Selecting a Location Near a University to Open a Student Coffee Shop

(Applied Data Science Capstone - Final Report)

Hiba Jadallah

October, 2019

Table of Contents

Tables list

Figures List

Selecting a Location Near a University to Open a Student Coffee Shop

(Applied Data Science Capstone - Final Report)

Hiba Jadallah, October, 2019

Introduction

The idea of studying in a coffee shop can be more inviting than studying at a library. Researchers at the University of Illinois [1] found that ambient noise, such as the sound of coffee machines, boost the creativity. In their own words "…instead of burying oneself in a quiet room trying to figure out a solution, walking out of one's comfort zone and getting into a relatively noisy environment (such as a cafe´) may trigger the brain to think abstractly, and thus generate creative ideas.".

Based on that study and on students' constant need for coffee as a primary source of caffeine [2], investors can find a profitable business opportunity by opening a coffee shop that is focused primarily on serving students. Hence, the location of this coffee shop should be near or around a university campus. The selection of near which campus to start a coffee shop business that turns into a successful business can rely on several factors including the number of enrolled students at the university, number of coffee shops in the area and their distance from the campus, expected customers spending habits…etc.

Business Problem

An investor who is interested in opening a coffee shop in Ontario, Canada, that is focused primarily on serving college students would need an answer to the following question: Near which university campus in Ontario should the investor consider starting his business?

Target Audience

This study is targeted towards investors who are planning on starting a coffee shop business that is focused on serving college students. Also, real estate agencies could be interested in the results as to know best places to recommend for their customers who want to start a coffee shop business.

Data Acquisition and Cleaning

Data Sources

In order to answer our question, we will need the following data:

1. List of universities in Ontario, Canada. This Wikipedia page (https://en.wikipedia.org/wiki/List_of_universities_in_Canada) contains a list of all universities in Canada, from which we will only considered universities in Ontario.

2. Although the Wikipedia page contains information about enrollment, it is not updated. We want most updated enrollment data, so it will be taken from Universities Canada website: (https://www.univcan.ca/universities/facts-and-stats/enrolment-by-university)

3. We also consider the tuition fees data for each university as an indicator to income and spending habits for prospect customers. This data will also be taken from Universities Canada website: (https://www.univcan.ca/universities/facts-and-stats/tuition-fees-by-university/)

4. The population of the cities in which the universities are located will be added for an overview of overall population in the city of the university. This data will be taken from the World Population Review website: (http://worldpopulationreview.com/countries/canada-population/cities/).

5. **Foursquare API** will be used to obtain data on coffee shops around target universities.

Data Cleaning

Data was scraped from the above mentioned data sources starting from the Wikipedia page that contains the list of universities in Canada (table 1). The target is to study universities in Ontario, so for each data frame we obtained it was necessary to filter it out based on the province. Additionally, we only considered the city of the main campus of each university (table 2).

	Name	City	Province	Language	Est.	Under-graduates	Post-graduates	Total students	Notes
0	Athabasca University	Athabasca, Calgary, Edmonton	Alberta	English	1970	36240	3460	39700	[38]
1	MacEwan University	Edmonton	Alberta	English	1971	18897	0	18897	[39]
2	Mount Royal University	Calgary	Alberta	English	1910	24768	0	24768	[40]
3	University of Alberta	Edmonton, Camrose, Calgary	Alberta	Bilingual	1906	31904	7598	39502	[41]
4	University of Calgary	Calgary, Edmonton	Alberta	English	1966	23320	6540	29860	[42]

Table 1 Data scraped from Wikipedia

	University	Total students	Main Campus City
36	Algoma University	1150	Sault Ste. Marie
37	Brock University	17006	St. Catharines
38	Carleton University	24250	Ottawa
39	Dominican University College	244	Ottawa
40	Lakehead University	8050	Thunder Bay

Table 2 Wikipedia data after cleaning

The data also required some cleaning, for example the tuition fees data was read as a strings (objects) and had some characters in addition to the digits, such as asterisks and dollar sign (table 3). This would not allow casting to float type in order to be able to perform mathematical operations or to simply deal with the numeric value of the data.

