Food around the Universities in Canada

Applied Data Science Capstone
Coursera Capstone Project - The Battle of Neighborhoods

Ripple Shi
5/20/2020

This report is drafted for the capstone project of the specialization Applied Data Science Capstone offered on Coursera by IBM. In this report, we will explore the recommended food options near the universities in Canada using the machine learning algorithm.

Table of Contents

1. INTRODUCTION

I believe food is an important part of life. Food provides you with nutrition, energy and ideally, satisfaction. As someone who will soon enroll in a university in Canada, I am really curious about what types of food I can get there. I was born and raised in Asia, where the diets are quite different from those in North America, so it will be a great relief if I know whether I could easily find my familiar types of meals nearby.

I am sure this is also a concern for others who are seeking to study, work or live in a different country. However, limit to the scale of the project, we will only explore different food suppliers around Canadian universities. They could be restaurants, café or any other venues that could satisfy people's needs in food.

I hope this project could provide the readers with some insights on this subject. Although here we will focus on food, universities and Canada, I think the idea behind this project can also be applied to any similar intention.

To carry out the project, we are going to rely on the recommendations provided by Foursquare's API. By specifying the coordinate of the university, Foursquare will return us some recommended venues that are in the food section within the limit and distance we set. Meanwhile, the category of a venue will also be returned, so we could use that to guess the main cuisine of the venue.

Also, we will use K-Means, a clustering algorithm, to cluster the universities in groups and explore the features of each group. Hopefully, we could get some interesting findings out of that.

2. DATA

The data used in this project come from three sources.

We start by getting a list of universities in Canada from Wikipedia[1]. Although we are not sure whether the list is complete, it should be enough to represent the population we are interested in. Here we define the universities included in this list to be the study object of this project. There are 91 universities contained in the list, most of them are public universities, 16 are private universities. The list also provides the provinces and the cities where the universities locate. This information will play an important role in determining the location of the universities.

Next, we will use the name and the province of the universities to get their coordinates. We obtain the latitude and longitude of each university using the dataset on http://py4edata.dr-chuck.net/. This is a subset of data from the Google Geocoding API, established by Dr. Charles R. Severance from University of Michigan. This data set is built to facilitate the study of Python courses taught by Dr. Chuck. Please note that this dataset is not my first choice to get the coordinates. Recall that we only have the name of a university and the city and the province where it locates. Due to the limits of the available geocoding APIs, I cannot get the coordinates of the universities only using that information as the parameters. It turned out that the data set built by Dr. Chuck is the only option I am aware of that could help me attain my goal. We will further explain this problem in the Data Preprocessing section afterwards. Anyway, the coordinates retrieving from this data set allow us to specify the location in the search queries of Foursquare.

Finally, we use Foursquare's API to get the recommendations in the food section and do the analysis. According to the documentation of Foursquare's API, by using the endpoint "explore" we could get a list of recommended venues near the current location. The list includes much information, but we only need the venue name and the venue type. This information will be enough for us to summarize what types of venues we could find around the universities. Based on that, we will build our variables of what percent a venue category is taking among all the recommended venues that meet the conditions we set. These self-created variables will be used to train a model using K-Means algorithm to get the clusters of the universities.

---

[1] The link of the page is https://en.wikipedia.org/wiki/List_of_universities_in_Canada.

3. METHODOLOGY

3.1. Data Preprocessing

3.1.1. List of the Universities

As discussed in the Data section, we need the tables on a Wikipedia page. The function of Pandas read_html makes this process quite easy. There are 14 tables on that page, and what we need is the first 11 tables, 10 for public universities and one for private universities.

	Name	City	Province	Language	Est.			Students	Notes
	Name	City		Language	Est.	Undergrad.	Postgrad.	Total	Notes
0	Alberta University of the Arts	Calgary	Alberta	English	1926	NaN	NaN	1323	NaN
1	Athabasca University	Athabasca, Calgary, Edmonton	Alberta	English	1970	36240.0	3460.0	39700	[38]
2	MacEwan University	Edmonton	Alberta	English	1971	18897.0	0.0	18897	[39]
3	Mount Royal University	Calgary	Alberta	English	1910	24768.0	0.0	24768	[40]
4	University of Alberta	Edmonton, Camrose, Calgary	Alberta	Bilingual	1906	31904.0	7598.0	39502	[41]
...	...	...	...	...	...	...	...	...	...
12	Université du Québec à Rimouski[note 3]	Rimouski and Lévis	Quebec	French	1969	4620.0	810.0	5430	[75]
13	Université du Québec à Trois-Rivières[note 3]	Trois-Rivières	Quebec	French	1969	9160.0	1450.0	10610	[76]
14	Université Laval	Quebec City	Quebec	French	1663	27530.0	10270.0	37800	[77]
0	University of Regina	Regina, Saskatoon, Swift Current	Saskatchewan	Bilingual	1911	10690.0	1480.0	12170	[78]
1	University of Saskatchewan	Saskatoon	Saskatchewan	English	1907	16430.0	2190.0	18620	[79]

