1. Introduction: Business Problem

Toronto is the most populous city in Canada, a multicultural city, and the Canadian financial and commercial centre. Several famous companies are located in Toronto. Moreover, Toronto also owns the top education resources in Canada. Toronto attracts more and more young people to pursue their dreams and start their new life.

Social activity is one of important part in young people's life. No matter how the society changes, the bar is always an option for social life. The big requirement will attract people to start their business of Bar. But before they can take action, it's important to figure out where they should open it? What should they consider when selecting a location? I hope to figure out whether the neighborhood of bars is an essential factor for the success of a bar with sound analysis.

Target Audience:

- The people that want to find a bar for social activity in Toronto.

- Investors who want to set up a Bar.

2. Data

The data used in this analysis is gathered from the following sources.
- **Location information of Neighborhoods in Toronto :**
The location information of Toronto neighborhoods is available from the assignment in this course, which includes the data of postal code, longitude and latitude.
- **Venue information of interest :**
Venue information of interest is obtained by using the Foursquare API based on the longitude and latitude coordinates of the neighborhood.

3. Methodology

I used BeautifulSoup to scrape boroughs from Wikipedia, and organized a table containing Postal Code and Neighborhoods information of Toronto City. I used Geospatial data from the course assignment to get the geological location of each community board.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
1	M7A	Downtown Toronto	Queen's Park , Ontario Provincial Government	43.662301	-79.389494
2	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
3	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
4	M4E	East Toronto	The Beaches	43.676357	-79.293031
5	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
6	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383
7	M6G	Downtown Toronto	Christie	43.669542	-79.422564
8	M5H	Downtown Toronto	Richmond , Adelaide , King	43.650571	-79.384568
9	M6H	West Toronto	Dufferin , Dovercourt Village	43.669005	-79.442259
10	M5J	Downtown Toronto	Harbourfront East , Union Station , Toronto Is...	43.640816	-79.381752
11	M6J	West Toronto	Little Portugal , Trinity	43.647927	-79.419750
12	M4K	East Toronto	The Danforth West , Riverdale	43.679557	-79.352188
13	M5K	Downtown Toronto	Toronto Dominion Centre , Design Exchange	43.647177	-79.381576

I utilized the **Foursquare API** to explore the boroughs and segment them for each borough from their given latitude and longitude information. The limit is 100 venues and the radius is 500 meters. The results are shown as below, adding venue id, venue name, category, latitude, and longitude information from **Foursquare API**.

It returns with 1622 records. I summarize venues by category. Among these 1622 records, **bar** counts 30. I utilized the **Foursquare API** again by Bar places ID to explore the detailed record of these bars including Rating, Price, Likes, Photos, Tips into a data frame.

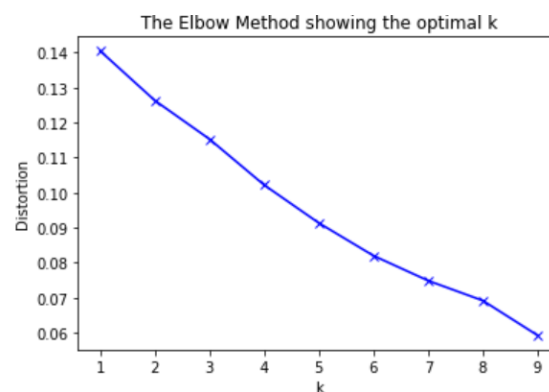|	Neighborhood	ID	Rating	Price	Likes	Photos	Tips	Latitude	Longitude
0	SUDS	4cc5e76091413704e955cc55	6.0	2	5	1	3	43.659880	-79.394712
1	Duke's Refresher + Bar	529fa9e111d2159d1f3caa0e	8.1	2	92	217	36	43.658980	-79.382949
2	Duke's Refresher + Bar	529fa9e111d2159d1f3caa0e	8.1	2	92	217	36	43.658980	-79.382949
3	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956
4	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706
5	The Greater Good Bar	5753753b498eeb535c53aed5	9.3	2	47	49	6	43.669409	-79.439267
6	Corks Beer & Wine Bar	4d0faed7d467236a7291bf4a	7.5	2	10	19	5	43.642493	-79.381540
7	The Rec Room	5952cf02829b0c59fa430f63	8.1	2	115	103	13	43.641110	-79.386763
8	Reposado	4ae662e8f964a520cfa621e3	8.5	2	86	162	45	43.647321	-79.420032
9	The Communist's Daughter	4adfd8c0f964a520bf7d21e3	8.4	1	68	40	23	43.649362	-79.420963
10	Dakota Tavern	4ad4c05df964a52067f620e3	8.3	1	74	139	33	43.649680	-79.420838
11	apt 200	55fe3af5498edd651ce3c385	8.3	2	41	5	6	43.644026	-79.420063
12	Bar Fancy	545d6eeb498ea42297425c1e	8.2	2	68	31	16	43.643734	-79.421326
13	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956
14	Pharmacy	50dfbed9e4b01ebc5ced2ae3	8.8	2	44	51	5	43.638090	-79.431810
15	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956
16	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706
17	The Roy Public House	4ad79986f964a5209d0c21e3	6.7	2	42	79	35	43.660452	-79.342994
18	Hole in the Wall	4baebdabf964a520ddd33be3	8.9	2	85	82	35	43.665296	-79.465118
19	Shoxs	4b26d1aff964a520698124e3	5.7	2	5	11	11	43.665353	-79.463563
20	The Local Pub and Restaurant	4ad4c05df964a52068f620e3	7.4	2	41	63	37	43.651017	-79.450911
21	Harbord House	4ade455bf964a520837421e3	8.8	2	52	79	29	43.662466	-79.405410
22	Charlie's Gallery	4af11491f964a520aee021e3	6.2	2	5	3	3	43.662810	-79.403822
23	A Dark Horse	4af6185af964a5201c0122e3	6.0	2	19	19	16	43.649533	-79.483056
24	Cold Tea	4e9a55b68231878c156f04bf	8.5	1	162	140	48	43.654193	-79.401075
25	Trinity Common	55175269498ef06e96eeeb98	8.0	2	84	113	13	43.656590	-79.402761
26	The Supermarket	4ad4c05df964a5206af620e3	5.8	2	40	118	31	43.656680	-79.402954
27	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706
28	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956
29	Loose Moose	4ae603a5f964a52010a421e3	7.8	2	272	515	109	43.645281	-79.383966

