Introduction

High yielding investment plans, multimillionaire business deals, smoothening relationship feuds or a hearty chat with a soul-mate- all of these can find a perfect ambience in a local coffee shop over a cup of fresh coffee beans roasted, grounded and brewed to perfection. Since a coffee shop is a hotspot which can caters to the need for relaxation and refreshment of folks of all ages, we decided to speculate the business proposition of opening a coffee shop in the Canadian city of Toronto.

The Covid19 pandemic has spread with an alarming rate, infecting millions and bringing economic activity to a near standstill, having witnessed the deepest global recession in decades. So in this grim scenario, an entrepreneurial venture to establish a coffee shop shall be an effort to attain monetary growth and more importantly generate employment and contributing to the local tax base, thus uplifting our community.

So, in the given project the problem statement that has been focused upon is - " Opening a coffee shop". The next question that arises out of the aforementioned statement is – " Where?" The location that has been decided upon is Toronto. This geographical location has been narrowed down after exploring and analysing its business growth opportunity, vividly comparing the same with its neighbourhood. The said analysis has been methodically approached with the help of Foursquare Location data, in addition to other datasets.

Aspiring entrepreneurs shall positively find this business proposition interesting as well as promising for several reasons, the foremost being the demand for higher quality or specialty coffee is developing and growing in the U.S, the Europe and other nations as well. In addition to this, starting a coffee shop shall be highly rewarding in terms of socialization and unleashing one's own creativity. IR shall also prove to be a profitable business due to the emerging popularity of coffee among various segments of the society and thus result in self empowerment and financial independence.

So in this project, we will explore the neighbourhood of Toronto for finding best business opportunity for opening a Coffee Shop.

A description of the data:

The data used for analysing the business opportunity of opening a coffee shop is as follows:

1. Neighbourhood data of Toronto.
2. Population data of each neighbourhood of Toronto.
3. Latitude and Longitude information of each neighbourhood.
4. Number of coffee shop at each neighbourhood and their average user rating.

How the data will be used to solve the problem:

1. Location wise number of coffee shop and their mean user rating has been prepared.
2. Each coffee shop in each neighbourhood caters how many populations have been calculated.
3. Then the mean user rating of each area is divided by the number of population per coffee shop to get the business opportunity index. The lower the business opportunity index, higher is the opportunity of success if we open a coffee shop in that area.

Sources of data used in the Capstone Project:

1. Neighbourhood Data:
   URL: https://open.toronto.ca/dataset/wellbeing-toronto-demographics/
2. Population Data:
   URL: https://open.toronto.ca/dataset/wellbeing-toronto-demographics/
3. Location data of Toronto neighbourhood:
   https://www.here.com/
4. Location data and user rating of each coffee shop:
   https://foursquare.com/developers/apps

Methodology

The neighbourhood data and the respective population data was derived from the official open source data of Canada. The dataset available in their website was used for the reliability of the data.

	HoodID	Neighbourhood	Pop2016	Pop2011	PopChg11t16
0	1	West Humber-Clairville	33312	34100	-0.023109
1	2	Mount Olive-Silverstone-Jamestown	32954	32788	0.005063
2	3	Thistletown-Beaumond Heights	10360	10138	0.021898
3	4	Rexdale-Kipling	10529	10488	0.003909
4	5	Elms-Old Rexdale	9456	9550	-0.009843
...	...	...	...	...	...
135	136	West Hill	27392	26547	0.031830
136	137	Woburn	53485	53350	0.002530
137	138	Eglinton East	22776	22829	-0.002322
138	139	Scarborough Village	16724	16609	0.006924
139	140	Guildwood	9917	9816	0.010289

140 rows × 5 columns

A Snapshot of the Dataset

The latitude and longitude data of each neighbourhood was obtained from “HERE Map” by using a free account. As access was limited, the downloaded data was stored in a local “CSV” file and the same was used afterwards.