	University	Undergraduate Canadian Students	Undergraduate Foreign Students	Graduate Canadian Students	Graduate Foreign Students	Province
0	Memorial University of Newfoundland	2, 550−2,550	11, 460−11,460	1, 905−2,859	3, 222−4,833	Newfoundland & Labrador
1	University of Prince Edward Island	6, 150−6,150	13, 326−13,326	4, 134−5,000 *	11, 310−12,176 *	Prince Edward Island
2	Acadia University	7, 390−8,673	17, 363−17,363	7, 341−8,624	17, 532−17,532	Nova Scotia
3	Université Sainte-Anne	6, 277−7,560	10, 340−10,340	6, 277−7,560 *	10, 340−10,340 *	Nova Scotia
4	Dalhousie University	6, 355−7,638	17, 148−17,148	6, 805−8,088	17, 598−17,598	Nova Scotia

Table 3 Tuition fees data as scraped from source

Therefore, it was necessary to remove all additional characters, and then split each range to two columns, and finally cast to float. I have decided to only consider the tuition fees of the undergraduate programs as data showed that the majority of students in these universities are undergraduate students (table 4).

	University	Average Tuition
0	Algoma University	11884.50
1	Brock University	15981.50
2	Carleton University	15848.25
3	University of Guelph	11284.00
4	Lakehead University	14462.25

Table 4 Tuition fees data after cleaning and averaging

It was then essential to make sure that the universities names and the cities names match across the different data frames as we want to merge them all into one data frame. That required removing all added text or characters around universities names. For instance, some universities names were followed by some text within parenthesis (table 5). So, I had to remove the parenthesis and the text and any additional spaces after a university name (table 6).

	University	Full-time Undergraduate	Full-time Graduate	Part-time Undergraduate	Part-time Graduate	Province
33	Algoma University	1000	0	300	0	Ontario
34	Brock University	15300	1400	2100	350	Ontario
35	Carleton University (includes Dominican Univer...	22100	3600	5200	600	Ontario
36	University of Guelph	24000	2700	3200	220	Ontario
37	Lakehead University	6000	1300	1300	20	Ontario

Table 5 Enrollment data as scraped from source

	University	Total Students 2019
0	Algoma University	1300
1	Brock University	19150
2	Carleton University	31500
3	University of Guelph	30120
4	Lakehead University	8620

Table 6 Enrollment data after cleaning universities names and finding total

After merging data from all the data frames, the final frame contained all the data and the coordinates of each university and city (table 7).

	University	Main Campus City	Total Students	Average Tuition	City Population	City_Latitude	City_Longitude	Uni_Latitude	Uni_Longitude
0	Algoma University	Sault Ste. Marie	1300	11884.50	74948	46.513600	-84.335800	46.5015	-84.2879
1	Brock University	St. Catharines	19150	15981.50	131989	43.159400	-79.246900	43.1176	-79.2477
2	Carleton University	Ottawa	31500	15848.25	812129	45.416667	-75.700000	45.3876	-75.6960
3	Lakehead University	Thunder Bay	8620	14462.25	99334	48.400000	-89.233333	48.4215	-89.2619
4	Laurentian University	Sudbury	9020	15043.25	157857	46.500000	-80.966667	46.4668	-80.9742
5	McMaster University	Hamilton	33960	15529.25	519949	43.256101	-79.857484	43.2609	-79.9192
6	Nipissing University	North Bay	5100	12874.00	50170	46.300000	-79.450000	46.3373	-79.4887
7	Ontario College of Art and Design University	Toronto	4510	14486.25	2600000	43.666667	-79.416667	43.6530	-79.3912