Then I tried to find correlations among these variables. As shown in the correlation matrix below, Likes, Photos, and Tips are highly correlated with each other. But these factors are not highly related to Rating. The possible explanation is that customers who may submit these feedbacks for some specific reasons but give lower ratings to the general performance might cause this low correlation. Therefore, Rating is chosen to represent the bar. Rating is somewhat correlated to Price but not significantly.

	Rating	Price	Likes	Photos	Tips
Rating	1.000000	-0.118195	0.284868	0.171478	0.106293
Price	-0.118195	1.000000	0.454389	0.452030	0.522452
Likes	0.284868	0.454389	1.000000	0.934885	0.940157
Photos	0.171478	0.452030	0.934885	1.000000	0.932678
Tips	0.106293	0.522452	0.940157	0.932678	1.000000

I utilized the **Foursquare API** centering these bars to explore their neighborhoods within radius of 500-meter.

|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue ID	Venue Latitude	Venue Longitude	Venue Category
0	SUDS	43.659880	-79.394712	Prenup Pub	54238a83498e5efd4a50435b	43.658882	-79.394828	Gastropub
1	SUDS	43.659880	-79.394712	Starbucks	576d30ed498ef2e95e26e6d0	43.659456	-79.390411	Coffee Shop
2	SUDS	43.659880	-79.394712	Kekou Gelato House	51e87a62498e8eea21f772d3	43.656099	-79.392715	Ice Cream Shop
3	SUDS	43.659880	-79.394712	McCafé	5b57b67fc53093003750b093	43.657943	-79.398993	Café
4	SUDS	43.659880	-79.394712	Vegetarian Haven	4aeb711ef964a52017c221e3	43.656016	-79.392758	Vegetarian / Vegan Restaurant
5	SUDS	43.659880	-79.394712	Queen's Park	4b9d206bf964a520e69136e3	43.663946	-79.392180	Park
6	SUDS	43.659880	-79.394712	Jimmy's Coffee	57fe5f64498e08c9fc55cb87	43.655827	-79.392042	Café
7	SUDS	43.659880	-79.394712	Hart House Theatre	4ce47a4bc850721e2b8bc38a	43.663571	-79.394616	Theater
8	SUDS	43.659880	-79.394712	Fat Bastard Burrito	55e1f58b498ed51280251aa4	43.658900	-79.394530	Burrito Place
9	SUDS	43.659880	-79.394712	The Clubhouse	51b8ebe8e4b051460c6dfa24	43.658246	-79.399894	Sandwich Place
10	SUDS	43.659880	-79.394712	Koh Lipe	5cb0d23175dcb7002cb5e3ad	43.655933	-79.393480	Thai Restaurant
11	SUDS	43.659880	-79.394712	A & C World	4dbc33b90cb691071c9c7d29	43.657409	-79.399847	Gaming Cafe
12	SUDS	43.659880	-79.394712	Sambuca Grill	4b758acdf964a520e0122ee3	43.656110	-79.392946	Italian Restaurant
13	SUDS	43.659880	-79.394712	Toronto Vegetarian Association	4b117fa4f964a520df7d23e3	43.655953	-79.392854	Office
14	SUDS	43.659880	-79.394712	Charidise	588541549343e011e4a22fbb	43.655985	-79.393233	Bubble Tea Shop
15	SUDS	43.659880	-79.394712	Mother's Dumplings	4b7716a9f964a520297c2ee3	43.657105	-79.399552	Dumpling Restaurant
16	SUDS	43.659880	-79.394712	Second Cup	4b72de05f964a520668c2de3	43.659018	-79.394152	Coffee Shop
17	SUDS	43.659880	-79.394712	Booster Juice	4baa6c4af964a520bb683ae3	43.658016	-79.399300	Juice Bar
18	SUDS	43.659880	-79.394712	Flock Rotisserie + Greens	59ea2931f96b2c78d41cc639	43.659167	-79.389475	Fried Chicken Joint
19	SUDS	43.659880	-79.394712	Subway	4bce296fcc8cd13aeddfc3cf	43.658683	-79.394896	Sandwich Place
20	SUDS	43.659880	-79.394712	Midi Bistro	4ba8f061f964a520e8fb39e3	43.655871	-79.392091	French Restaurant