	Neighbourhood	Latitude	Longitude	Population (As on 2016)
0	West Humber-Clairville	43.71516	-79.59251	33312
1	Mount Olive-Silverstone-Jamestown	43.74559	-79.58605	32954
2	Thistletown-Beaumond Heights	43.73755	-79.56456	10360
3	Rexdale-Kipling	43.72424	-79.56643	10529
4	Elms-Old Rexdale	43.71964	-79.55064	9456
...	...	...	...	...
135	West Hill	43.76436	-79.16839	27392
136	Woburn	42.47927	-71.15016	53485
137	Eglinton East	43.74264	-79.24017	22776
138	Scarborough Village	43.73626	-79.21600	16724
139	Guildwood	43.74596	-79.19747	9917

140 rows × 4 columns

A Snapshot of the Dataset

“Foursquare API” was used to get the venue details around each neighbourhood. Further the dataset was filtered by “Coffee Shop”. The user rating of each coffee shop was also obtained by the “Foursquare API”. This is a premium call and only 50 calls are allowed per day. As the dataset after filtering was having 122 coffee shop, the data was obtained over a period of 3 days and the same was saved in a “.csv” file. This CSV file was used afterwards. The venues which are not rated yet were removed from the dataset. This value was only 10.65% of the total dataset.

	Neighbourhood	Latitude	Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
0	West Humber-Clairville	43.71516	-79.59251	Tim Hortons	4be452122457a593c40eaa15	43.714657	-79.593716	Coffee Shop
1	West Humber-Clairville	43.71516	-79.59251	Xawaash	561d7a54498ed5c87ce4159d	43.715786	-79.593053	Mediterranean Restaurant
2	West Humber-Clairville	43.71516	-79.59251	Staples Rexdale	4d6ec39e12196ea832fa548e	43.718539	-79.594570	Paper / Office Supplies Store
3	West Humber-Clairville	43.71516	-79.59251	Swiss Pick	4cc21af082388cfae8f25d35	43.716150	-79.593843	Swiss Restaurant
4	West Humber-Clairville	43.71516	-79.59251	Enterprise Rent-A-Car	4d764ad4f5e4b1f7b07e0187	43.715260	-79.589320	Rental Car Location

	Venue ID	Average Rating
0	4be452122457a593c40eaa15	6.8
1	4cd9d00734bb8cfa6576babf	5.5
2	4d84b99a02eb5481040f34f5	6.2
3	50d4866ee4b0c1554d94fef7	7.4
4	4c92baf61221b1f7c64514e2	NaN

A Snapshot of the Dataset

Neighbourhood-wise no of coffee shop and their mean average user rating was calculated.

	Neighbourhood	No of Coffee Shop	Average Rating
0	Alderwood	1	5.8000
1	Annex	1	6.3000
2	Bay Street Corridor	8	7.7875
3	Bedford Park-Nortown	1	7.0000
4	Briar Hill-Belgravia	1	7.3000

A Snapshot of the Dataset

The population data as per the neighbourhood was appended and population catered by each coffee shop in those neighbourhoods was calculated in a separate column.

	Neighbourhood	No of Coffee Shop	Average Rating	Population	Population / Coffee Shop
0	Alderwood	1	5.8000	12054	12054
1	Annex	1	6.3000	30526	30526
2	Bay Street Corridor	8	7.7875	25797	3224.62
3	Bedford Park-Nortown	1	7.0000	23236	23236
4	Briar Hill-Belgravia	1	7.3000	14257	14257

A Snapshot of the Dataset

The “Population/Coffee Shop” data was scaled in the range 0 – 10 for better visualisation and understanding. The mean average rating and the population / coffee shop are the two deciding factors in considering a specific neighbourhood for opening a coffee shop. The idea behind this is as follows:

1. Higher the population/coffee shop better is the opportunity.
2. Lower the average rating higher is the chance of success.

