

# Capstone Project Report

## Project: Exploring Toronto for Opening a New Chain of Gourmet Pizza Place with Bar

### 1. Introduction:

#### 1.1 Description of the Problem, Background

For the Capstone project, a hypothetical scenario is being created for an Italian restaurateur who wants to open a chain of a gourmet pizza place with a bar in Toronto in multiple locations. Pizza is probably the most popular food among Canadians. There is already plethora of pizza places that offer bar facility as well. The idea behind this project is that there may not be enough pizza places in Toronto that offer the facility of a bar as well. Therefore, a pizza place with a bar might present a great opportunity for this restaurateur who is, now, living in Canada. Since there are pizza places everywhere and bars are also located all across Toronto area, this entrepreneur is thinking of opening this restaurant in locations or neighborhoods where there are less or no pizza places and bars. With the purpose in mind, finding the location to open such a chain restaurant is one of the most important decisions for this entrepreneur.

The objective of this capstone project is to find the most suitable locations for the entrepreneur to open this chain of gourmet pizza place with a bar in Toronto, Canada. By using data science methodology and machine learning (e.g. clustering), this project aims to provide solutions to answer the business question: What are the best possible locations to open a chain of gourmet pizza places with the facility of a bar.

##### 1.1.2 Target Audience

The entrepreneurs who want to find the most suitable location to open a pizza place with a bar.

### 2. Data

#### 2.1 Required Data

In order to answer the questions for the project, the following data is required:

- A List of neighborhoods in Toronto to categorize Toronto into smaller areas.
- Geographical locations (Latitudes and Longitudes) of these neighborhoods to know the venues in the neighbourhoods.
- The distribution of population by different ethnicities to identify the neighborhoods which are densely populated with Italians as it might be helpful in identifying the suitable neighborhood to open a gourmet pizza place with a bar because Italians, specifically, might be more interested in a gourmet pizza than other ethnicities.
- Data of the venues (pizza places and bars) in Toronto which will help me determine the places with the least number of pizza places and bars by using data science methodology.

#### 2.2 Data Sources

The following data sources were used for this project.

- “[https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M))” Wiki page to get the information about the neighborhoods in Toronto.

- “[https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)” csv file to get the geographical coordinates (longitudes and latitudes) of the neighborhoods.
- A wiki page to get information about the distribution of population by their ethnicity. Using this page I’m going to identify the neighborhoods which are densely populated with Italians. “Demographics of Toronto” ([https://en.m.wikipedia.org/wiki/Demographics\\_of\\_Toronto#Ethnic\\_diversity](https://en.m.wikipedia.org/wiki/Demographics_of_Toronto#Ethnic_diversity))
- Foursquare’s API ((<https://developer.foursquare.com/docs>)) to get location and other information about various venues in Toronto Foursquare API

### 3. Methodology

The first step was data wrangling/cleaning. “List of Postal code of Canada: M” wikipedia page was scraped in order to obtain the data about the Neighborhoods in Toronto. Then, geographical coordinates were added to these neighborhoods. The data present in Geospatial Data csv file was used for this purpose. Another factor that can help in deciding which locations will be most suitable options to open a gourmet pizza place with a bar is the distribution of population based on the ethnic diversity for each neighborhood. It might be helpful in identifying the suitable neighborhood to open a gourmet pizza place with a bar because Italians, specifically, might be more interested in a gourmet pizza than other ethnicities. The Wikipedia page, “Demographics of Toronto” was scraped in order to obtain the data about the Toronto neighborhoods and most populated neighbourhoods with Italian ethnicity were identified. Foursquare API is very useful online tool and is used by many soft wares and applications like Uber etc. In this project, it was used to retrieve information about the venues present in the neighborhoods of Toronto. Foursquare API was used to pull the list of top 100 venues within 1 Km radius. Foursquare developer account was created in order to obtain account ID and API key to pull the data. From Foursquare, the names, categories, latitude and longitude of the venues were pulled. From this data, unique categories of the venues were checked. Then, each neighborhood was analysed by grouping the rows by neighborhood and taking the mean on the frequency of occurrence of each venue category to prepare the data for clustering later on.

Pizza place and bar were found as two unique categories in the venues in the neighbourhoods. The idea is to find locations or neighbourhoods that lack both pizza places and bars. For data exploration purposes, the number of pizza places and bars were plotted in each neighbourhood and neighbourhoods were identified with the most number of bars and pizza places as shown in the graphs below in Figures 1 and 2.