Figure 1 Original Table of Public Universities in Canada

	Name	City	Province	Language	Established	Undergraduates	Post-graduates	Total students	Notes
0	Fairleigh Dickinson University (branch)	Vancouver	British Columbia	English	2007	78[failed verification]	50.0	78[failed verification]	[80]
1	New York Institute of Technology (branch)	Vancouver	British Columbia	English	2007	70[failed verification]	40.0	70[failed verification]	[81]
2	Quest University	Squamish	British Columbia	English	2007	700	0.0	700	[82]
3	Niagara University (branch)	Vaughan	Ontario	English	2019	NaN	NaN	NaN	[83]
4	Trinity Western University	Langley	British Columbia	English	1962	2130	730.0	2860	[84]
5	University Canada West	Victoria	British Columbia	English	2005	350[needs update]	0.0	350[needs update]	[85]
6	Booth University College	Winnipeg	Manitoba	English	1982	250	0.0	250	[86]
7	Canadian Mennonite University	Winnipeg	Manitoba	English	1944	600	0.0	600	[56]
8	Kingswood University	Sussex	New Brunswick	English	1945	300	0.0	300	[87][needs update]
9	Crandall University	Moncton	New Brunswick	English	1949	685	0.0	685	[88][needs update]
10	St. Stephen's University	St. Stephen	New Brunswick	English	1975	100	0.0	100	[89][needs update]
11	University of Fredericton	Fredericton	New Brunswick	English	2005	NaN	NaN	NaN	[59][needs update]
12	Atlantic School of Theology	Halifax	Nova Scotia	English	1971	0	124.0	124	[59]
13	Tyndale University	Toronto	Ontario	English	1894	850	0.0	850	[90]
14	Redeemer University College	Ancaster	Ontario	English	1982	955	0.0	955	NaN
15	The King's University	Edmonton	Alberta	English	1979	790	0.0	790	[91]

Figure 2 Original Table of Private Universities in Canada

We will not perform separate operations on public universities and private universities, so it is better to combine these two tables before cleaning. It is clear that the two tables have different formats. The one with public universities has multiple indexes, so the names of the columns are not complete. We refer to the private universities' table to adjust the format of the first table.

Furthermore, we will not use all the columns in the table. We drop the unnecessary columns and leave Name, City and Province as a result. A complete list of universities is as follows. It contains 91 records in total.

	Name	City	Province	Address
0	Alberta University of the Arts	Calgary	Alberta	1407 14 Ave NW, Calgary, AB T2N 4R3, Canada
1	Athabasca University	Athabasca, Calgary, Edmonton	Alberta	1 University Dr, Athabasca, AB T9S 3A3, Canada
2	MacEwan University	Edmonton	Alberta	10700 104 Ave NW, Edmonton, AB T5J 4S2, Canada
3	Mount Royal University	Calgary	Alberta	30 Mt Royal Cir SW, Calgary, AB T3E 7C9, Canada
4	University of Alberta	Edmonton, Camrose, Calgary	Alberta	116 St & 85 Ave, Edmonton, AB T6G 2R3, Canada
...	...	...	...	...
86	University of Fredericton	Fredericton	New Brunswick	3 Bailey Dr, Fredericton, NB E3B 5A3, Canada
87	Atlantic School of Theology	Halifax	Nova Scotia	660 Francklyn St, Halifax, NS B3H 3B6, Canada
88	Tyndale University	Toronto	Ontario	3377 Bayview Ave, North York, ON M2M 3S4, Canada
89	Redeemer University College	Ancaster	Ontario	777 Garner Rd E, Ancaster, ON L9K 1J4, Canada
90	The King's University	Edmonton	Alberta	9125 50 St NW, Edmonton, AB T6B 2H3, Canada

Figure 3 List of Universities in Canada

3.1.2. Coordinates of the Universities

At this stage, we are ready to retrieve the coordinates of the universities. We will send requests to the address where the copy of the previous Google Geocoding API is stored, sending the API key and the addresses of the universities as parameters. Notice that we do not have the formatted addresses of the universities, so we build some informal addresses instead. These informal addresses will be strings contain the name of a university, the province where it locates and the country. An example of the query is "Alberta University of the Arts, Alberta, Canada". Such information will be enough to get the coordinates.

In this process, we got the warning that we failed to retrieve the coordinate of Trent University. So we remove any records that contain at least a NA value. The result after this step is as follows. Now we have only 90 records in total.

	Name	City	Province	Address	Latitude	Longitude
0	Alberta University of the Arts	Calgary	Alberta	1407 14 Ave NW, Calgary, AB T2N 4R3, Canada	51.061571	-114.092098
1	Athabasca University	Athabasca, Calgary, Edmonton	Alberta	1 University Dr, Athabasca, AB T9S 3A3, Canada	54.714955	-113.308545
2	MacEwan University	Edmonton	Alberta	10700 104 Ave NW, Edmonton, AB T5J 4S2, Canada	53.547054	-113.506372
3	Mount Royal University	Calgary	Alberta	30 Mt Royal Cir SW, Calgary, AB T3E 7C9, Canada	51.011010	-114.129778
4	University of Alberta	Edmonton, Camrose, Calgary	Alberta	116 St & 85 Ave, Edmonton, AB T6G 2R3, Canada	53.523219	-113.526319
...	...	...	...	...	...	...
85	University of Fredericton	Fredericton	New Brunswick	3 Bailey Dr, Fredericton, NB E3B 5A3, Canada	45.945570	-66.640826
86	Atlantic School of Theology	Halifax	Nova Scotia	660 Francklyn St, Halifax, NS B3H 3B6, Canada	44.626826	-63.580503
87	Tyndale University	Toronto	Ontario	3377 Bayview Ave, North York, ON M2M 3S4, Canada	43.796851	-79.392186
88	Redeemer University College	Ancaster	Ontario	777 Garner Rd E, Ancaster, ON L9K 1J4, Canada	43.208677	-79.949140
89	The King's University	Edmonton	Alberta	9125 50 St NW, Edmonton, AB T6B 2H3, Canada	53.525418	-113.416752

