Applied Data Science Capstone

Determining the Best Location for a Toronto Coffee Shop

1. Introduction

1.1 Background

Coffee shops are a very common venue in many cities around the world. Many people drink coffee almost every morning, making these shops a great place to stop by on the way to work. However, because there are so many choices for where to purchase coffee, it's important for an entrepreneur to do the research and determine the ideal location to open up a coffee shop of their own. This way, he or she can avoid the risk of losing too many potential customers to an already-established competitor, or even not having enough potential customers in the area to begin with.

1.2 Business Problem

For an entrepreneur looking to open up a coffee shop in Toronto, is important for them to determine where the best location to open up their shop would be, in order to get a good amount of business. In order to determine where the best location to open up their coffee shop would be, there are multiple factors that they should consider. For instance, they should look at what areas currently have coffee shops, and how many. It could also be helpful to look at where office buildings are located – if there is an area that has many office buildings, but no nearby coffee shop, opening up a coffee shop in this area could provide a good business advantage. This is because coffee is a very common item for people to grab on their way into the office, or even throughout the workday.

1.3 Interest

This project would be of great interest to any entrepreneur looking to open up a coffee shop in Toronto. The same methods could also be used for entrepreneurs looking to open up businesses of any variety, by making just a few necessary tweaks to fit the category.

2. Data

2.1 Data Sources

Given the business problem stated above, the factors that would determine where an ideal place to open up a coffee shop would be the number/location of existing coffee shops in

Toronto, as well as the number/location of existing office buildings in Toronto. Therefore, the following data sources will be used for this project:

- Postal code data will be sourced from the following Wikipedia page: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Geographical coordinates will be sourced from the following .csv link: http://cocl.us/Geospatial_data
- Coffee shop and office building data will be sourced from Foursquare API

2.2 Data Cleaning

The data obtained from the multiple sources listed above was combined into a single table. It was necessary to clean the data as well, in order to get an accurate analysis.

First, the neighborhoods were grouped together by postal code, and any rows lacking a neighborhood entry used the borough in its place. This ensured that the project would have a reasonable amount of locations to work with, and the location data obtained for the postal codes ensured that it would be possible to plot their locations on a map.

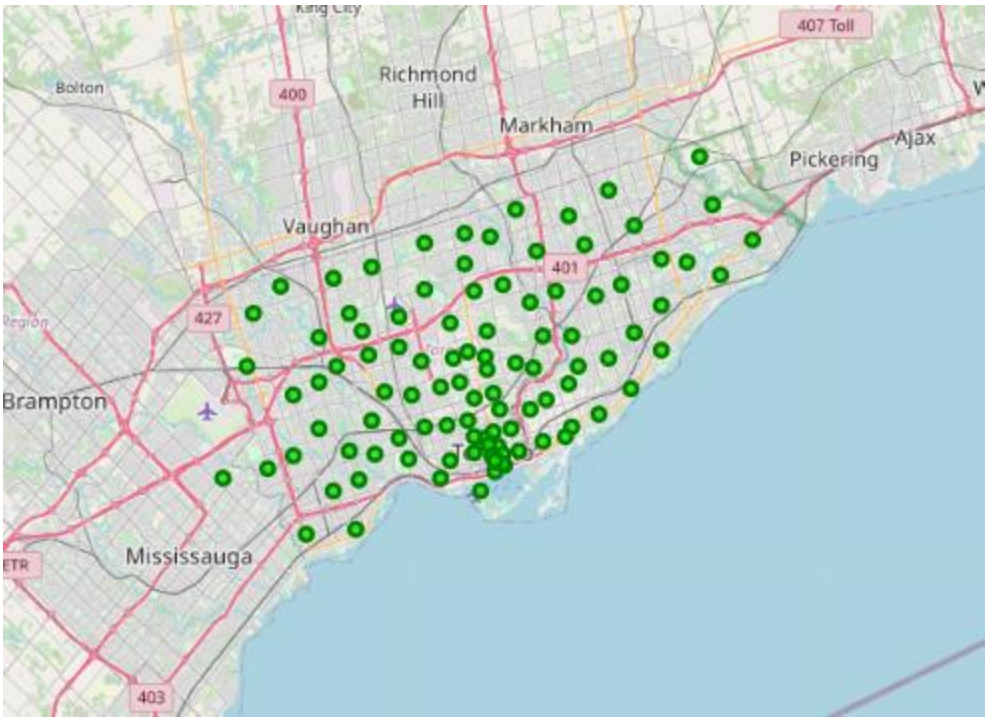
Next, the venues that appeared upon searching for “coffee” needed to be filtered down to only include those with the category “coffee shop.” This is because multiple other businesses, such as tech companies, might have shown up in the search due to having “coffee” in their business name. The same was done for the office buildings, in which the results from searching for “office” were filtered down to only those in the “office” category.