Table 7 Final data frame including all collected data

This is the visualization of the universities' locations on Ontario map.
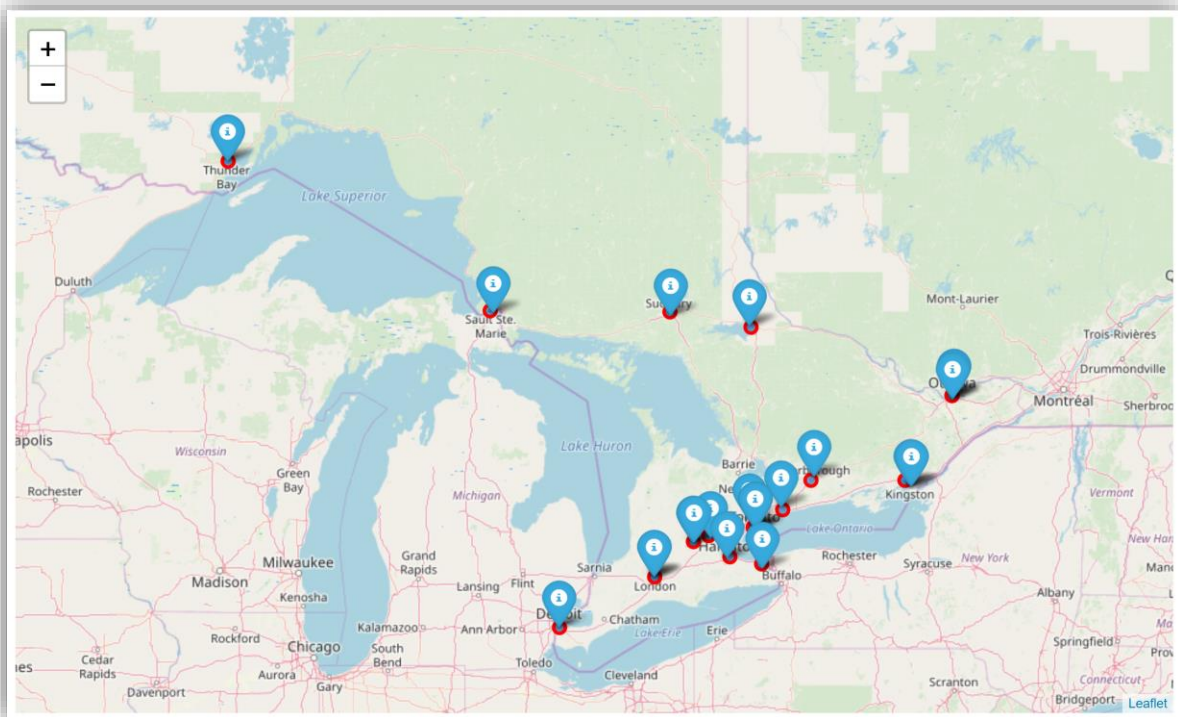

Figure 1 Universities on Ontario Map

I have used Foursquare API to retrieve coffee shops around the location of each university. Sample of the results is shown in table 8.

	Location	LocationLatitude	LocationLongitude	Venue	VenueLatitude	VenueLongitude	Venue Category
0	Algoma University	46.5015	-84.2879	Tim Hortons	46.501103	-84.287194	Coffee Shop
1	Brock University	43.1176	-79.2477	Tim Hortons	43.119688	-79.248627	Coffee Shop
2	Brock University	43.1176	-79.2477	Tim Hortons	43.118233	-79.241331	Coffee Shop
3	Carleton University	45.3876	-75.6960	Rooster's Coffeehouse	45.383025	-75.698155	Coffee Shop
4	Carleton University	45.3876	-75.6960	Tim Hortons	45.386054	-75.693545	Coffee Shop

Table 8 Example of the coffee shops venues received from Foursquare API

Methodology and Analysis

I started the analysis by looking at the basic statistics of the data frame as shown in table 9.

