DETERMINING THE BEST LOCATION FOR A COFFEE SHOP IN TORONTO

Applied Data Science Capstone – Final Presentation

Predicting the best location is important

- For an entrepreneur looking to open up a new coffee shop, it is important for them to determine where the best location would be in order to:
  - Avoid the risk of losing too many potential customers to an already-established competitor
  - Avoid the risk of not having enough potential customers in the area to begin with
- The same methods could also be used for entrepreneurs looking to open up a business of any variety, by making just a few necessary tweaks to fit the category.

Data Acquisition

◦ The factors that would determine where an ideal place to open up a coffee shop in Toronto would be:
  ◦ The number/location of existing coffee shops in Toronto
  ◦ The number/location of existing office buildings in Toronto
    ◦ This is based on the idea that coffee is a very common item to grab on the way into work, or throughout the workday

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

Data Cleaning

◦ The following steps were taken to clean the data obtained from these sources:
   ◦ The neighborhoods were grouped together by postal code
   ◦ Any rows lacking a neighborhood entry used the borough in its place
   ◦ The venues that appeared upon searching for "coffee" were filtered down to only include those with the category "coffee shop"
   ◦ The venues that appeared upon searching for "office" were filtered down to only include those with the category "office"
   ◦ The tables were altered to display the number of coffee shops and/or office buildings in each area, rather than a full list of them
   ◦ Any locations that contained neither coffee shops nor office buildings were dropped from the table
   ◦ Locations that had only one or the other were still included, and instances of "NaN" were replaced with "0"

Obtaining Postal Code Data and Geographical Coordinates
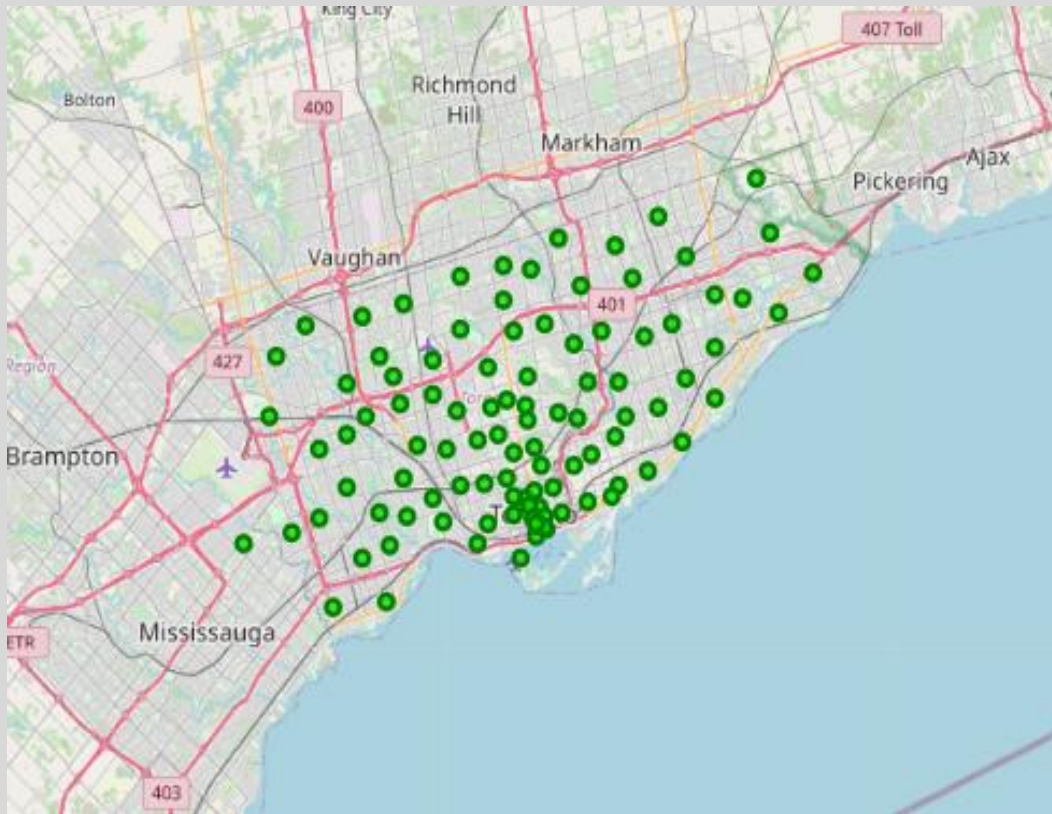
○ Postal code data was taken from the Wikipedia link mentioned previously