The exact venues were not important in this project, only the amount of them in each area. Therefore, the tables were altered so that they displayed the number of coffee shops and/or office buildings, rather than a full list of them. Any locations that did not contain any coffee shops or office buildings were irrelevant to this project, and were dropped from the table. However, any locations that had one or the other were included, and instances of “NaN” were replaced with “0,” accordingly.

3. Methodology

3.1 Postal Code Data and Geographical Coordinates

First, it was necessary to obtain the postal code data and geographical coordinates. The postal coded data was taken from the Wikipedia link and the geographical coordinates were taken from the .csv file, both mentioned previously. The data was cleaned, and a simple map showing each location was created.



3.2 Obtaining Data for Existing Coffee Shops and Office Buildings

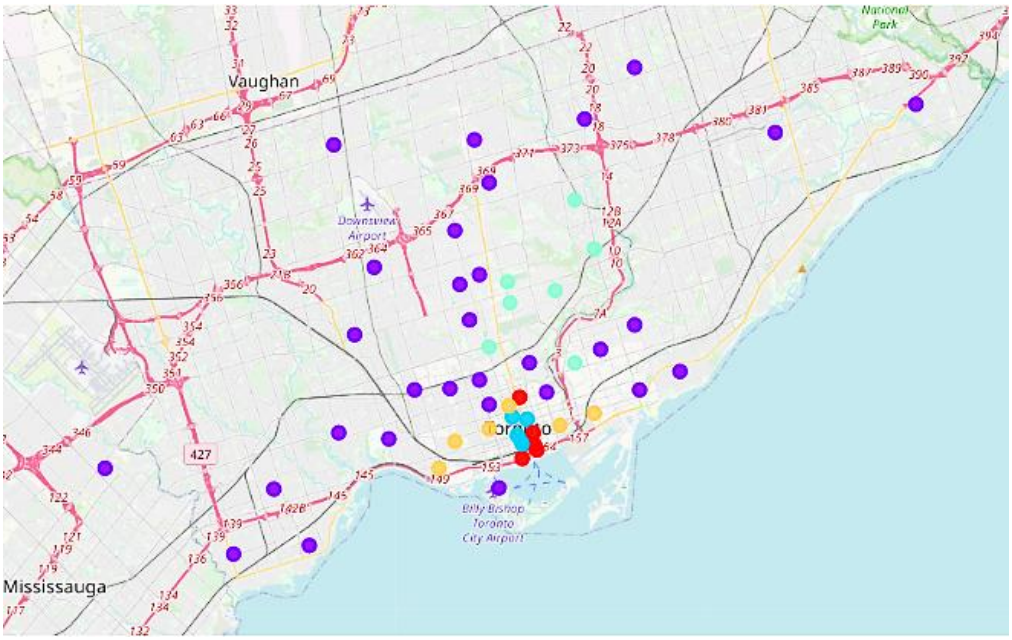
The data for existing coffee shops and office buildings was obtained using the Foursquare API. For each location, a simple search was run for “coffee,” with the venues filtered by category “coffee shop” to ensure that any other business types were not included. Likewise, a simple search was run for “office,” with the venues filtered by category “office” to ensure that any other business types were not included. The important data in this case was the number of coffee shops and/or offices in each location, not the specific venues themselves. Therefore, the dataframes were altered to include only the number of existing coffee shops and offices in each location. These dataframes were then combined into a single table, a portion of which is shown in the following figure:

	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
0	Alderwood, Long Branch	1.0	0.0	M8W	Etobicoke	43.602414	-79.543484
1	Berczy Park	8.0	22.0	M5E	Downtown Toronto	43.644771	-79.373306
2	Brockton, Parkdale Village, Exhibition Place	5.0	4.0	M6K	West Toronto	43.636847	-79.428191
3	CN Tower, King and Spadina, Railway Lands, Har...	1.0	1.0	M5V	Downtown Toronto	43.628947	-79.394420
4	Cedarbrae	1.0	1.0	M1H	Scarborough	43.773136	-79.239476
5	Central Bay Street	22.0	15.0	M5G	Downtown Toronto	43.657952	-79.387383
6	Church and Wellesley	8.0	21.0	M4Y	Downtown Toronto	43.665860	-79.383160
7	Commerce Court, Victoria Hotel	26.0	20.0	M5L	Downtown Toronto	43.648198	-79.379817
8	Davisville	2.0	6.0	M4S	Central Toronto	43.704324	-79.388790
9	Del Ray, Mount Dennis, Keelsdale and Silverthorn	1.0	0.0	M6M	York	43.691116	-79.476013
10	Dufferin, Dovercourt Village	2.0	3.0	M6H	West Toronto	43.669005	-79.442259
11	East Toronto, Broadview North (Old East York)	1.0	0.0	M4J	East York	43.685347	-79.338106
12	Fairview, Henry Farm, Oriole	3.0	2.0	M2J	North York	43.778517	-79.346556
13	First Canadian Place, Underground city	27.0	23.0	M5X	Downtown Toronto	43.648429	-79.382280
14	Garden District, Ryerson	25.0	16.0	M5B	Downtown Toronto	43.657162	-79.378937
15	Harbourfront East, Union Station, Toronto Islands	11.0	16.0	M5J	Downtown Toronto	43.640816	-79.381752
16	India Bazaar, The Beaches West	1.0	1.0	M4L	East Toronto	43.668999	-79.315572

Any locations that did not contain either coffee shops or offices were dropped, as they were not applicable to this project. However, any locations that contained only one or the other were still included. The number of coffee shops in each location ranged from 0 – 27, while the number of offices in each location ranged from 0 – 23.

3.3 K-Means Clustering with Coffee Shop and Office Building Data

K-means clustering was used to group the locations into five clusters, based on only the number of coffee shops and office buildings. The different clusters were displayed on a map of Toronto, as shown in the following figure:



Each of these clusters were then examined to see how they were differentiated from one another, based on the number of coffee shops and office buildings.

First Cluster (Label 0):

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
1	0	Berczy Park	8.0	22.0	M5E	Downtown Toronto	43.644771	-79.373306
6	0	Church and Wellesley	8.0	21.0	M4Y	Downtown Toronto	43.665860	-79.383160
15	0	Harbourfront East, Union Station, Toronto Islands	11.0	16.0	M5J	Downtown Toronto	43.640816	-79.381752
27	0	St. James Town	17.0	23.0	M5C	Downtown Toronto	43.651494	-79.375418
30	0	Stn A PO Boxes	17.0	20.0	M5W	Downtown Toronto	43.646435	-79.374846

Second Cluster (Label 1):