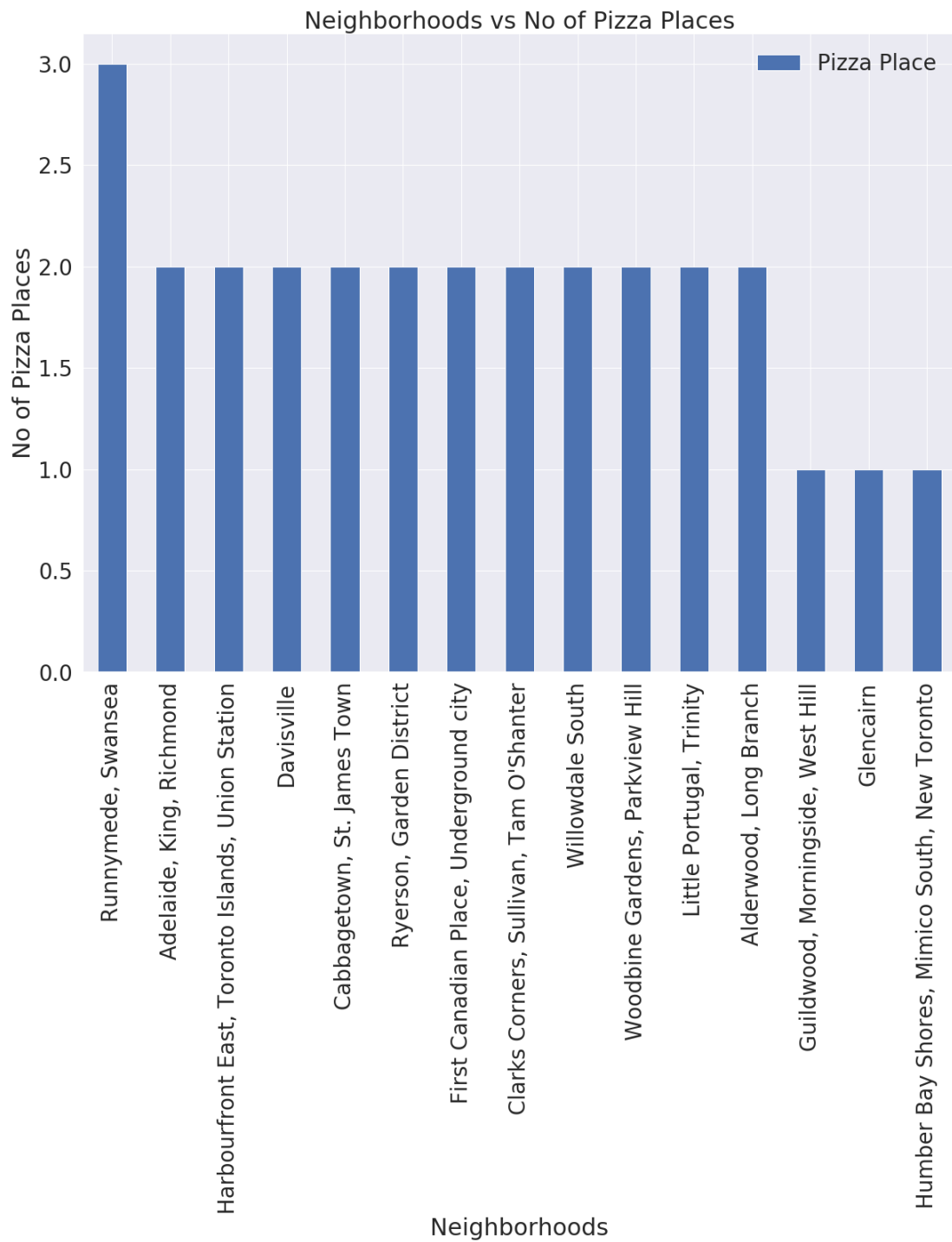


Figure 1 Number of pizza places in different neighbourhoods

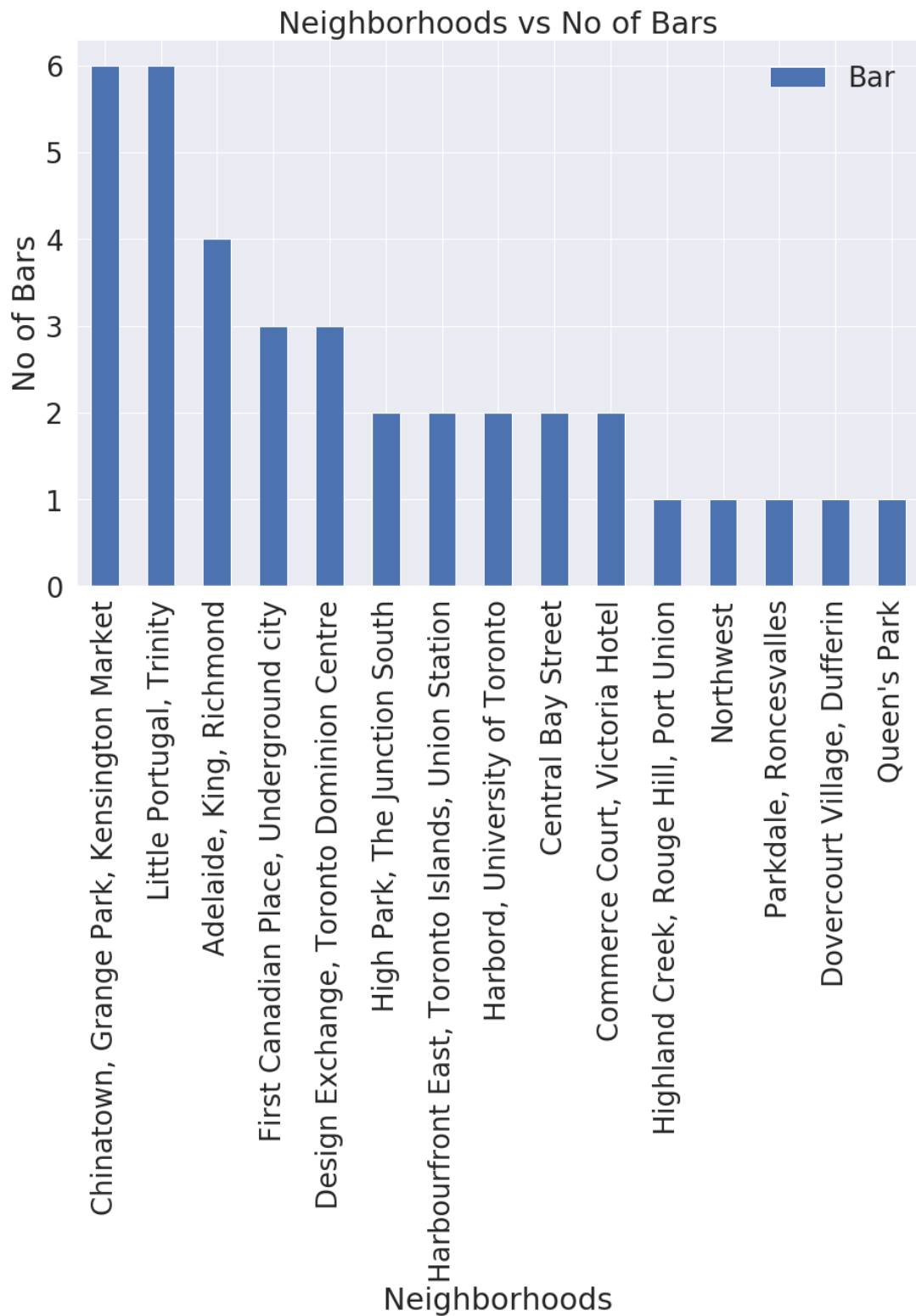


Figure 2 Number of bars in different neighbourhoods

In order to find out the distribution of bars and pizza places in different boroughs, violin plots were also plotted as shown in figures 3 and 4. Figure 3 shows that Scarborough, downtown Toronto and East Toronto have very scarce distribution of pizza places.

