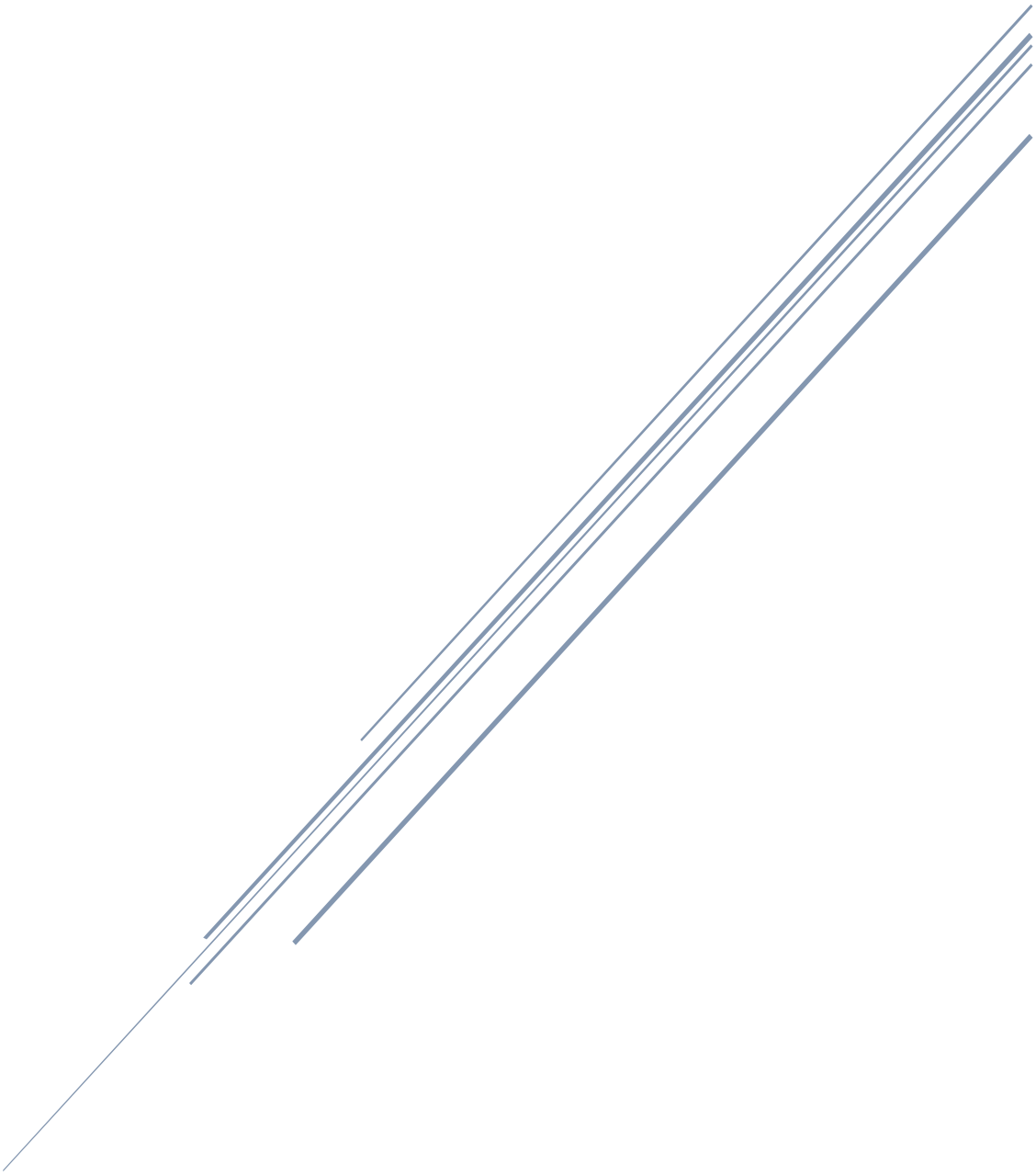
BEST LOCATION FOR OPENING A COFFEE SHOP IN TORONTO



Data Science Capstone project

Raghav Arora - 29/12/2019

1 Introduction

1.1 Background

In Toronto which is the financial capital of Canada is very famous across the world. Toronto welcome approx. 43.7 million visitors annually and 49% of their money is being spent on Food Services and Drinking Places. Due to this, there are approx. 400K people are employed. Also, there is coffee consumption of 4-5 cups of everyday. Therefore, it is important to know whether there are enough coffee shops present in all the Neighbourhood or not. Which coffee shop has the highest count?

1.2 Problem

Data that might contribute to determine the need of coffee shop. Which neighbourhood or Borough needs a coffee shop and which coffee shops has the highest count?