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
0	1	Alderwood, Long Branch	1.0	0.0	M8W	Etobicoke	43.602414	-79.543484
3	1	CN Tower, King and Spadina, Railway Lands, Har...	1.0	1.0	M5V	Downtown Toronto	43.628947	-79.394420
4	1	Cedarbrae	1.0	1.0	M1H	Scarborough	43.773136	-79.239476
9	1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	1.0	0.0	M6M	York	43.691116	-79.476013
10	1	Dufferin, Dovercourt Village	2.0	3.0	M6H	West Toronto	43.669005	-79.442259
11	1	East Toronto, Broadview North (Old East York)	1.0	0.0	M4J	East York	43.685347	-79.338106
12	1	Fairview, Henry Farm, Oriole	3.0	2.0	M2J	North York	43.778517	-79.346556
16	1	India Bazaar, The Beaches West	1.0	1.0	M4L	East Toronto	43.668999	-79.315572
19	1	Mimico NW, The Queensway West, South of Bloor,...	1.0	0.0	M8Z	Etobicoke	43.628841	-79.520999
20	1	New Toronto, Mimico South, Humber Bay Shores	2.0	0.0	M8V	Etobicoke	43.605647	-79.501321
21	1	Parkdale, Roncesvalles	2.0	0.0	M6R	West Toronto	43.648960	-79.456325
25	1	Rosedale	1.0	2.0	M4W	Downtown Toronto	43.679563	-79.377529
26	1	Runnymede, Swansea	2.0	0.0	M6S	West Toronto	43.651571	-79.484450
28	1	St. James Town, Cabbagetown	1.0	2.0	M4X	Downtown Toronto	43.667967	-79.367675
29	1	Steeles West, L'Amoreaux West	2.0	0.0	M1W	Scarborough	43.799525	-79.318389
33	1	The Annex, North Midtown, Yorkville	2.0	1.0	M5R	Central Toronto	43.672710	-79.405678
34	1	The Beaches	1.0	0.0	M4E	East Toronto	43.676357	-79.293031
37	1	University of Toronto, Harbord	2.0	1.0	M5S	Downtown Toronto	43.662696	-79.400049
38	1	Willowdale, Willowdale East	3.0	3.0	M2N	North York	43.770120	-79.408493
39	1	Bedford Park, Lawrence Manor East	0.0	1.0	M5M	North York	43.733283	-79.419750
40	1	Canada Post Gateway Processing Centre	0.0	2.0	M7R	Mississauga	43.636966	-79.615819
41	1	Christie	0.0	3.0	M6G	Downtown Toronto	43.669542	-79.422564
45	1	Forest Hill North & West, Forest Hill Road Park	0.0	1.0	M5P	Central Toronto	43.696948	-79.411307
46	1	Lawrence Manor, Lawrence Heights	0.0	2.0	M6A	North York	43.718518	-79.464763
48	1	North Toronto West, Lawrence Park	0.0	1.0	M4R	Central Toronto	43.715383	-79.405678
49	1	Northwood Park, York University	0.0	2.0	M3J	North York	43.767980	-79.487262
50	1	Roselawn	0.0	2.0	M5N	Central Toronto	43.711695	-79.416936
51	1	Rouge Hill, Port Union, Highland Creek	0.0	1.0	M1C	Scarborough	43.784535	-79.160497
52	1	Woodbine Heights	0.0	1.0	M4C	East York	43.695344	-79.318389
53	1	York Mills West	0.0	1.0	M2P	North York	43.752758	-79.400049

Third Cluster (Label 2):

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
5	2	Central Bay Street	22.0	15.0	M5G	Downtown Toronto	43.657952	-79.387383
7	2	Commerce Court, Victoria Hotel	26.0	20.0	M5L	Downtown Toronto	43.648198	-79.379817
13	2	First Canadian Place, Underground city	27.0	23.0	M5X	Downtown Toronto	43.648429	-79.382280
14	2	Garden District, Ryerson	25.0	16.0	M5B	Downtown Toronto	43.657162	-79.378937
24	2	Richmond, Adelaide, King	27.0	20.0	M5H	Downtown Toronto	43.650571	-79.384568
36	2	Toronto Dominion Centre, Design Exchange	23.0	23.0	M5K	Downtown Toronto	43.647177	-79.381576

Fourth Cluster (Label 3):

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
8	3	Davisville	2.0	6.0	M4S	Central Toronto	43.704324	-79.388790
32	3	Summerhill West, Rathnelly, South Hill, Forest...	1.0	5.0	M4V	Central Toronto	43.686412	-79.400049
35	3	The Danforth West, Riverdale	3.0	5.0	M4K	East Toronto	43.679557	-79.352188
42	3	Davisville North	0.0	6.0	M4P	Central Toronto	43.712751	-79.390197
43	3	Don Mills	0.0	8.0	M3B	North York	43.745906	-79.352188
44	3	Don Mills	0.0	8.0	M3C	North York	43.725900	-79.340923
47	3	Leaside	0.0	4.0	M4G	East York	43.709060	-79.363452

Fifth Cluster (Label 4):