○ Geographical coordinates were taken from the .csv file mentioned previously

○ The data was cleaned, and a simple map showing each location was created

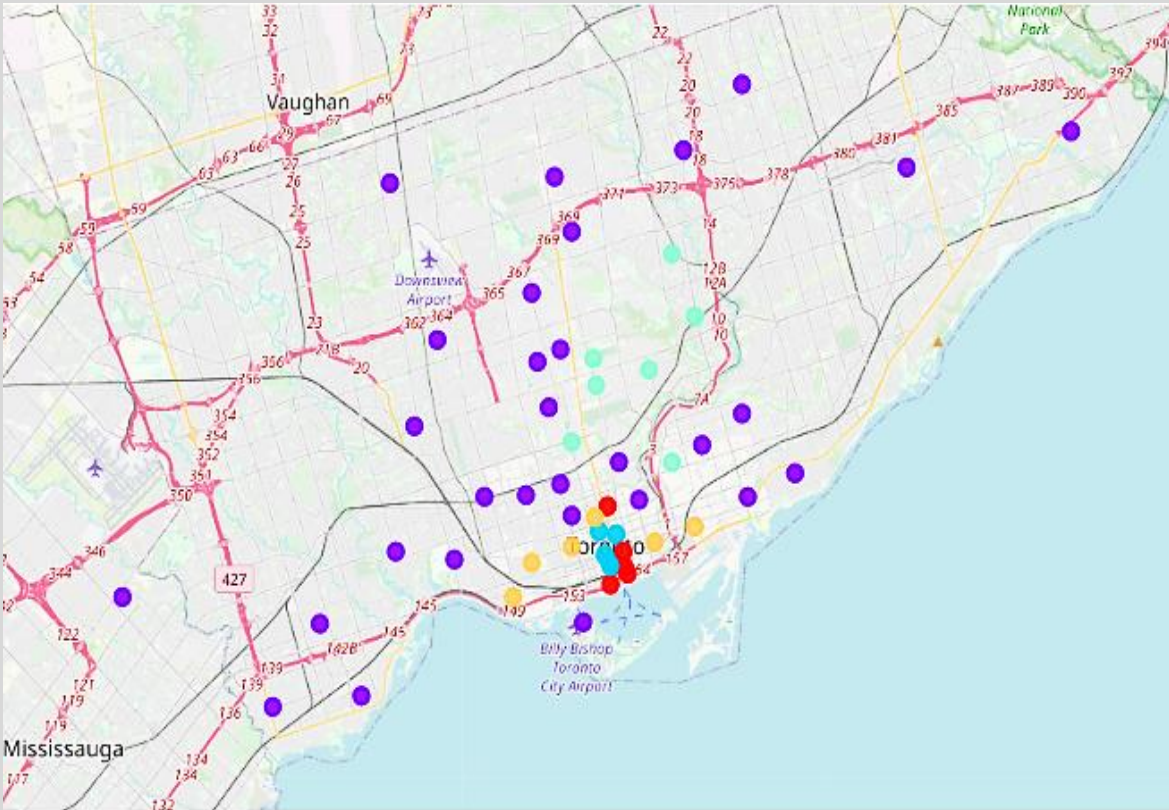# Obtaining Data for Existing Coffee Shops and Office Buildings

	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

(partial dataframe)

- Data for existing coffee shops and office buildings was obtained using the Foursquare API
  - For each location, a search was run for "coffee," with venues filtered by category "coffee shop" to ensure any other business types were not included
  - Likewise, a search was run for "office," with venues filtered by category "office"
- The dataframes were altered to include only the number of existing coffee shops and offices in each location, and combined into a single table
- Any locations that did not contain either coffee shops or offices were dropped, but those containing only one or the other were still included

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

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

Results: 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

◦ Observations:
  ◦ 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

Results: Cluster 2

(partial dataframe, full one is too large to display)

	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

○ Observations:

  ◦ 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

Results: 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

◦ Observations:
  ◦ 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

Results: 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

◦ Observations:
  ◦ 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

Results: 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

◦ Observations:
  ◦ 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

Results: Office Buildings per Coffee Shop

Office Buildings per Coffee Shop in each Cluster
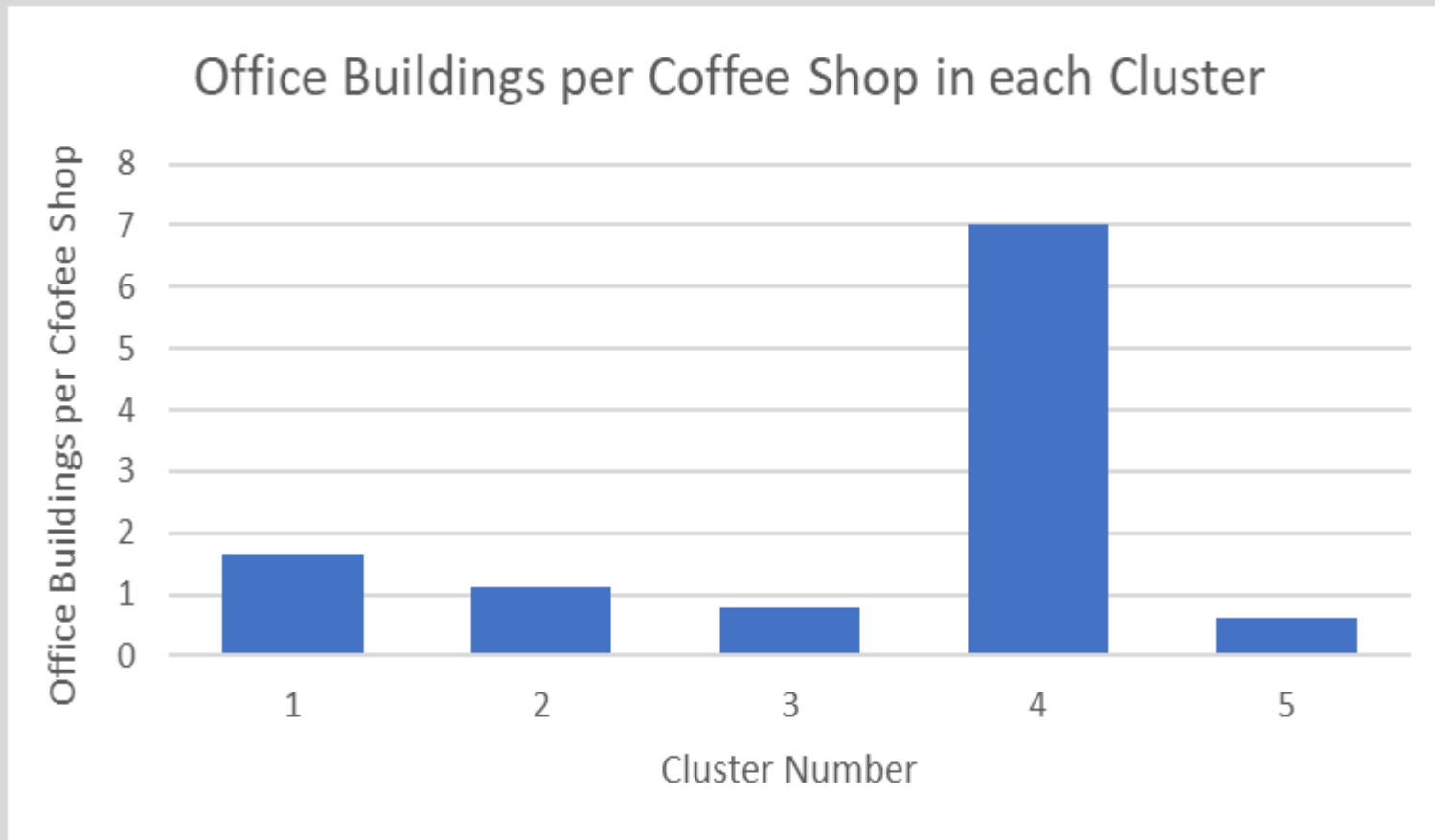
- Clusters 1, 2, 3, and 5 contained roughly 1-2 office buildings per coffee shop (on average, rounded to the nearest whole number)

- Cluster 4 contained roughly 7 office buildings per coffee shop

Discussion

◦ As seen in the results, four of the five clusters in this project had a ratio of coffee shops to office buildings of roughly 1:1 or 2:1…
  ◦ Clusters 1, 2, 3, and 5

◦ …while only one cluster had a drastically different ratio of 1:7
  ◦ Cluster 4

◦ Overall, the data makes it clear that the locations in cluster 4 are the most in need of a new coffee shop, as the ratio of coffee shops to office buildings is by far the lowest.
  ◦ This means that if an entrepreneur were to open up a new coffee shop in any of the areas grouped into cluster 4, they 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.

Recommendations

◦ 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

◦ The locations in Cluster 1 would be the second option, with the next-lowest ratio of 1 coffee shop for every 2 office buildings

◦ The locations in Cluster 5 would be the worst option, with the largest ratio of 1-2 coffee shops per office building

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

◦ By using K-means clustering, it was possible to determine the best locations for a new coffee shop, based on the fact that it contained 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 bel likely to have a much higher number of potential customers