	Neighbourhood	No of Coffee Shop	Average Rating	Population	Population / Coffee Shop	Pop/CofShop Scaled
0	Malvern	1	6.9	43794	43794	10
1	West Humber-Clairville	1	6.8	33312	33312	7.60652
2	Mount Olive-Silverstone-Jamestown	1	5.5	32954	32954	7.52478
3	Annex	1	6.3	30526	30526	6.97036
4	York University Heights	1	7.2	27593	27593	6.30063

A Snapshot of the Dataset

So we have created a BO Index (business opportunity index) which is population / coffee shop divided by mean of average user rating. So, lower is the BO Index higher is the opportunity. The resultant data was sorted in ascending order for better visualisation.

	Neighbourhood	No of Coffee Shop	Average Rating	Population	Population / Coffee Shop	Pop/CofShop Scaled	Business Opportunity Index
0	Malvern	1	6.9	43794	43794	10	0.69
2	Mount Olive-Silverstone-Jamestown	1	5.5	32954	32954	7.52478	0.730919
1	West Humber-Clairville	1	6.8	33312	33312	7.60652	0.89397
3	Annex	1	6.3	30526	30526	6.97036	0.903827
4	York University Heights	1	7.2	27593	27593	6.30063	1.14274

A Snapshot of the Dataset

Note: As in my course, I have learnt only K-means clustering and as it may not be applicable in this problem, I have not used any machine learning algorithm.

Results

	Neighbourhood	No of Coffee Shop	Average Rating	Population	Population / Coffee Shop	Pop/CofShop Scaled	Business Opportunity Index
0	Malvern	1	6.9	43794	43794	10	0.69
2	Mount Olive-Silverstone-Jamestown	1	5.5	32954	32954	7.52478	0.730919
1	West Humber-Clairville	1	6.8	33312	33312	7.60652	0.89397
3	Annex	1	6.3	30526	30526	6.97036	0.903827
4	York University Heights	1	7.2	27593	27593	6.30063	1.14274

It is evident from the resultant dataset that chances of success in coffee shop business are highest in Malvern area and then it follows as shown in the table snapshot.

Discussion

1. The resultant dataset clearly shows the neighbourhood of Toronto with higher opportunity for coffee shop business.
2. The methodology can be adopted for any other business ideas.
3. The Limitation:
    a. The user rating used in the project is irrespective of the no of user rated. Number of user rating also plays an important role in deciding the quality and popularity of any business.
    b. The neighbourhood area was not considered while arriving at the result. The neighbourhood with small area may be catered by smaller no of coffee shop.
4. Future Scope of Improvement:
    a. The population age category, no of user rating along with average user rating, footfall data, area of each neighbourhood may be included in the analysis.
    b. Some statistical analysis and machine algorithm may be used for achieving the control over the data and arriving at better result.
    c. Considering the business specific data which have a deciding role in the success of the particular business.

Conclusion

If you can't explain it simply, you don't understand it well enough.

—Albert Einstein

The practice of data science can best be described as a combination of analytical engineering and exploration. The business presents a problem we would like to solve. Rarely is the business problem directly one of our basic data mining tasks. We decompose the problem into subtasks that we think we can solve, usually starting with existing tools. For some of these tasks we may not know how well we can solve them, so we have to mine the data and conduct evaluation to see. If that does not succeed, we may need to try something completely different. In the process we may discover knowledge that will help us to solve the problem we had set out to solve, or we may discover something unexpected that leads us to other important successes. Neither the analytical engineering nor the exploration should be omitted when considering the application of data science methods to solve a business problem. Omitting the engineering aspect usually makes it much less likely that the results of mining data will actually solve the business problem. Omitting the understanding of process as one of exploration and discovery often keeps an organization from putting the right management, incentives, and investments in place for the project to succeed.

Finally I would like to thank the instructors of the course for beautifully designing the course as this course is well designed for people from non-computer background.