`all_data.describe()`

	Total Students	Average Tuition	City Population	City_Latitude	City_Longitude	Uni_Latitude	Uni_Longitude
count	23.000000	23.000000	2.300000e+01	23.000000	23.000000	23.000000	23.000000
mean	23422.173913	15194.413043	6.828323e+05	44.373369	-79.783040	44.375348	-79.781143
std	21957.898549	4063.501868	9.310037e+05	1.473245	3.031486	1.471303	3.047042
min	520.000000	5720.000000	5.017000e+04	42.301649	-89.233333	42.304300	-89.261900
25%	6860.000000	13252.250000	1.067645e+05	43.464300	-80.743533	43.472350	-80.759550
50%	19150.000000	15075.500000	2.479890e+05	43.666667	-79.416667	43.773500	-79.488700
75%	32915.000000	16555.625000	8.121290e+05	45.416667	-78.600000	45.397750	-78.593050
max	91300.000000	26630.000000	2.600000e+06	48.400000	-75.700000	48.421500	-75.676500

Table 9 Basic statistics analysis

The student enrollment goes as low as 520 students at some universities. I decided to only consider universities with enrollment greater than 3000 students. This is basically the number of target customers.

I counted the number of coffee shops around each university and around each city center. The idea of having the city center venues in addition to the universities' areas venues is to have an idea of the coffee shops market in the city of the university in general in addition to the neighborhood of the university in specific. However, it shall not be included in the clustering data.

I also computed the distance between each coffee shop and the university close to it. Then, I computed the average distance of coffee shops around each university. All the data was then merged into the data frame (table 10).

	University	Main Campus City	Total Students	Average Tuition	City Population	City_Latitude	City_Longitude	Uni_Latitude	Uni_Longitude	NumberOfUniCafes	NumberOfCityCafes	DistanceToUni
0	Brock University	St. Catharines	19150	15981.50	131989	43.159400	-79.246900	43.1176	-79.2477	2	3.0	0.456648
1	Carleton University	Ottawa	31500	15848.25	812129	45.416667	-75.700000	45.3876	-75.6960	4	51.0	0.487904
2	Lakehead University	Thunder Bay	8620	14462.25	99334	48.400000	-89.233333	48.4215	-89.2619	3	2.0	0.261785
3	Laurentian University	Sudbury	9020	15043.25	157857	46.500000	-80.966667	46.4668	-80.9742	2	1.0	0.336817
4	McMaster University	Hamilton	33960	15529.25	519949	43.256101	-79.857484	43.2609	-79.9192	8	14.0	0.660482
5	Nipissing University	North Bay	5100	12874.00	50170	46.300000	-79.450000	46.3373	-79.4887	2	1.0	0.374839
6	Ontario College of Art and Design University	Toronto	4510	14486.25	2600000	43.666667	-79.416667	43.6530	-79.3912	13	52.0	0.689551
7	Queen's University	Kingston	29880	20992.00	114195	44.300000	-76.566667	44.2253	-76.4951	8	0.0	0.673909
8	Ryerson University	Toronto	43160	15606.75	2600000	43.666667	-79.416667	43.6577	-79.3788	10	52.0	0.728134
9	Trent University	Peterborough	10350	13280.50	75877	44.300000	-78.333333	44.3564	-78.2902	1	1.0	0.797926
10	University of Guelph	Guelph	30120	11284.00	115760	43.544800	-80.248200	43.5327	-80.2262	5	9.0	0.359826
11	University of Ontario Institute of Technology	Oshawa	10120	13224.00	247989	43.900000	-78.866667	43.9458	-78.8959	3	5.0	0.472030
12	University of Ottawa	Ottawa	43400	19104.50	812129	45.416667	-75.700000	45.4231	-75.6831	15	51.0	0.600505

Table 10 Data frame after adding Foursquare data, number of coffee shop venues and average distance