1.3 Interest

- Potential Buyers who can estimate the value of a house based on the surrounding venues.
- Tourist who are looking to explore the place
- Investors and Real Estate Planners
- People who are looking to settle down in the place

2 Data

2.1 Data Sources

- Scraped Toronto Borough and Neighbourhood from Wikipedia link: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

	Postcode	Borough	Neighborhood
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

- Used Foursquare explore API to fetch all the venues and venue categories in each neighbourhood i.e. https://api.foursquare.com/v2/venues/explore

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Dip 'n Sip	43.678897	-79.297745	Coffee Shop
1	The Danforth West, Riverdale	43.679557	-79.352188	Marvel Coffee Co.	43.678630	-79.347460	Coffee Shop
2	The Danforth West, Riverdale	43.679557	-79.352188	Second Cup	43.677232	-79.352898	Coffee Shop
3	The Danforth West, Riverdale	43.679557	-79.352188	Cafe Frappe	43.678126	-79.348434	Coffee Shop
4	The Danforth West, Riverdale	43.679557	-79.352188	Leonidas Chocolates Cafe	43.678118	-79.349485	Café
...	...	...	...	...	...	...	...
812	Runnymede, Swansea	43.651571	-79.484450	Koffea	43.650193	-79.481715	Coffee Shop
813	Runnymede, Swansea	43.651571	-79.484450	Second Cup	43.650137	-79.480795	Café
814	Runnymede, Swansea	43.651571	-79.484450	Say Tea	43.650205	-79.481267	Tea Room
815	Runnymede, Swansea	43.651571	-79.484450	The Coffee Bouquets	43.648785	-79.485940	Coffee Shop
816	Runnymede, Swansea	43.651571	-79.484450	Coffee Time	43.649636	-79.483837	Coffee Shop

817 rows × 7 columns

2.1 Data Cleaning

- Downloaded data from Wikipedia has to be cleaned because it has few missing values.

	Postcode	Borough	Neighborhood
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

	Postcode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Harbourfront
3	M6A	North York	Lawrence Heights
4	M6A	North York	Lawrence Manor

- Merging all the categories against same postcode.

	Postcode	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
...	...	...	...
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...
102	M9W	Etobicoke	Northwest

103 rows × 3 columns

- Merging latitude and longitude against each neighbourhood.

	Postcode	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
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724
101	M9V	Etobicoke	Albion Gardens, Beaumond Heights, Humbergate, ...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest	43.706748	-79.594054