Figure 4 List of Universities with Address and Coordinate

Here we quickly explain why we give up using the module geocoder to retrieve the latitude and longitude of a university. Below is a sample of the coordinates we got for the first six universities using geocoder.

	0	1
Alberta University of the Arts	51.06438	-114.09212
Athabasca University	53.52289	-113.52627
MacEwan University	53.53948	-113.49236
Mount Royal University	51.01228	-114.13239
University of Alberta	53.52289	-113.52627
University of Calgary	51.07663	-114.13209

Figure 5 Sample of results got using geocoder module

We could see from the result that Athabasca University and University of Alberta share the same coordinate. It is impossible. We further check this location on the map, and it is clear that [53.52289, -113.52627] is the coordinate of University of Alberta.
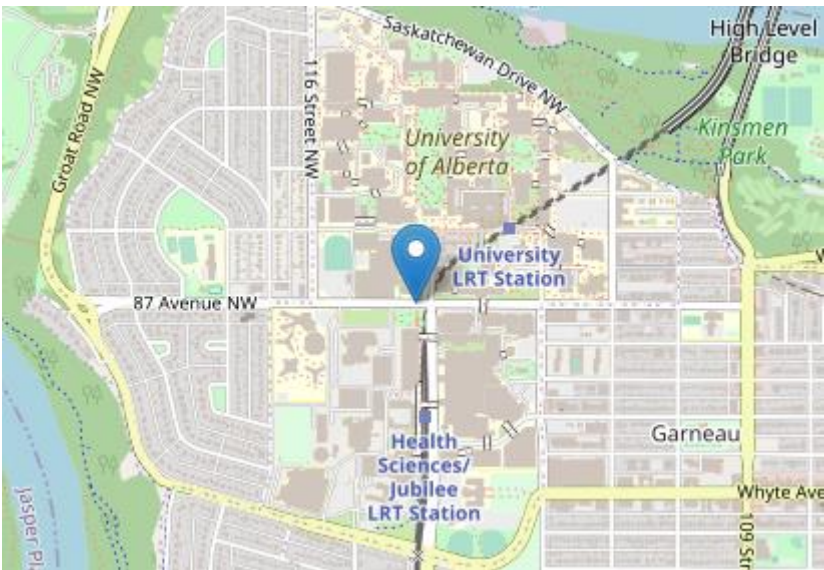

Figure 6 Coordinate of 53.52289, -113.52627

We use the coordinate we got for Athabasca University and display it on the map. In fact, these two universities are quite far from each other.
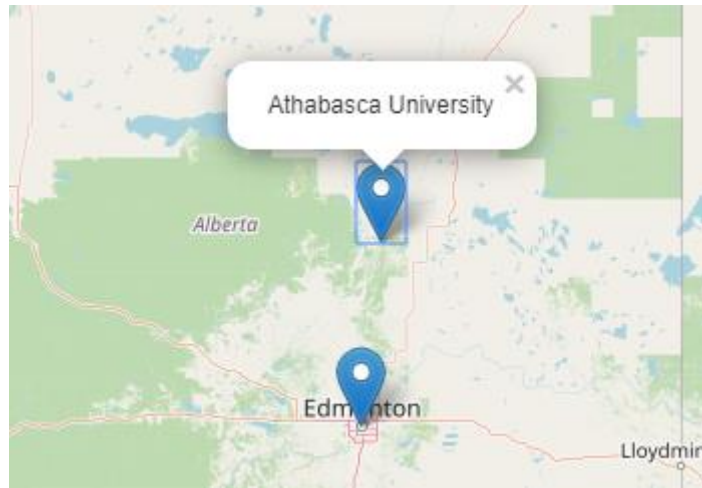
Figure 7 Coordinate of Athabasca University

The result told us that using this way to find the coordinate is not accurate. In fact, I have checked the complete result of the coordinates of all the universities obtained using the geocoder, and have noticed many duplicate values. This is not the fault of the module. The reason is that I did not provide the formatted address but only the name of a university to the API. According to the documentation of the geocoder, the address is necessary to get the correct result.

I also tried finding another API as the replacement. Unfortunately, my attempts all failed. All the available geocoding APIs I found ask for formatted address as a required parameter to carry out the search. Meanwhile, among the API candidates, the accuracy of the result when only providing the name of the university is almost the same. What is worse, I cannot find a list of formatted addresses of the universities in Canada on the Internet. So instead, I decide to use the database provided on http://py4e-data.dr-chuck.net/json. Thank Dr. Chuck for this work.