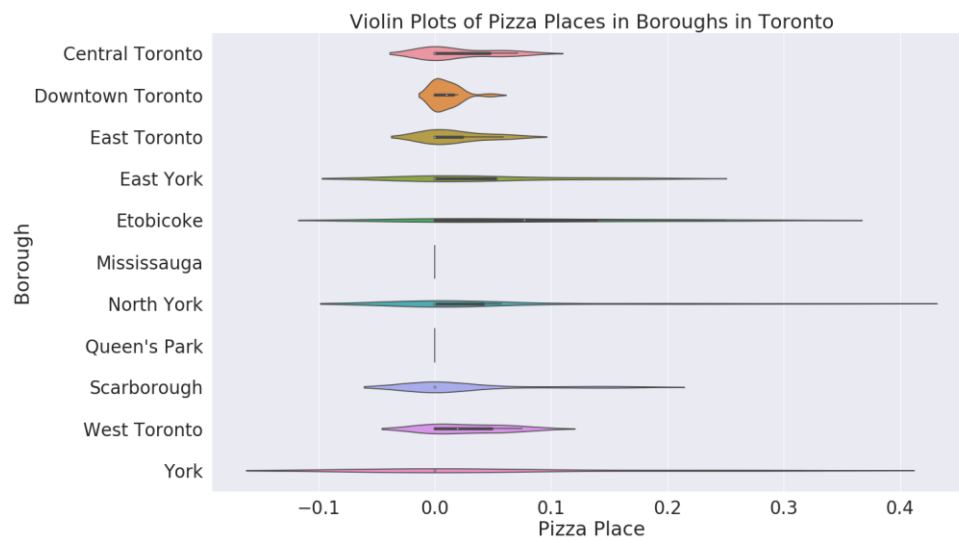


Figure 3 Pizza places distribution in different boroughs

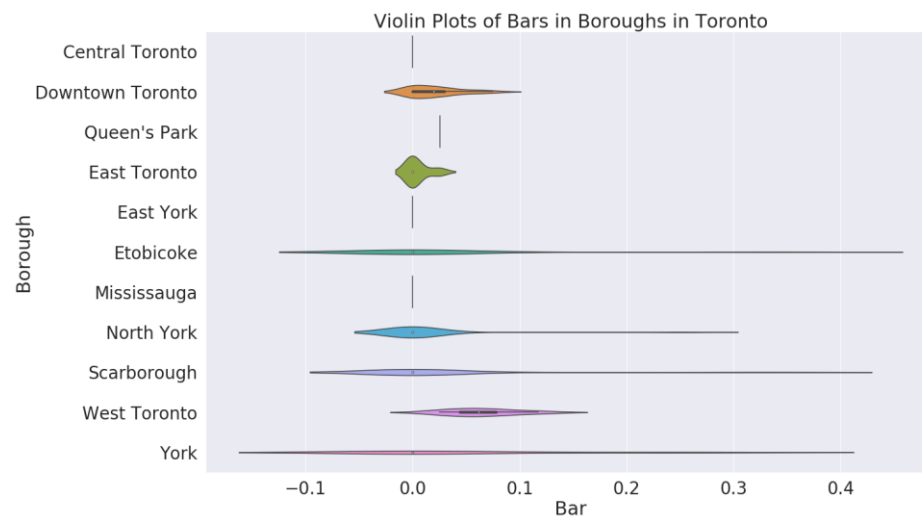


Figure 4 Bars distribution in different boroughs

Then Italian population was looked at in each neighbourhood in order to find a relationship between number pizza places in a neighbourhood and Italian population which needs more investigation.

