District
5	Stn A PO Boxes 25 The Esplanade
6	St. James Town
7	Ryerson, Garden District
8	Runnymede, Swansea
9	Parkdale, Roncesvalles
10	North Toronto West
11	Little Portugal, Trinity
12	Harbourfront, Regent Park
13	First Canadian Place, Underground city
14	Berczy Park
15	Design Exchange, Toronto Dominion Centre
16	Deer Park, Forest Hill SE, Rathnelly, South Hi...
17	Davisville
18	Commerce Court, Victoria Hotel
19	Church and Wellesley
20	Christie
21	Chinatown, Grange Park, Kensington Market
22	Central Bay Street
23	Cabbagetown, St. James Town
24	CN Tower, Bathurst Quay, Island airport, Harbo...
25	Brockton, Exhibition Place, Parkdale Village
26	The Danforth West, Riverdale

The clustering algorithm resulted the following 5 clusters using K-Cluster algorithm

Once all the defined clusters are mapped on the city of Toronto:



Examine clusters and predict

- Defined clusters

Cluster 1

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 0, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

[illegible]

Cluster 2

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 1, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

	Borough	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
--	---------	----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------------------------

Cluster 3

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 2, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

	Borough	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
--	---------	----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------------------------

Cluster 4

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 3, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

	Borough	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
--	---------	----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------------------------

Cluster 5

```
Toronto_merged.loc[Toronto_merged['Cluster Labels'] == 4, Toronto_merged.columns[[1] + list(range(5, Toronto_merged.shape[1]))]]
```

	Borough	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
--	---------	----------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------------------------

- Identify the clusters and neighborhoods with Maximum number of Coffee shops

```
---- Cluster 1----
Out of 27 Neighborhoods in the Cluster 27 Neighborhoods have Coffee Shop
Selected Boroughs with No. of Neighborhoods are:
  Downtown Toronto      16
  West Toronto          4
  Central Toronto       4
  East Toronto          3

---- Cluster 2----
Out of 0 Neighborhoods in the Cluster 0 Neighborhoods have Coffee Shop

---- Cluster 3----
Out of 0 Neighborhoods in the Cluster 0 Neighborhoods have Coffee Shop

---- Cluster 4----
Out of 0 Neighborhoods in the Cluster 0 Neighborhoods have Coffee Shop

---- Cluster 5----
Out of 0 Neighborhoods in the Cluster 0 Neighborhoods have Coffee Shop
```

DISCUSSION

Based on our assumptions of the cluster with maximum number of coffee shops would have the best possibility of having a successful coffee shop, Cluster 1 is THE ONLY possible cluster to setup a coffee shop. From this data it is also clear that Downtown Toronto would be the best neighborhood to setup a coffee shop. While exploring the neighborhoods, the Top 5 neighborhoods identified for setting up a coffee shop are:

- a) Adelaide, King, Richmond
- b) Harbourfront East, Toronto Islands, Union Station
- c) The Beaches West, India Bazaar
- d) The Annex, North Midtown, Yorkville
- e) Studio District

CONCLUSION

Clustering as an algorithm gave a good split among the clusters and was able to identify uniqueness well. With the current data used, this looks like one of the best possible solution to setup a coffee shop in Cluster 1. If a borough is to be chosen, the best bet would be on Downtown, Toronto and possibly one of the top five group of neighborhoods as explained in the Discussion section.

Please note that there might be a different assumption or calculation logic used to identify the right neighborhoods which might significantly change the result of this model.