There is another important point that we should clarify here. We understand that many universities may have two or more campuses. Limit to the operation here, we can only get the coordinate of one of the multiple campuses of each university. Since we already selected a subset of the universities in Canada to proceed with the project, we decide to ignore this issue and go with no matter what campus's coordinate we got from the API for each university to carry out the analysis.

3.1.3. Recommendations of Food around the Universities

After obtaining the coordinates of the universities, we use Foursquare's API to get the recommendations in the food section.

We will use the endpoint "explore" to achieve this goal. There are some parameters needed to be sent with the queries and these parameters will heavily impact the result of our study. First, the version of the API is set to be 16 May 2020, which means any following updates in the API will not be applied in our queries. Next, we narrow down the search to the food section. The radius and the limit parameters are very important. After careful consideration, we decide to use 3,000 meters as the radius and get recommendations not exceed 10 venues. Locations with distances

within 3 kilometres are prone to be considered as ideal options for daily meals. We also consider the situation where the coordinate we got from the previous step is inside the campus while the area of the campus may be big. Taking all these factors into consideration, we think 3 kilometres is a reasonable figure. A maximum of 10 venues near each of the 90 universities will constitute a large enough data set for us to carry out the analysis. Last but not least, include the latitude and the longitude of the university in the query. In this way, the API will look for venues in the food section that are located within 3,000 meters around a circular area around the coordinate given. If 10 or more venues found, the API will only return the first 10 results. When no result got, a blank entry will be recorded except the name for that university.

Below is the result we got. There are 856 records in total, with the venue category we care most in the last column of the table.

	University	Venue name	Venue location lat	Venue location lng	Venue location distance	Venue categories
0	Alberta University of the Arts	Jimmy's A&A Deli	51.070299	-114.092472	971.0	Mediterranean Restaurant
1	Alberta University of the Arts	Vendome Cafe	51.055138	-114.083323	943.0	Café
2	Alberta University of the Arts	Hayden Block	51.052595	-114.088226	1035.0	BBQ Joint
3	Alberta University of the Arts	Wow Chicken	51.054881	-114.085833	864.0	Korean Restaurant
4	Alberta University of the Arts	Peppino	51.052509	-114.090946	1011.0	Diner
...	...	...	...	...	...	...
851	The King's University	A&W	53.541615	-113.417087	1803.0	Fast Food Restaurant
852	The King's University	Sabu Sushi Bar	53.518032	-113.442067	1866.0	Sushi Restaurant
853	The King's University	Subway	53.525481	-113.444089	1809.0	Sandwich Place
854	The King's University	Fargo's	53.540761	-113.424622	1785.0	Steakhouse
855	The King's University	Sawmill Banquet & Catering Centre	53.512940	-113.401028	1735.0	Breakfast Spot

Figure 8 Recommended Venues

The final step of our data preprocessing is to calculate the percentages we needed for K-Means to process.

We build two tables via calculations done on the table we got in the last step, one displays the counts of the recommended venues and one displays the percentages. Other information will all be dropped in this process but only be available in the last table for reference only.

University	Afghan Restaurant	American Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	Brazilian Restaurant	Breakfast Spot	...
Alberta University of the Arts	0	0	0	1	0	1	0	0	0	0	...
Athabasca University	0	1	1	0	0	0	0	0	0	0	...
MacEwan University	0	0	0	0	0	0	0	0	0	0	...
Mount Royal University	0	0	0	0	0	1	0	0	0	1	...
University of Alberta	0	0	0	1	0	0	0	0	0	0	...

Figure 9 Recommended Venues in counts

University	Afghan Restaurant	American Restaurant	Asian Restaurant	BBQ Joint	Bagel Shop	Bakery	Belgian Restaurant	Bistro	Brazilian Restaurant	Breakfast Spot	...
Alberta University of the Arts	0.0	0.000000	0.000000	0.1	0.0	0.1	0.0	0.0	0.0	0.0	...
Athabasca University	0.0	0.166667	0.166667	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
MacEwan University	0.0	0.000000	0.000000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...
Mount Royal University	0.0	0.000000	0.000000	0.0	0.0	0.1	0.0	0.0	0.0	0.1	...
University of Alberta	0.0	0.000000	0.000000	0.1	0.0	0.0	0.0	0.0	0.0	0.0	...

Figure 10 Recommended Venues in percentages

With these two tables, we can move forward to do the analysis.