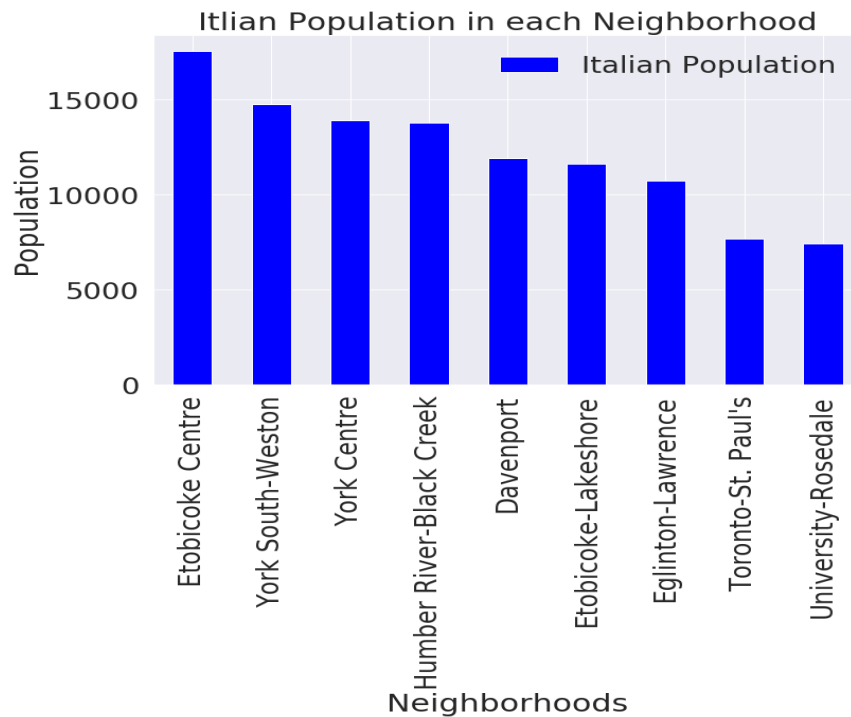


Figure 5 The distribution of Italian population in different neighbourhoods

Keeping in mind that idea is to find places with least number of bars and pizza places and the data is very scattered and varies from location to location, it will be good idea to group places with different density of bars and pizza places. It will make it easier to spot places with the least or minimum number of pizza places and bars. K-means clustering provides this ability to divide data into groups of different characteristics. K-means clustering algorithm works by identifying k number of centeriods, and allocating every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the easiest and most popular unsupervised machine learning algorithms. It is very well suited for this project as well. First data was grouped together based on number of pizza places. By using and elbow visualizer, as shown below, K = 7 was picked.

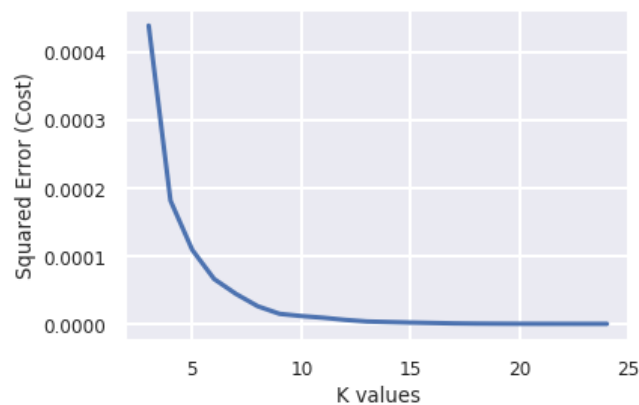


Figure 6 Squared error for different K values

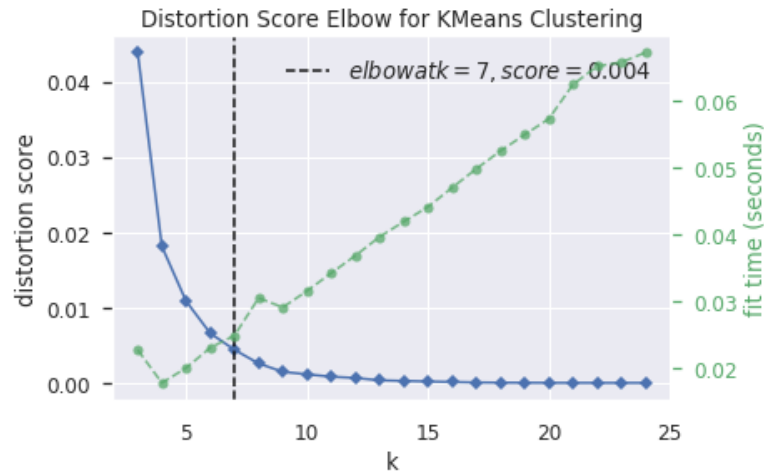


Figure 7 Elbow visualizer to determine K-value

The K-means clustering algorithm was run and those clusters were mapped. The same analysis was performed on number of bars by using k-means clustering (first by finding suitable number for K and then running the algorithm). The clusters were mapped as well. After analyzing clusters locations were identified for the least number of bars and pizza places and recommendations were made.

## 4. Results

The K-means clustering algorithm was run for pizza places in different neighbourhoods and those clusters were mapped as below.