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
2	4	Brockton, Parkdale Village, Exhibition Place	5.0	4.0	M6K	West Toronto	43.636847	-79.428191
17	4	Kensington Market, Chinatown, Grange Park	7.0	5.0	M5T	Downtown Toronto	43.653206	-79.400049
18	4	Little Portugal, Trinity	6.0	1.0	M6J	West Toronto	43.647927	-79.419750
22	4	Queen's Park, Ontario Provincial Government	10.0	5.0	M7A	Downtown Toronto	43.662301	-79.389494
23	4	Regent Park, Harbourfront	6.0	6.0	M5A	Downtown Toronto	43.654260	-79.360636
31	4	Studio District	4.0	3.0	M4M	East Toronto	43.659526	-79.340923

4. Results

As can be seen in the tables above, it was very possible to tell the clusters apart based on the number of coffee shops and office buildings. The results for each cluster were as follows:

Cluster 1

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
1	0	Berczy Park	8.0	22.0	M5E	Downtown Toronto	43.644771	-79.373306
6	0	Church and Wellesley	8.0	21.0	M4Y	Downtown Toronto	43.665860	-79.383160
15	0	Harbourfront East, Union Station, Toronto Islands	11.0	16.0	M5J	Downtown Toronto	43.640816	-79.381752
27	0	St. James Town	17.0	23.0	M5C	Downtown Toronto	43.651494	-79.375418
30	0	Stn A PO Boxes	17.0	20.0	M5W	Downtown Toronto	43.646435	-79.374846

Contained a high amount of office buildings, and a moderate amount of coffee shops.

Locations in this cluster had roughly 2 office buildings per coffee shop, on average.

Cluster 2

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
0	1	Alderwood, Long Branch	1.0	0.0	M8W	Etobicoke	43.602414	-79.543484
3	1	CN Tower, King and Spadina, Railway Lands, Har...	1.0	1.0	M5V	Downtown Toronto	43.628947	-79.394420
4	1	Cedarbrae	1.0	1.0	M1H	Scarborough	43.773136	-79.239476
9	1	Del Ray, Mount Dennis, Keelsdale and Silverthorn	1.0	0.0	M6M	York	43.691116	-79.476013
10	1	Dufferin, Dovercourt Village	2.0	3.0	M6H	West Toronto	43.669005	-79.442259
11	1	East Toronto, Broadview North (Old East York)	1.0	0.0	M4J	East York	43.685347	-79.338106
12	1	Fairview, Henry Farm, Oriole	3.0	2.0	M2J	North York	43.778517	-79.346556
16	1	India Bazaar, The Beaches West	1.0	1.0	M4L	East Toronto	43.668999	-79.315572
19	1	Mimico NW, The Queensway West, South of Bloor,...	1.0	0.0	M8Z	Etobicoke	43.628841	-79.520999
20	1	New Toronto, Mimico South, Humber Bay Shores	2.0	0.0	M8V	Etobicoke	43.605647	-79.501321
21	1	Parkdale, Roncesvalles	2.0	0.0	M6R	West Toronto	43.648960	-79.456325
25	1	Rosedale	1.0	2.0	M4W	Downtown Toronto	43.679563	-79.377529
26	1	Runnymede, Swansea	2.0	0.0	M6S	West Toronto	43.651571	-79.484450
28	1	St. James Town, Cabbagetown	1.0	2.0	M4X	Downtown Toronto	43.667967	-79.367675
29	1	Steeles West, L'Amoreaux West	2.0	0.0	M1W	Scarborough	43.799525	-79.318389
33	1	The Annex, North Midtown, Yorkville	2.0	1.0	M5R	Central Toronto	43.672710	-79.405678
34	1	The Beaches	1.0	0.0	M4E	East Toronto	43.676357	-79.293031
37	1	University of Toronto, Harbord	2.0	1.0	M5S	Downtown Toronto	43.662696	-79.400049
38	1	Willowdale, Willowdale East	3.0	3.0	M2N	North York	43.770120	-79.408493
39	1	Bedford Park, Lawrence Manor East	0.0	1.0	M5M	North York	43.733283	-79.419750
40	1	Canada Post Gateway Processing Centre	0.0	2.0	M7R	Mississauga	43.636966	-79.615819
41	1	Christie	0.0	3.0	M6G	Downtown Toronto	43.669542	-79.422564
45	1	Forest Hill North & West, Forest Hill Road Park	0.0	1.0	M5P	Central Toronto	43.696948	-79.411307
46	1	Lawrence Manor, Lawrence Heights	0.0	2.0	M6A	North York	43.718518	-79.464763
48	1	North Toronto West, Lawrence Park	0.0	1.0	M4R	Central Toronto	43.715383	-79.405678
49	1	Northwood Park, York University	0.0	2.0	M3J	North York	43.767980	-79.487262
50	1	Roselawn	0.0	2.0	M5N	Central Toronto	43.711695	-79.416936
51	1	Rouge Hill, Port Union, Highland Creek	0.0	1.0	M1C	Scarborough	43.784535	-79.160497
52	1	Woodbine Heights	0.0	1.0	M4C	East York	43.695344	-79.318389
53	1	York Mills West	0.0	1.0	M2P	North York	43.752758	-79.400049