3.2. Exploratory Data Analysis

Before carrying out the clustering, we take two universities and obtain the recommendations nearby to get some insights.

```
------- Alberta University of the Arts ---------
Gastropub                   1
Diner                       1
BBQ Joint                   1
Italian Restaurant          1
Bakery                      1
Japanese Restaurant         1
Korean Restaurant           1
Sushi Restaurant            1
Café                        1
Mediterranean Restaurant    1
Name: Alberta University of the Arts, dtype: uint8
------- Athabasca University ---------
Sandwich Place              2
American Restaurant         1
Asian Restaurant            1
Gastropub                   1
Burger Joint                1
Name: Athabasca University, dtype: uint8
```

Figure 11 Recommendations near Alberta University of the Arts and Athabasca University in counts

```
------- Alberta University of the Arts ---------
Gastropub                   0.1
Diner                       0.1
BBQ Joint                   0.1
Italian Restaurant          0.1
Bakery                      0.1
Japanese Restaurant         0.1
Korean Restaurant           0.1
Sushi Restaurant            0.1
Café                        0.1
Mediterranean Restaurant    0.1
Name: Alberta University of the Arts, dtype: float64
------- Athabasca University ---------
Sandwich Place              0.333333
American Restaurant         0.166667
Asian Restaurant            0.166667
Gastropub                   0.166667
Burger Joint                0.166667
Name: Athabasca University, dtype: float64
```

Figure 12 Recommendations near Alberta University of the Arts and Athabasca University in percentages

From these two outcomes, we could see that the counts and the percentages are equally important. Although the percentage of each venue type near Alberta University of the Arts is lower than that of Athabasca University, in return it means that there are more food options available near Alberta University of the Arts. But still, we will use the percentages to train our model. Notice that the counts are not continuous and the number of venues returned by the API for each university varies. Hence, using percentages allow us to compare the similarities of food options available near the universities between groups, so we will get better performance from the model and the result is more meaningful.

In addition, recall that we meant to use the name of the category to guess the main cuisine of the venue. However, it turns out that some venues only have vague category names like diner or restaurant. Even though, I think it is still worthwhile to proceed.

3.3. Model training

We already got the percentage of each venue category near each university in previous steps. At

this stage, we are going to use the K-Means clustering algorithm to label the universities.

There are many other clustering algorithms provided by the sklearn package. K-Means is a classic one that clusters the data by calculating the distance between points. When using K-Means to do the clustering, we can define how many clusters we would like the algorithm to return. We have 90 universities in our data set, so we decide to have 6 clusters. If no significant variance observed among universities, theoretically we will have 15 universities in each group. This is a reasonable size for us to carry out exploring the features of each group.

Another parameter we could set is the method for initialization. This method will be used to initialize the cluster centers. Due to the mechanism of the algorithm, the result of the clustering heavily relies on the locations of the initial centers. We use the default "k-means ++" to determine the initial centers.

The result of the model is as follows. Here we only display the first 5 universities to illustrate. The labels returned by the model are in the last column.

	Name	City	Province	Address	Latitude	Longitude	Label
0	Alberta University of the Arts	Calgary	Alberta	1407 14 Ave NW, Calgary, AB T2N 4R3, Canada	51.061571	-114.092098	5
1	Athabasca University	Athabasca, Calgary, Edmonton	Alberta	1 University Dr, Athabasca, AB T9S 3A3, Canada	54.714955	-113.308545	4
2	MacEwan University	Edmonton	Alberta	10700 104 Ave NW, Edmonton, AB T5J 4S2, Canada	53.547054	-113.506372	5
3	Mount Royal University	Calgary	Alberta	30 Mt Royal Cir SW, Calgary, AB T3E 7C9, Canada	51.011010	-114.129778	1
4	University of Alberta	Edmonton, Camrose, Calgary	Alberta	116 St & 85 Ave, Edmonton, AB T6G 2R3, Canada	53.523219	-113.526319	4

Figure 13 Labelled Universities

4. RESULTS

A complete list of the labelled universities is as follows.

```
-------- Label 0 --------
                                                     Name          Province
13                          University of British Columbia  British Columbia
14                                   University of Victoria  British Columbia
21                                   University of Winnipeg          Manitoba
23                                   St. Thomas University     New Brunswick
24                           University of New Brunswick       New Brunswick
25                                   Université de Moncton     New Brunswick
27                                       Acadia University       Nova Scotia
37                                    Carleton University            Ontario
43                             Queen's University at Kingston         Ontario
44                           Royal Military College of Canada        Ontario
45                                      Ryerson University            Ontario
46                           Université de l'Ontario français        Ontario
47                                    University of Guelph            Ontario
50                                    University of Toronto           Ontario
52                              University of Western Ontario         Ontario
58                                    Concordia University            Quebec
59                              École de technologie supérieure        Quebec
60                     École nationale d'administration publique      Quebec
61  Institut national de la recherche scientifique              Quebec
62                                       McGill University            Quebec
63                                    Université de Montréal          Quebec
68                              Université du Québec à Montréal        Quebec
69                              Université du Québec à Rimouski        Quebec
85                                   University of Fredericton   New Brunswick
```

Figure 14 Label 0

```
-------- Label 1 --------
                                                     Name              Province
3                                    Mount Royal University            Alberta
7                                       Capilano University    British Columbia
9                            Kwantlen Polytechnic University   British Columbia
15                             University of the Fraser Valley  British Columbia
20                                   University of Manitoba           Manitoba
26                    Memorial University of Newfoundland  Newfoundland and Labrador
31  Nova Scotia College of Art and Design University           Nova Scotia
38                                     Lakehead University            Ontario
39                                    Laurentian University            Ontario
42       Ontario College of Art and Design University             Ontario
49                                    University of Ottawa            Ontario
51                                    University of Waterloo          Ontario
54                              Wilfrid Laurier University            Ontario
55                                         York University            Ontario
66                        Université du Québec en Outaouais          Quebec
70                    Université du Québec à Trois-Rivières          Quebec
73                                University of Saskatchewan     Saskatchewan
74             Fairleigh Dickinson University (branch)      British Columbia
77                              Niagara University (branch)            Ontario
79                                University Canada West      British Columbia
80                                Booth University College          Manitoba
```