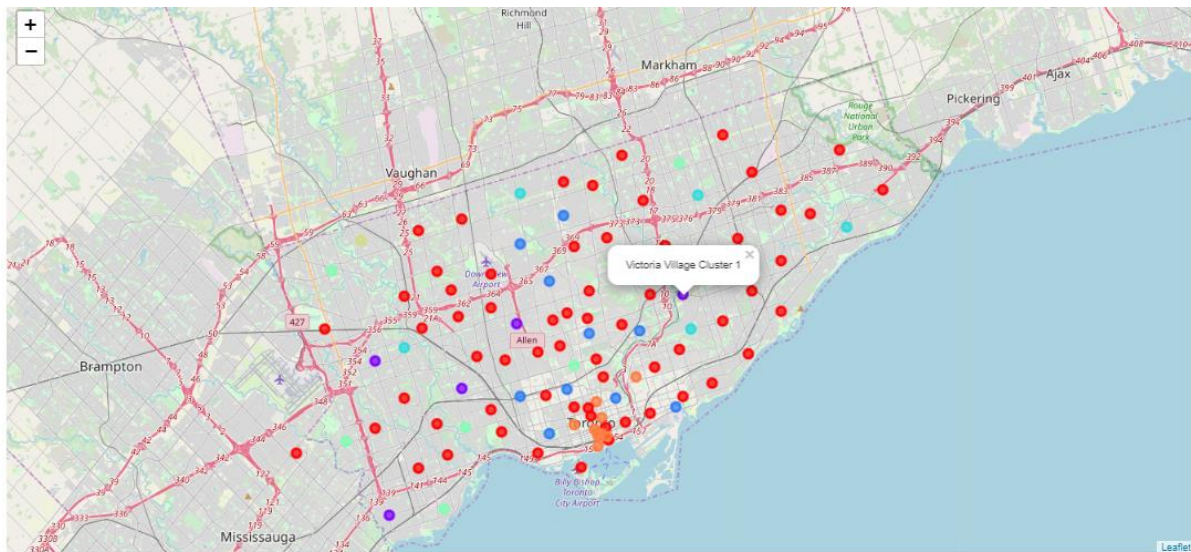


Figure 8 K-means clusters map for pizza places

The red dots represent cluster 0 and purple dots represent cluster 1. The clusters were examined for pizza places distribution. The first cluster is shown below.