I plotted a scatter plot between the number of enrolled students and the number of coffee shops around each university to see the relation between them. Though not persistent, the number of coffee shop around a university does increase with the increase in the number of enrolled students (figure 2).
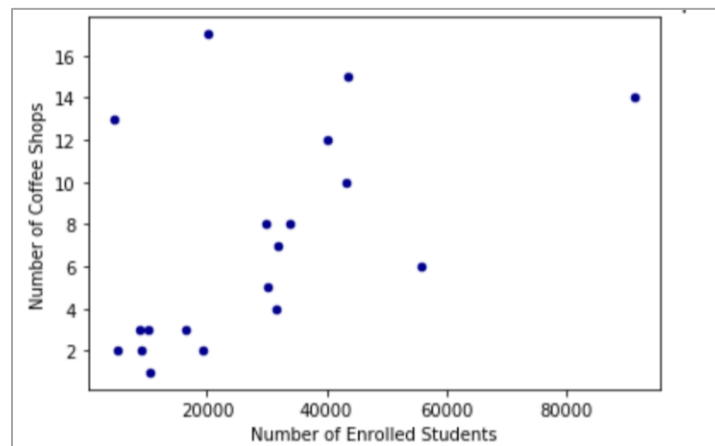

Figure 2 Relation between enrolled students and number of coffee shops around the university

I also plotted histograms of the students' enrollment (figure 3) and the tuition fees (figure 4) to see their distribution. It was noted that one university had significantly higher number of enrolled students in comparison to the other universities in the study. The tuition fees in most universities

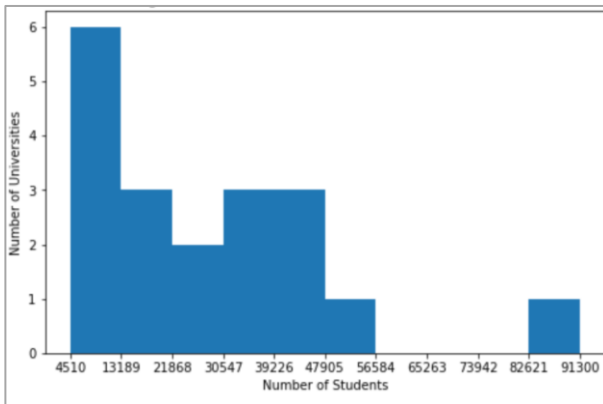in the study is around 15000. Again, there was one university with higher tuition fees than all the others.

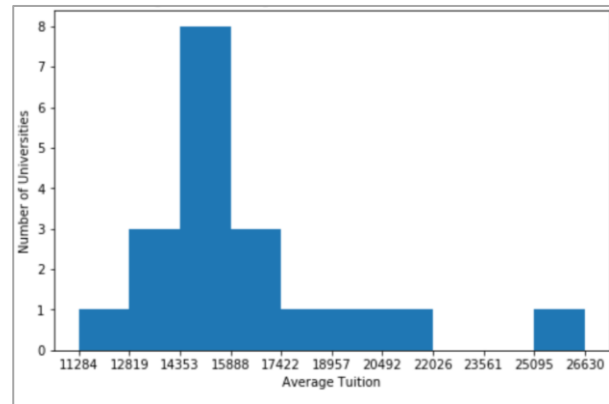*Figure 3 Histogram of students' enrollment in Ontario universities*