Figure 15 Label 1

```
-------- Label 2 --------
                              Name  Province
81  Canadian Mennonite University  Manitoba
```

Figure 16 Label 2

```
-------- Label 3 --------
                                            Name              Province
5                           University of Calgary              Alberta
18                            Brandon University             Manitoba
28                          Cape Breton University          Nova Scotia
32                    Saint Francis Xavier University      Nova Scotia
35                             Algoma University             Ontario
36                              Brock University             Ontario
41                            Nipissing University            Ontario
56                University of Prince Edward Island  Prince Edward Island
57                            Bishop's University             Quebec
64                        Université de Sherbrooke            Quebec
65   Université du Québec en Abitibi-Témiscamingue        Quebec
72                          University of Regina          Saskatchewan
82                          Kingswood University          New Brunswick
83                           Crandall University          New Brunswick
84                       St. Stephen's University        New Brunswick
88                   Redeemer University College            Ontario
```

Figure 17 Label 3

```
-------- Label 4 --------
                                            Name              Province
1                            Athabasca University            Alberta
4                            University of Alberta           Alberta
6                          University of Lethbridge          Alberta
11                        Simon Fraser University      British Columbia
16   University of Northern British Columbia      British Columbia
19              University College of the North         Manitoba
22                      Mount Allison University       New Brunswick
34                        Université Sainte-Anne         Nova Scotia
48                        Ontario Tech University         Ontario
67              Université du Québec à Chicoutimi          Quebec
78                     Trinity Western University    British Columbia
89                       The King's University           Alberta
```

Figure 18 Label 4

```
-------- Label 5 --------
                                            Name              Province
0                       Alberta University of the Arts       Alberta
2                             MacEwan University            Alberta
8       Emily Carr University of Art and Design   British Columbia
10                        Royal Roads University     British Columbia
12                      Thompson Rivers University   British Columbia
17                    Vancouver Island University    British Columbia
29                          Dalhousie University         Nova Scotia
30                   Mount Saint Vincent University    Nova Scotia
33                        Saint Mary's University        Nova Scotia
40                          McMaster University           Ontario
53                         University of Windsor           Ontario
71                            Université Laval             Quebec
75    New York Institute of Technology (branch)   British Columbia
76                             Quest University     British Columbia
86                     Atlantic School of Theology     Nova Scotia
87                           Tyndale University            Ontario
```

Figure 19 Label 5

While only one universities were assigned to group 2, the other universities have been evenly assigned to different groups.

Label	Name	City	Province	Address	Latitude	Longitude
0	24	24	24	24	24	24
1	21	21	21	21	21	21
2	1	1	1	1	1	1
3	16	16	16	16	16	16
4	12	12	12	12	12	12
5	16	16	16	16	16	16

Figure 20 Summary of the Labels

The distribution of the groups across the provinces is as follows. Subject to the total number of universities in each province, almost every province has three or more categories. It can be inferred that the smaller area unit such as neighbourhood or block matters more than the province.

Label	Province Alberta	British Columbia	Manitoba	New Brunswick	Newfoundland and Labrador	Nova Scotia	Ontario	Prince Edward Island	Quebec	Saskatchewan
0	0	2	1	4	0	1	8	0	8	0
1	1	5	2	0	1	1	8	0	2	1
2	0	0	1	0	0	0	0	0	0	0
3	1	0	1	3	0	2	4	1	3	1
4	4	3	1	1	0	1	1	0	1	0
5	2	6	0	0	0	4	3	0	1	0

Figure 21 Label Distribution across the Provinces

Since there are 90 universities in total, the best way to present our findings is visualization. From the map generated using the folium module, we can see six different colours on the map, each representing a different cluster.
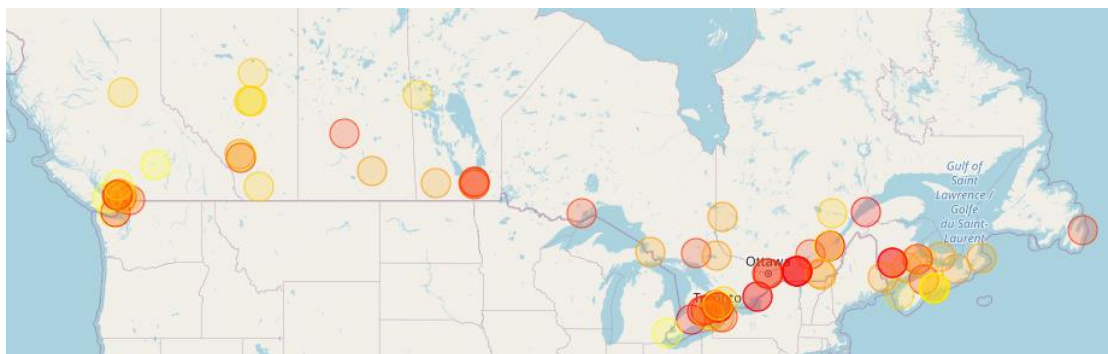