According to venue categories and its appearance frequency surrounding each bar, I use k-means to cluster bar into several groups. The "elbow" method is used to select the optimal number of clusters by fitting the model with a range of values for k. In the visualizer k=3 is shown as the elbow.

The Elbow Method showing the optimal k

I merged cluster labels of each bar place with its geological location.

	Neighborhood	ID	Rating	Price	Likes	Photos	Tips	Latitude	Longitude	Cluster Labels
0	SUDS	4cc5e76091413704e955cc55	6.0	2	5	1	3	43.659880	-79.394712	2
1	Duke's Refresher + Bar	529fa9e111d2159d1f3caa0e	8.1	2	92	217	36	43.658980	-79.382949	2
2	Duke's Refresher + Bar	529fa9e111d2159d1f3caa0e	8.1	2	92	217	36	43.658980	-79.382949	2
3	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956	2
4	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706	2
5	The Greater Good Bar	5753753b498eeb535c53aed5	9.3	2	47	49	6	43.669409	-79.439267	1
6	Corks Beer & Wine Bar	4d0faed7d467236a7291bf4a	7.5	2	10	19	5	43.642493	-79.381540	2
7	The Rec Room	5952cf02829b0c59fa430f63	8.1	2	115	103	13	43.641110	-79.386763	2
8	Reposado	4ae662e8f964a520cfa621e3	8.5	2	86	162	45	43.647321	-79.420032	1
9	The Communist's Daughter	4adfd8c0f964a520bf7d21e3	8.4	1	68	40	23	43.649362	-79.420963	1
10	Dakota Tavern	4ad4c05df964a52067f620e3	8.3	1	74	139	33	43.649680	-79.420838	1
11	apt 200	55fe3af5498edd651ce3c385	8.3	2	41	5	6	43.644026	-79.420063	1
12	Bar Fancy	545d6eeb498ea42297425c1e	8.2	2	68	31	16	43.643734	-79.421326	1
13	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956	2
14	Pharmacy	50dfbed9e4b01ebc5ced2ae3	8.8	2	44	51	5	43.638090	-79.431810	1
15	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956	2
16	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706	2
17	The Roy Public House	4ad79986f964a5209d0c21e3	6.7	2	42	79	35	43.660452	-79.342994	1
18	Hole in the Wall	4baebdabf964a520ddd33be3	8.9	2	85	82	35	43.665296	-79.465118	0
19	Shoxs	4b26d1aff964a520698124e3	5.7	2	5	11	11	43.665353	-79.463563	0
20	The Local Pub and Restaurant	4ad4c05df964a52068f620e3	7.4	2	41	63	37	43.651017	-79.450911	0
21	Harbord House	4ade455bf964a520837421e3	8.8	2	52	79	29	43.662466	-79.405410	0
22	Charlie's Gallery	4af11491f964a520aee021e3	6.2	2	5	3	3	43.662810	-79.403822	0
23	A Dark Horse	4af6185af964a5201c0122e3	6.0	2	19	19	16	43.649533	-79.483056	0
24	Cold Tea	4e9a55b68231878c156f04bf	8.5	1	162	140	48	43.654193	-79.401075	0
25	Trinity Common	55175269498ef06e96eeeb98	8.0	2	84	113	13	43.656590	-79.402761	0
26	The Supermarket	4ad4c05df964a5206af620e3	5.8	2	40	118	31	43.656680	-79.402954	0
27	Earls Kitchen & Bar	4d2b615e342d6dcb2b8115cb	7.9	3	262	365	116	43.647946	-79.383706	2
28	Boxcar Social Temperance	5894c1f9266c1121f0a757d1	8.0	2	35	36	7	43.650557	-79.381956	2
29	Loose Moose	4ae603a5f964a52010a421e3	7.8	2	272	515	109	43.645281	-79.383966	2