Figure 4 Histogram of average tuition fees in Ontario universities

I have used k-Means clustering to cluster the universities areas to three clusters. I have used the total students, average tuition fees, number of coffee shops around a university, and average distance between coffee shops and a university data for clustering. The map in figure 5 shows the visualization of the clusters and table 11 shows the labels in the data frame.
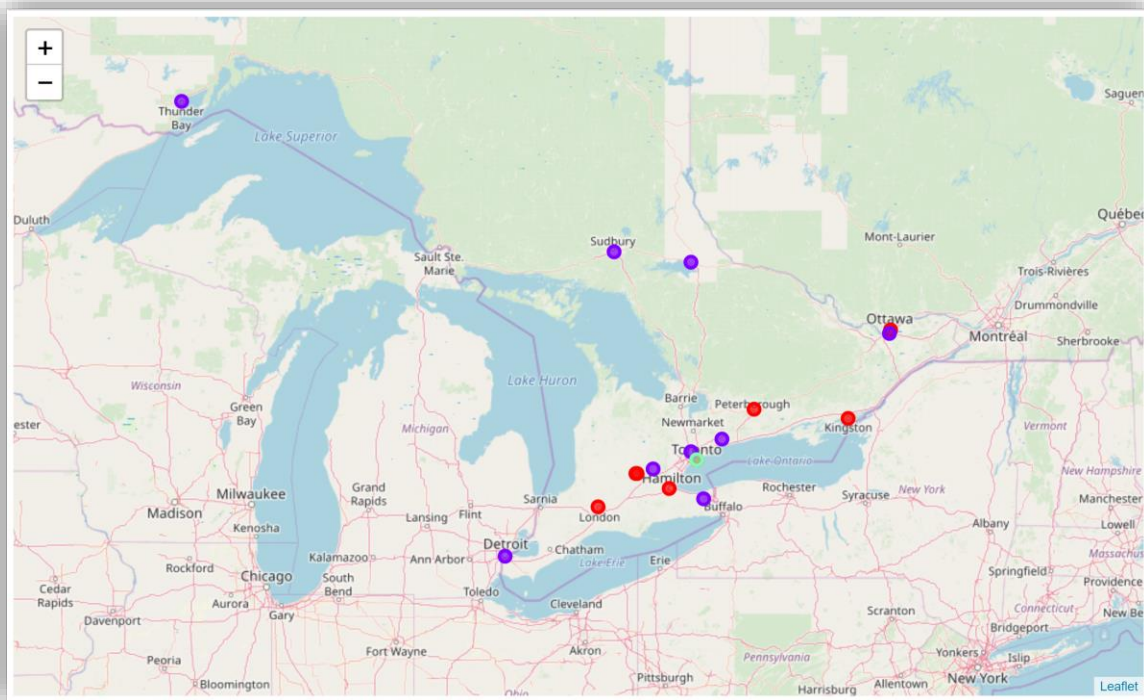

Figure 5 Map showing clusters of universities after applying k-Means

	University	Main Campus City	Total Students	Average Tuition	City Population	City_Latitude	City_Longitude	Uni_Latitude	Uni_Longitude	NumberOfUniCafes	NumberOfCityCafes	DistanceToUni	Labels
9	Trent University	Peterborough	10350	13280.50	75877	44.300000	-78.333333	44.3564	-78.2902	1	1.0	0.797926	0
15	University of Western Ontario	London	31870	16963.75	346765	42.983333	-81.250000	43.0096	-81.2737	7	20.0	0.656134	0
14	University of Waterloo	Waterloo	40000	18333.00	97475	43.464300	-80.520400	43.4723	-80.5449	12	26.0	0.533785	0
4	McMaster University	Hamilton	33960	15529.25	519949	43.256101	-79.857484	43.2609	-79.9192	8	14.0	0.660482	0
12	University of Ottawa	Ottawa	43400	19104.50	812129	45.416667	-75.700000	45.4231	-75.6831	15	51.0	0.600505	0
6	Ontario College of Art and Design University	Toronto	4510	14486.25	2600000	43.666667	-79.416667	43.6530	-79.3912	13	52.0	0.689551	0
7	Queen's University	Kingston	29880	20992.00	114195	44.300000	-76.566667	44.2253	-76.4951	8	0.0	0.673909	0
8	Ryerson University	Toronto	43160	15606.75	2600000	43.666667	-79.416667	43.6577	-79.3788	10	52.0	0.728134	0
17	Wilfrid Laurier University	Waterloo	20100	15075.50	97475	43.464300	-80.520400	43.4724	-80.5263	17	26.0	0.595299	0
16	University of Windsor	Windsor	16410	14762.00	278013	42.301649	-83.030744	42.3043	-83.0660	3	0.0	0.317714	1
11	University of Ontario Institute of Technology	Oshawa	10120	13224.00	247989	43.900000	-78.866667	43.9458	-78.8959	3	5.0	0.472030	1
0	Brock University	St. Catharines	19150	15981.50	131989	43.159400	-79.246900	43.1176	-79.2477	2	3.0	0.456648	1
5	Nipissing University	North Bay	5100	12874.00	50170	46.300000	-79.450000	46.3373	-79.4887	2	1.0	0.374839	1
3	Laurentian University	Sudbury	9020	15043.25	157857	46.500000	-80.966667	46.4668	-80.9742	2	1.0	0.336817	1
2	Lakehead University	Thunder Bay	8620	14462.25	99334	48.400000	-89.233333	48.4215	-89.2619	3	2.0	0.261785	1
1	Carleton University	Ottawa	31500	15848.25	812129	45.416667	-75.700000	45.3876	-75.6960	4	51.0	0.487904	1
10	University of Guelph	Guelph	30120	11284.00	115760	43.544800	-80.248200	43.5327	-80.2262	5	9.0	0.359826	1
18	York University	Toronto	55900	16147.50	2600000	43.666667	-79.416667	43.7735	-79.5019	6	52.0	0.378519	1
13	University of Toronto	Toronto	91300	26630.00	2600000	43.666667	-79.416667	43.6629	-79.3957	14	52.0	0.596968	2