Figure 22 Visualization via the map

We zoom in towards Toronto to get a close view. There are four universities in this area. Although by comparison, they are very close to each other, their labels are not the same. Three of them with red markers are classified to cluster 0, and the one with the orange marker is classified to cluster 1.
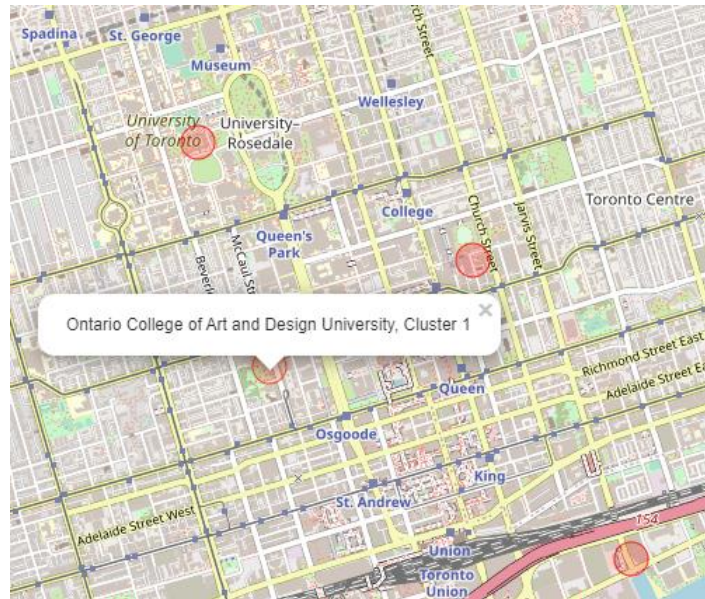
Figure 23 Universities in Toronto

Clearly, we cannot observe any clear geographic pattern of the clustering. Almost every province has universities from each label. This is consistent with what we concluded from the province distribution table that no apparent geographic pattern is observed. Therefore, our next step is to go back to our data frames and see which venue category is more common under each group.

Here we will use two formats to present the result. The first one summarizes the total number of venues under each category and displays them by groups. The second one is based on the result of the first one and is further divided by the number of universities in each group. This allows us to consider the different sizes of the groups when drawing any conclusions. Considering there are too many different venue categories, we only display the top 10 for each group.

```
-------- Label 0 --------              -------- Label 3 --------
Café                            64     Restaurant                 33
Restaurant                      36     Fast Food Restaurant       22
Bakery                          13     Pizza Place                16
Gastropub                       12     Sandwich Place             13
Vegetarian / Vegan Restaurant   11     Diner                       8
Pizza Place                     11     Breakfast Spot              7
French Restaurant               10     Café                        6
Italian Restaurant               7     Wings Joint                 5
Japanese Restaurant              6     Greek Restaurant            4
Breakfast Spot                   5     Gastropub                   3
Name: 0, dtype: uint8                  Name: 3, dtype: uint8
-------- Label 1 --------              -------- Label 4 --------
Restaurant                      29     Sandwich Place             17
Sandwich Place                  15     Fast Food Restaurant       14
Café                            13     Café                       11
Pizza Place                     10     Pizza Place                10
Breakfast Spot                  10     Sushi Restaurant            6
Sushi Restaurant                 8     Burger Joint                6
Bakery                           8     Steakhouse                  5
New American Restaurant          7     American Restaurant         5
Indian Restaurant                7     Diner                       4
Italian Restaurant               6     Italian Restaurant          3
Name: 1, dtype: uint8                  Name: 4, dtype: uint8
-------- Label 2 --------              -------- Label 5 --------
nan                              1     Burger Joint               14
Fast Food Restaurant             0     Pizza Place                13
Chinese Restaurant               0     Sushi Restaurant           13
Creperie                         0     Italian Restaurant         11
Deli / Bodega                    0     Mexican Restaurant          9
Diner                            0     Breakfast Spot              9
Donut Shop                       0     Turkish Restaurant          6
Eastern European Restaurant      0     Japanese Restaurant         6
Fish & Chips Shop                0     Bakery                      6
Wings Joint                      0     Indian Restaurant           5
Name: 2, dtype: uint8                  Name: 5, dtype: uint8
```