	Borough	Postcode	Neighborhood	Latitude	Longitude	Cluster Labels	Pizza Place
0	Central Toronto	M4N	Lawrence Park	43.728020	-79.388790	0.0	0.0
1	Central Toronto	M4P	Davisville North	43.712751	-79.390197	0.0	0.0
2	Central Toronto	M4R	North Toronto West	43.715383	-79.405878	0.0	0.0
4	Central Toronto	M4T	Moore Park, Summerhill East	43.689574	-79.383160	0.0	0.0
6	Central Toronto	M5N	Roselawn	43.711695	-79.416936	0.0	0.0
7	Central Toronto	M5P	Forest Hill North, Forest Hill West	43.686948	-79.411307	0.0	0.0
9	Downtown Toronto	M4W	Rosedale	43.679583	-79.377529	0.0	0.0
12	Downtown Toronto	M5A	Harbourfront	43.654280	-79.380836	0.0	0.0
14	Downtown Toronto	M5C	St. James Town	43.651494	-79.375418	0.0	0.0
15	Downtown Toronto	M5E	Berczy Park	43.644771	-79.373306	0.0	0.0
16	Downtown Toronto	M5G	Central Bay Street	43.657952	-79.387383	0.0	0.0
21	Downtown Toronto	M5S	Harbord, University of Toronto	43.662896	-79.400049	0.0	0.0
23	Downtown Toronto	M5V	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0.0	0.0
26	Downtown Toronto	M6G	Christie	43.669542	-79.422564	0.0	0.0
27	Downtown Toronto	M7A	Queen's Park	43.662301	-79.389494	0.0	0.0
28	East Toronto	M4E	The Beaches	43.676357	-79.293031	0.0	0.0
30	East Toronto	M4L	The Beaches West, India Bazaar	43.668999	-79.315572	0.0	0.0
31	East Toronto	M4M	Studio District	43.659528	-79.340823	0.0	0.0
34	East York	M4C	Woodbine Heights	43.695344	-79.318389	0.0	0.0
35	East York	M4G	Leaside	43.709060	-79.363452	0.0	0.0
37	East York	M4J	East Toronto	43.685347	-79.338106	0.0	0.0
40	Etobicoke	M8X	The Kingsway, Montgomery Road, Old Mill North	43.653654	-79.506944	0.0	0.0
41	Etobicoke	M8Y	Humber Bay, King's Mill Park, Kingsway Park So...	43.636258	-79.498509	0.0	0.0
42	Etobicoke	M8Z	Kingsway Park South West, Mimico NW, The Queen...	43.628841	-79.520999	0.0	0.0
43	Etobicoke	M8B	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724	0.0	0.0
48	Etobicoke	M9W	Northwest	43.706748	-79.594054	0.0	0.0
49	Mississauga	M7R	Canada Post Gateway Processing Centre	43.636986	-79.615819	0.0	0.0
50	North York	M2H	Hillcrest Village	43.803762	-79.363452	0.0	0.0
51	North York	M2J	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0.0	0.0
60	North York	M3C	Flemington Park, Don Mills South	43.725900	-79.340823	0.0	0.0
62	North York	M3J	Northwood Park, York University	43.767980	-79.487282	0.0	0.0
63	North York	M3K	CFB Toronto, Downsview East	43.737473	-79.464783	0.0	0.0
64	North York	M3L	Downsview West	43.739015	-79.506944	0.0	0.0
65	North York	M3M	Downsview Central	43.728496	-79.495997	0.0	0.0
66	North York	M3N	Downsview Northwest	43.761631	-79.520999	0.0	0.0
69	North York	M8A	Lawrence Heights, Lawrence Manor	43.718518	-79.464783	0.0	0.0
71	North York	M8L	Downsview, North Park, Upwood Park	43.713756	-79.490074	0.0	0.0
73	North York	M9M	Emery, Humberlea	43.724786	-79.532242	0.0	0.0
74	Queen's Park	M9A	Queen's Park	43.667856	-79.532242	0.0	0.0
75	Scarborough	M1B	Rouge, Malvern	43.806686	-79.194353	0.0	0.0
76	Scarborough	M1C	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0.0	0.0
78	Scarborough	M1G	Woburn	43.770992	-79.216917	0.0	0.0
79	Scarborough	M1H	Cedarbrae	43.773136	-79.239476	0.0	0.0
80	Scarborough	M1J	Scarborough Village	43.744734	-79.239476	0.0	0.0
81	Scarborough	M1K	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	0.0	0.0
82	Scarborough	M1L	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	0.0	0.0
83	Scarborough	M1M	Cliffcrest, Cliffside, Scarborough Village West	43.716316	-79.239476	0.0	0.0
84	Scarborough	M1N	Birch Cliff, Cliffside West	43.692657	-79.264848	0.0	0.0
85	Scarborough	M1P	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	0.0	0.0
86	Scarborough	M1R	Maryvale, Wexford	43.750072	-79.295849	0.0	0.0
87	Scarborough	M1S	Agincourt	43.794200	-79.262029	0.0	0.0
89	Scarborough	M1V	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	0.0	0.0
93	West Toronto	M8K	Brockton, Exhibition Place, Parkdale Village	43.636847	-79.428191	0.0	0.0
94	West Toronto	M8P	High Park, The Junction South	43.661608	-79.484763	0.0	0.0
95	West Toronto	M8R	Parkdale, Roncesvalles	43.648980	-79.456325	0.0	0.0
97	York	M6C	Humewood-Cedarvale	43.693781	-79.428191	0.0	0.0
98	York	M6E	Caledonia-Fairbanks	43.689026	-79.453512	0.0	0.0
99	York	M6M	Del Ray, Keelelale, Mount Dennis, Silverthorn	43.691116	-79.476013	0.0	0.0
101	York	M9N	Weston	43.706876	-79.518188	0.0	0.0

Figure 9 Cluster 0 for pizza places in Toronto neighbourhoods



The same analysis was performed on number of bars by using k-means clustering (first by finding suitable number for K and then running the algorithm). The clusters were mapped as given below.