Then I used folium to visualize the distribution of these bar in Toronto as below:
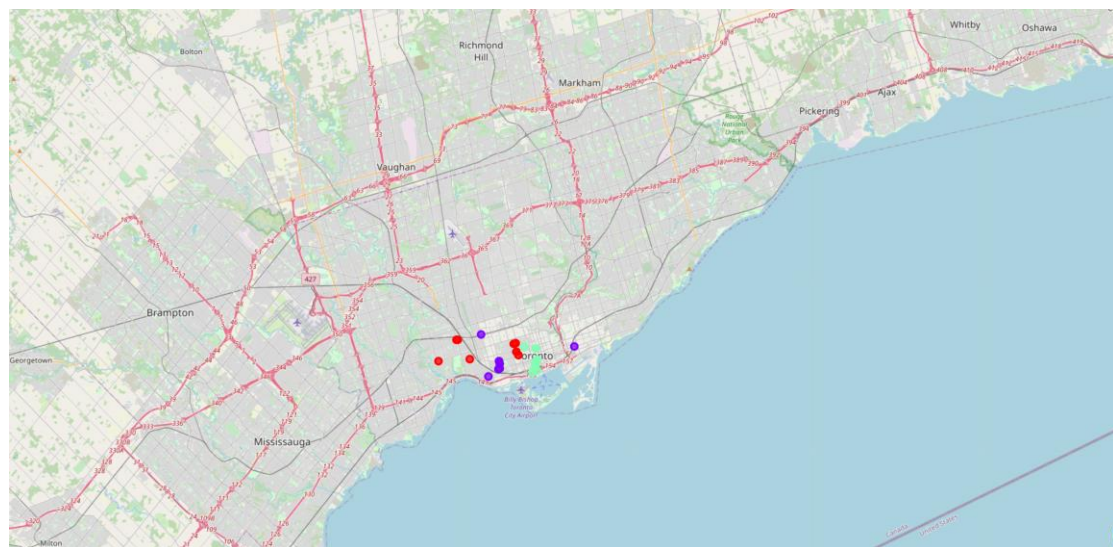
4. Results

Let's see if there are differences among performances of these bars with different clustering labels.

```
Cluster Labels
0   7.255556
1   8.312500
2   7.792308
Name: Rating, dtype: float64
```

Based on mean value of rating of each cluster, it seems that there is a difference. Let's run the one-way ANOVA analysis to check whether the difference is significant or not.

```
Multiple Comparison of Means - Tukey HSD,FWER=0.05
========================================
group1 group2 meandiff  lower   upper  reject
----------------------------------------
  0     1    1.0569  -0.0328 2.1467 False
  0     2    0.5368  -0.4358 1.5093 False
  1     2   -0.5202   -1.528 0.4876 False
----------------------------------------
```

As shown in the table, the performances of three types of bars are significantly different. Places with cluster labels 1 perform best while those with label 0 perform worst. Therefore, the different neighborhoods might affect the impression of the customers in the bar. Bar investors shall search for a similar location to start their business. In order to investigate what is special in these locations around bars, I selected bars with label 1, and sort venue types by average numbers of venues around bars in a descending way.

	Cluster 1
Venue Category	
Bar	3.875
Restaurant	2.500
Caйй	2.125
Coffee Shop	2.000
Vegetarian / Vegan Restaurant	1.250
Asian Restaurant	1.125
Ice Cream Shop	1.125
Brewery	1.125
Italian Restaurant	1.125
Cocktail Bar	1.000
Japanese Restaurant	1.000
Sandwich Place	1.000
Men's Store	1.000

Bar is usually a relaxed place before or after dinner. It explains why places with higher ratings gather around other places for eat and drink. Besides, people care much about the atmosphere of relaxing and freedom when they are at a bar. Hence, bars usually gather together also. Little Portugal, Trinity is one place of this kind.

5. Conclusion

As a result, for those who want to find a bar, they should consider locations with high density of bars and close to other restaurants or relaxing places serving light meals.

For those aiming to invest a bar, they should find places which have relaxing and free atmosphere.

6. Discussion

As a recommendation to those who plan to operate a bar, location selection is only one fundamental problem to think over. It can't solve the problem of whether the type of a bar is popular or not. And it doesn't consider how many customers will visit every day. For location proposal, it offers an opportunity analysis but lacks risk analysis, e.g. the competition in that area.

This project demonstrates the relations between location and ratings, but ratings can't show the whole picture. The suggestion is relatively narrow. More practical and profitable ideas need to be considered. Therefore, the relationship between financial performance of bar and customers' reactions is valuable to be analyzed.

With all of these analyses done, the report can finally become a constructive proposal for bar investors in the real business world.