Figure 24 Top 10 venue categories

```
-------- Label 0 --------                  -------- Label 3 --------
Café                           2.666667    Restaurant              2.0625
Restaurant                     1.500000    Fast Food Restaurant    1.3750
Bakery                         0.541667    Pizza Place             1.0000
Gastropub                      0.500000    Sandwich Place          0.8125
Vegetarian / Vegan Restaurant  0.458333    Diner                   0.5000
Pizza Place                    0.458333    Breakfast Spot          0.4375
French Restaurant              0.416667    Café                    0.3750
Italian Restaurant             0.291667    Wings Joint             0.3125
Japanese Restaurant            0.250000    Greek Restaurant        0.2500
Breakfast Spot                 0.208333    Gastropub               0.1875
Name: 0, dtype: float64                    Name: 3, dtype: float64
-------- Label 1 --------                  -------- Label 4 --------
Restaurant                     1.380952    Sandwich Place          1.416667
Sandwich Place                 0.714286    Fast Food Restaurant    1.166667
Café                           0.619048    Café                    0.916667
Pizza Place                    0.476190    Pizza Place             0.833333
Breakfast Spot                 0.476190    Sushi Restaurant        0.500000
Sushi Restaurant               0.380952    Burger Joint            0.500000
Bakery                         0.380952    Steakhouse              0.416667
New American Restaurant        0.333333    American Restaurant     0.416667
Indian Restaurant              0.333333    Diner                   0.333333
Italian Restaurant             0.285714    Italian Restaurant      0.250000
Name: 1, dtype: float64                    Name: 4, dtype: float64
-------- Label 2 --------                  -------- Label 5 --------
nan                            1.0         Burger Joint            0.8750
Fast Food Restaurant           0.0         Pizza Place             0.8125
Chinese Restaurant             0.0         Sushi Restaurant        0.8125
Creperie                       0.0         Italian Restaurant      0.6875
Deli / Bodega                  0.0         Mexican Restaurant      0.5625
Diner                          0.0         Breakfast Spot          0.5625
Donut Shop                     0.0         Turkish Restaurant      0.3750
Eastern European Restaurant    0.0         Japanese Restaurant     0.3750
Fish & Chips Shop              0.0         Bakery                  0.3750
Wings Joint                    0.0         Indian Restaurant       0.3125
Name: 2, dtype: float64                    Name: 5, dtype: float64
```

Figure 25 Top 10 venue categories on average

5. DISCUSSION

We summarize the features of each group as follows.

■ **Label 0**

Café dominates the list of the top 10 venue categories of this group. There are 64 cafés recommended in total and an average of 2.67 cafes near each university in this group. These two numbers are much higher than those of the rest of the groups. So the percentage of café must be an important factor that K-Means used when deciding which university should be assigned to cluster 0.

This group is also the only one that has Vegan Restaurant on the list. Good news for vegetarians. In addition, no fast-food was recommended. So basically this is a very healthy group.

The foreign restaurants recommended and are on the list include French Restaurant, Japanese Restaurant and Italian Restaurant.

The universities in this group include University of British Columbia, McGill University, Université de Montréal and other 21 universities.

■ **Label 1**

Restaurant is the top one on the list with an absolute advantage of 29 in total. Since we can infer nothing from it, we skip this category. Followed Restaurant are Sandwich Place and Café. Except that, the numbers of the rest of the categories in the list are quite even .

The foreign restaurants recommended and are on the list include Sushi Restaurant, New American Restaurant, Indian Restaurant and Italian Restaurant.

The universities in this group include University of Manitoba, University of Waterloo, York University and other 18 universities.

■ **Label 2**

Only Canadian Mennonite University was classified to this group with no recommendations available.

■ **Label 3**

The top one on the list of this group is still Restaurant, which tells us nothing. Other than that, this group got plenty of fast-food restaurants and nearly no foreign restaurants recommended. There are on average 1.4 fast-food restaurants, 0.8 sandwich place and 0.3 wings joint. In total,

there are about 40 venues in these kinds being recommended.

Greek Restaurant is the only foreign restaurants recommended and are on the list.

The universities in this group include University of Calgary, Université de Sherbrooke, and 14 other universities.

- **Label 4**

Only this group, group 2 and group 5 that Restaurant is not on the list of their top 10 common categories.

Similar to group 3, the universities in this group got a lot of fast-food recommendations. The difference is that Fast-Food Restaurant is the most common one in group 3 while in group 4 is Sandwich Place. There are on average 1.4 sandwich place, 1.2 fast-food restaurants and 0.5 Burger Joint. In total, there are about 37 venues in these kinds being recommended.

The foreign restaurants recommended and are on the list include Sushi Restaurant, American Restaurant and Italian Restaurant.

The universities in this group include University of Alberta, Simon Fraser University and 10 other universities.

- **Label 5**

What is special about this group is that it has the most diverse food supplier recommendations. The foreign restaurants recommended and are on the list include Sushi Restaurant, Italian Restaurant, Mexican Restaurant, Turkish Restaurant, Japanese Restaurant and Indian Restaurant, six in total.

Another feature of this group is that no one category can get an average number over one. Some actually there is not a venue category that can dominate the list of this group.

The universities in this group include Dalhousie University, McMaster University and 14 other universities.

6. CONCLUSION

Overall, it seems that finding delicious foods around the universities does not constitute a serious problem for students there. The surrounding of almost every university offers enough options to be considered.

Based on the result of our clustering, we could also see that the most common venue categories of each group are quite different. In general, venues that are classified into restaurants, fast-food restaurants, pizza places and cafés are most easily to be found nearby.

In addition, there are many foreign restaurants in the neighbourhood. Therefore, international students and domestic students who would like to have something different can all find good places to go.

There are some categories we did not mention in the last section like gastropub, bakery and breakfast spots are also great options.

So in summary, we hope this project could help you understand what the food map is like around the universities, and may inspire you on some similar or any other interesting ideas.

One kind warning is that the features of the groups are summarized based on groups in aggregate. Depending on the performance of the clustering algorithm, the individual university may not completely share the features of its group.