Table 11 Clusters labels added to the data frame

Labels	Total Students	Average Tuition	City Population	City_Latitude	City_Longitude	Uni_Latitude	Uni_Longitude	NumberOfUniCafes	NumberOfCityCafes	DistanceToUni
0	28581.111111	16596.833333	8.070961e+05	43.835337	-79.064624	43.836744	-79.055833	10.111111	26.888889	0.659525
1	20660.000000	14402.972222	4.992490e+05	44.798798	-80.684353	44.809678	-80.706500	3.333333	13.777778	0.382898
2	91300.000000	26630.000000	2.600000e+06	43.666667	-79.416667	43.662900	-79.395700	14.000000	52.000000	0.596968

Table 12 Mean of Clusters Data

Results

The analysis shows that cluster number 2 has only one university, that is University of Toronto. It was no surprise that no other university fitted within its cluster, taking into consideration the very high number of students in it. Although it has a relatively high number of coffee shops near, we still recommend the investor to consider it for further analysis as the number of students is significantly high and the tuition fees are also high indicating a possibly good market.

Cluster 1, as seen in tables 11 and 12, mostly includes universities with low number of coffee shops around universities reducing the risk of competition. On the other hand, cluster 0 has mostly universities with higher number of coffee shops around them and on average higher number of enrolled students than cluster 1.

Discussion

As a result of this analysis we can recommend the investor to look further around University of Toronto. However, we believe more features could serve in better clustering for the universities in this study, especially in clusters 1 and 0.

It would be beneficial to investigate the availability of cafeterias within campuses and how do students feed about them, and to study locations on campuses.

Additionally, it is possible to survey the students of the selected universities to study their interest in opening a new coffee shop. It would also be a good opportunity to ask students about how much on average they spend on coffee per week, for example, or their willingness to study at a coffee shop.

Conclusion and Future Directions

Considering students constant need for coffee and their interest in studying at coffee shops, starting a Student Coffee Shop near a university campus has a high potential for success. The main question is where to open it? In this study I worked on narrowing down the search area for a potential investor looking for a location in Ontario, Canada, to start such business. The analysis resulted in recommending University of Toronto and probably deeper research for the other clusters and the potential market.

Investors are recommended to thoroughly study locations within the streets around the selected campus. The analysis should take into consideration the success rates of existing businesses around the area as successful businesses usually attract other successful businesses.

References

[1] Mehta, Ravi, Rui Zhu, and Amar Cheema. "Is noise always bad? Exploring the effects of ambient noise on creative cognition." Journal of Consumer Research 39, no. 4 (2012): 784-799

[2] Mahoney, Caroline R., Grace E. Giles, Bernadette P. Marriott, Daniel A. Judelson, Ellen L. Glickman, Paula J. Geiselman, and Harris R. Lieberman. "Intake of caffeine from all sources and reasons for use by college students." Clinical Nutrition 38, no. 2 (2019): 668-675.