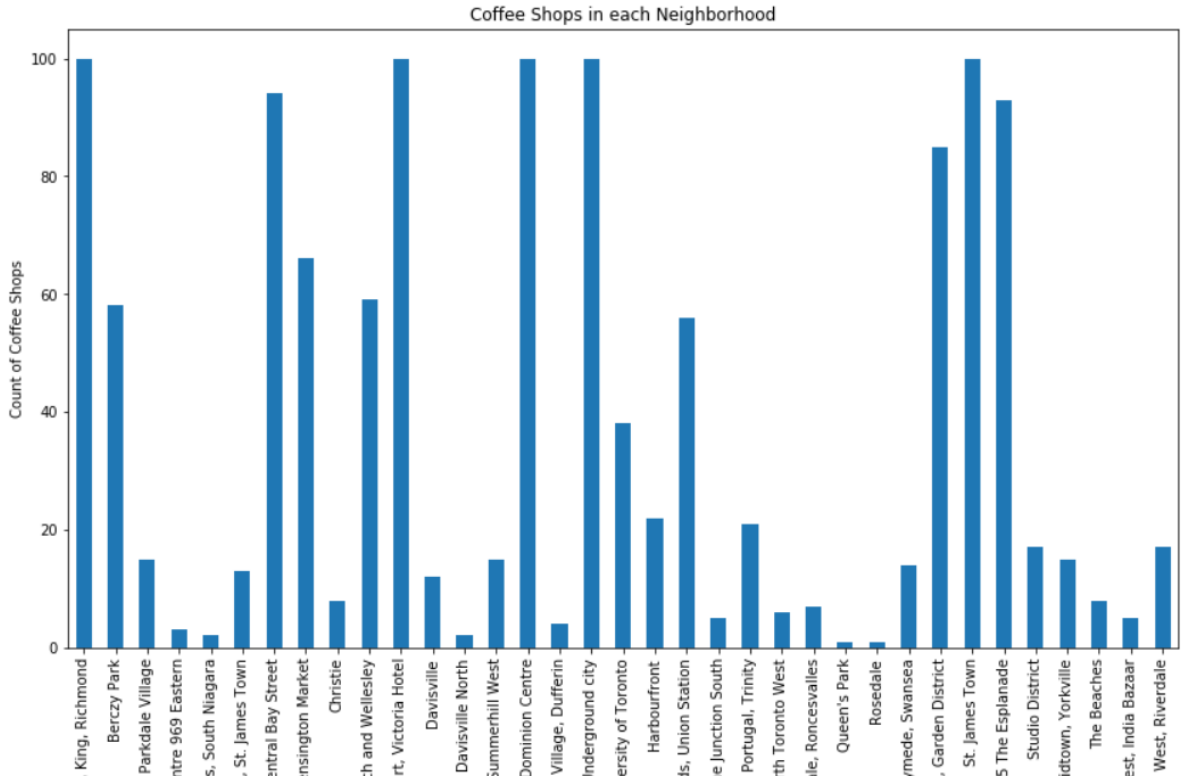
103 rows × 5 columns

- Fetching Borough which has Toronto text

	Postcode	Borough	Neighborhood	Latitude	Longitude
0	M5H	Downtown Toronto	Adelaide, King, Richmond	43.650571	-79.384568
1	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
2	M6K	West Toronto	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191
3	M7Y	East Toronto	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558
4	M5V	Downtown Toronto	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420
5	M4X	Downtown Toronto	Cabbagetown, St. James Town	43.667967	-79.367675
6	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
7	M5T	Downtown Toronto	Chinatown, Grange Park, Kensington Market	43.653206	-79.400049
8	M6G	Downtown Toronto	Christie	43.669542	-79.422564
9	M4Y	Downtown Toronto	Church and Wellesley	43.665860	-79.383160
10	M5L	Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817
11	M4S	Central Toronto	Davisville	43.704324	-79.388790
12	M4P	Central Toronto	Davisville North	43.712751	-79.390197
13	M4V	Central Toronto	Deer Park, Forest Hill SE, Rathnelly, South Hi...	43.686412	-79.400049
14	M5K	Downtown Toronto	Design Exchange, Toronto Dominion Centre	43.647177	-79.381576
15	M6H	West Toronto	Dovercourt Village, Dufferin	43.669005	-79.442259
16	M5X	Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280

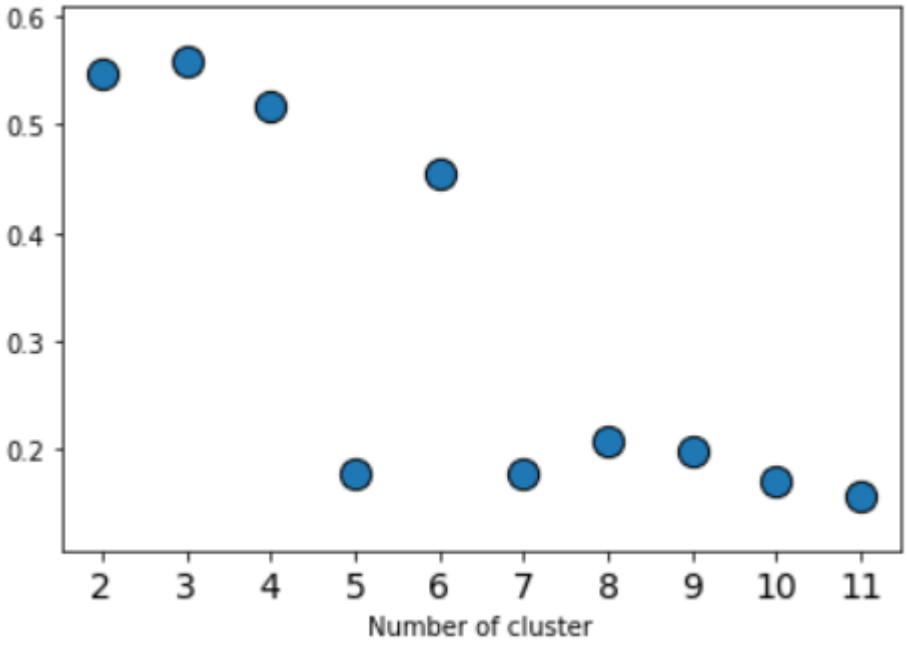
3 Methodology:

Exploratory Analysis: Exploring the dataset to analyse the count of coffee shops in each neighbourhood. Based on the chart I can say that there are quite a good no of coffee shops in few neighbourhood. There are few where I can see only 1 coffee shop.



Cluster Count:

To decide the number of clusters for KMeans algorithm. I have used Silhouette Coefficient. By plotting a graph, it was clearly stating the number of clusters to be used in this dataset.



4 Result:

The result of this project is that it produces a list of neighbourhood which is good to have more coffee shops. After running the model, we can say that Rosdale and Queen's Park are good location to have coffee shops. More insights can be depicted from the data analysed.

The following image shows the result:

	Neighborhood	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
34	Queen's Park	0	Java Joe's Village Cafe	None	None	None	None	None	None	None	None	None
8	Rosedale	2	Home Coffee Solutions	None	None	None	None	None	None	None	None	None

5 Conclusion:

This system considers factors such as existing coffee shops, neighbourhood to determine the place to open up a new coffee shop. More filters can be added in this with the help of more data like population in the neighbourhood or income of the person living in the neighbourhood. It is a powerful data driven model whose efficiency may decrease with more data but accuracy will increase.