Contained a low amount of both coffee shops and office buildings.

Locations in this cluster had roughly equal amounts of office buildings and coffee shops, on average.

Cluster 3

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
5	2	Central Bay Street	22.0	15.0	M5G	Downtown Toronto	43.657952	-79.387383
7	2	Commerce Court, Victoria Hotel	26.0	20.0	M5L	Downtown Toronto	43.648198	-79.379817
13	2	First Canadian Place, Underground city	27.0	23.0	M5X	Downtown Toronto	43.648429	-79.382280
14	2	Garden District, Ryerson	25.0	16.0	M5B	Downtown Toronto	43.657162	-79.378937
24	2	Richmond, Adelaide, King	27.0	20.0	M5H	Downtown Toronto	43.650571	-79.384568
36	2	Toronto Dominion Centre, Design Exchange	23.0	23.0	M5K	Downtown Toronto	43.647177	-79.381576

Contained a high amount of both coffee shops and office buildings.

Locations in this cluster had roughly equal amounts of office buildings and coffee shops, on average.

Cluster 4

	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
8	3	Davisville	2.0	6.0	M4S	Central Toronto	43.704324	-79.388790
32	3	Summerhill West, Rathnelly, South Hill, Forest...	1.0	5.0	M4V	Central Toronto	43.686412	-79.400049
35	3	The Danforth West, Riverdale	3.0	5.0	M4K	East Toronto	43.679557	-79.352188
42	3	Davisville North	0.0	6.0	M4P	Central Toronto	43.712751	-79.390197
43	3	Don Mills	0.0	8.0	M3B	North York	43.745906	-79.352188
44	3	Don Mills	0.0	8.0	M3C	North York	43.725900	-79.340923
47	3	Leaside	0.0	4.0	M4G	East York	43.709060	-79.363452

Contained a moderate amount of office buildings, and a low amount of coffee shops.

Locations in this cluster had roughly 7 office buildings per coffee shop, on average.

Cluster 5

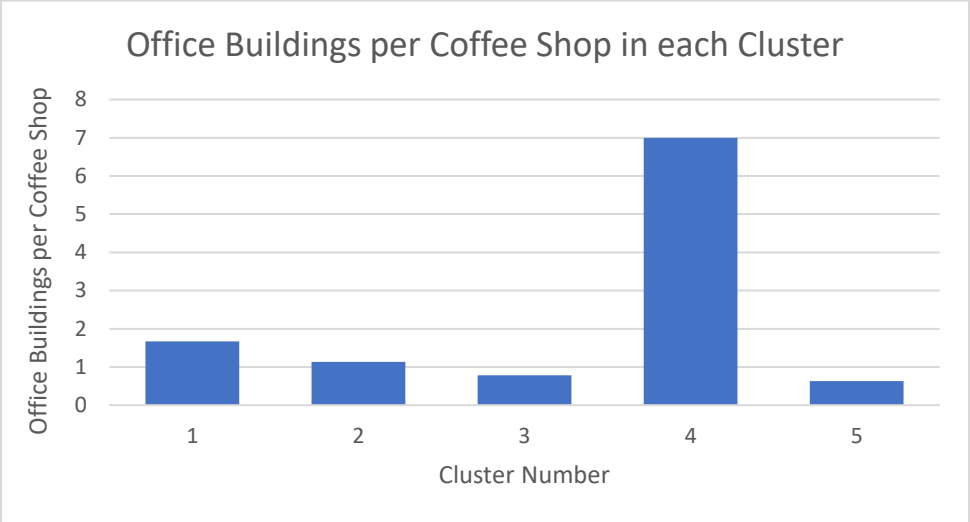
	Cluster Labels	Neighborhood	Coffee Shops	Office Buildings	Postal Code	Borough	Latitude	Longitude
2	4	Brockton, Parkdale Village, Exhibition Place	5.0	4.0	M6K	West Toronto	43.636847	-79.428191
17	4	Kensington Market, Chinatown, Grange Park	7.0	5.0	M5T	Downtown Toronto	43.653206	-79.400049
18	4	Little Portugal, Trinity	6.0	1.0	M6J	West Toronto	43.647927	-79.419750
22	4	Queen's Park, Ontario Provincial Government	10.0	5.0	M7A	Downtown Toronto	43.662301	-79.389494
23	4	Regent Park, Harbourfront	6.0	6.0	M5A	Downtown Toronto	43.654260	-79.360636
31	4	Studio District	4.0	3.0	M4M	East Toronto	43.659526	-79.340923

Contained a moderate amount of both coffee shops and office buildings.

Locations in this cluster had roughly 1-2 coffee shops per office building, on average.

Of the five clusters generated in this project, four of them (clusters 1, 2, 3, and 5) had a ratio of coffee shops to office buildings of roughly 1:1, 2:1, or 1:2. Only one cluster (cluster 4) had a

drastically different ratio of 1:7. These ratios were converted to the number of office buildings per coffee shop, and are shown plotted below for reference:



As seen in the above figure, clusters 1, 2, 3, and 5 had roughly 1-2 office buildings per coffee shop (on average, rounded to the nearest whole number). Cluster 4 had a drastically different number of 7 office buildings per coffee shop, on average.

5. Discussion

As seen in the results section, four of the five clusters generated in this project had a ratio of coffee shops to office buildings of roughly 1:1, 2:1, or 1:2, while only one cluster (cluster 4) had a drastically different ratio of 1:7. This means that for every coffee shop in these locations, there was an average of seven office buildings.

Overall, the data makes it clear that the locations in cluster 4 are in the most need of a new coffee shop, as the ratio of coffee shops to office buildings is by far the lowest (or, the ratio of office buildings to coffee shops is the highest). This means that if an entrepreneur were to open up a new coffee shop in any of the areas grouped into cluster 4, he or she would be much more likely to have a higher number of customers than if the same shop had opened up in clusters 1, 2, 3, or 5.

Therefore, if an entrepreneur is looking to open up a coffee shop in Toronto, the best locations to do so would be as follows:

- Central Toronto
    - Davisville
    - Summerhill West
    - Rathnelly
    - South Hill

    - Forest Hill SE
    - Deer Park
    - Davisville North
- East Toronto
    - The Danforth West
    - Riverdale
- North York
    - Don Mills
- East York
    - Leaside

While the remaining four clusters all had similar ratios, there was still a slight difference between them. Cluster 1 would be a decent second option, with the next-lowest ratio of 1 coffee shop for every 2 office buildings. However, it still falls far behind Cluster 4, making Cluster 4 by far the best choice. The worst choice would be Cluster 5, with 1-2 coffee shops per office building, making it the most saturated of the groups.

6. Conclusion

This project analyzed data from the Toronto area to determine where the best place for an entrepreneur to open up a coffee shop would be. In addition to the number of coffee shops in each area, the number of office buildings was considered, due to the idea that coffee shops located close to office buildings will gain a portion of those employees as customers, as coffee is a very common item to grab on the way into work or throughout the workday. By using K-means clustering, it was possible to determine the best locations for a new coffee shop, based on the fact the ideal cluster contained—by far—the lowest ratio of coffee shops to office buildings. Therefore, an entrepreneur opening up a new coffee shop in any of the locations grouped into this cluster would be likely to have a much higher number of potential customers.