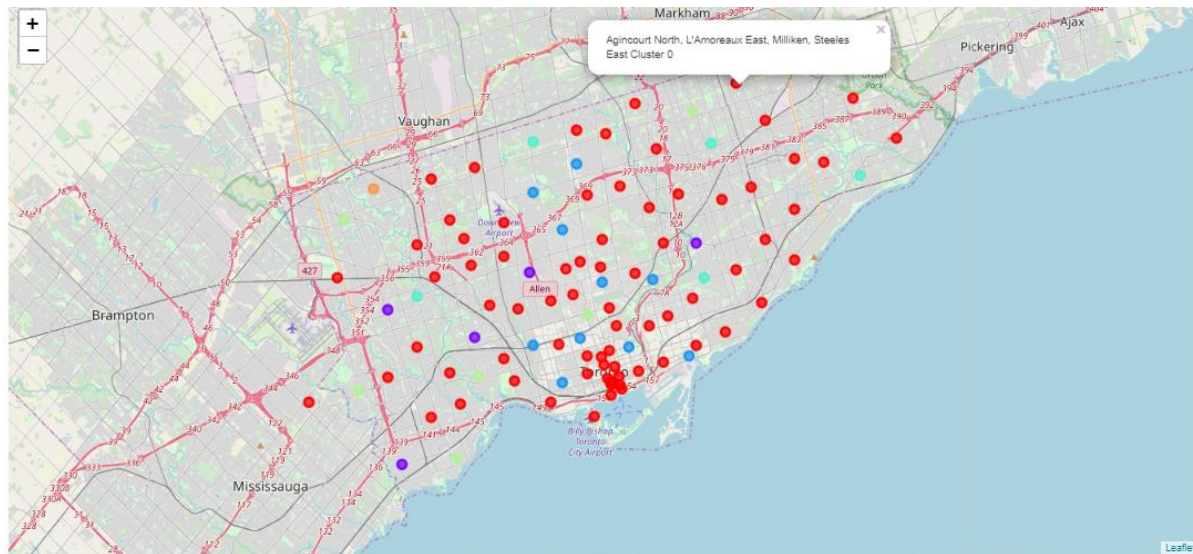


Figure 10 K-means clusters map for bars

Again, the red dots represent cluster 0 and purple dots represent cluster 1. The clusters were examined for bar places distribution. The cluster zero is shown below.

	Borough	Postcode	Neighborhood	Latitude	Longitude	Cluster Labels	Pizza Place	Bar
0	Central Toronto	M4N	Lawrence Park	43.728020	-79.388790	0.0	0.0	0.000000
1	Central Toronto	M4P	Davisville North	43.712751	-79.390197	0.0	0.0	0.000000
2	Central Toronto	M4R	North Toronto West	43.715383	-79.405678	0.0	0.0	0.000000
4	Central Toronto	M4T	Moore Park, Summerhill East	43.686574	-79.383160	0.0	0.0	0.000000
6	Central Toronto	M5N	Roselawn	43.711695	-79.416936	0.0	0.0	0.000000
7	Central Toronto	M5P	Forest Hill North, Forest Hill West	43.696948	-79.411307	0.0	0.0	0.000000
9	Downtown Toronto	M4W	Rosedale	43.679583	-79.377529	0.0	0.0	0.000000
12	Downtown Toronto	M5A	Harbourfront	43.654280	-79.380636	0.0	0.0	0.000000
14	Downtown Toronto	M5C	St. James Town	43.651484	-79.375418	0.0	0.0	0.000000
15	Downtown Toronto	M5E	Berczy Park	43.644771	-79.373306	0.0	0.0	0.000000
16	Downtown Toronto	M5G	Central Bay Street	43.657952	-79.387383	0.0	0.0	0.023810
21	Downtown Toronto	M5S	Harbord, University of Toronto	43.662896	-79.400049	0.0	0.0	0.055556
23	Downtown Toronto	M5V	CN Tower, Bathurst Quay, Island airport, Harbo...	43.628947	-79.394420	0.0	0.0	0.062500
26	Downtown Toronto	M6G	Christie	43.666542	-79.422584	0.0	0.0	0.000000
27	Downtown Toronto	M7A	Queen's Park	43.662301	-79.389484	0.0	0.0	0.025641
28	East Toronto	M4E	The Beaches	43.676357	-79.293031	0.0	0.0	0.000000
30	East Toronto	M4L	The Beaches West, India Bazaar	43.669999	-79.315572	0.0	0.0	0.000000
31	East Toronto	M4M	Studio District	43.659526	-79.340923	0.0	0.0	0.024390
34	East York	M4C	Woodbine Heights	43.695344	-79.318389	0.0	0.0	0.000000
35	East York	M4G	Leaside	43.709080	-79.363452	0.0	0.0	0.000000
37	East York	M4J	East Toronto	43.685347	-79.338106	0.0	0.0	0.000000
40	Etobicoke	M8X	The Kingsway, Montgomery Road, Old Mill North	43.653854	-79.506944	0.0	0.0	0.000000
41	Etobicoke	M8Y	Humber Bay, King's Mill Park, Kingsway Park So...	43.636258	-79.498509	0.0	0.0	0.000000
42	Etobicoke	M8Z	Kingsway Park South West, Mimico NW, The Queen...	43.628841	-79.520999	0.0	0.0	0.000000
43	Etobicoke	M9B	Cloverdale, Islington, Martin Grove, Princess ...	43.650943	-79.554724	0.0	0.0	0.000000
48	Etobicoke	M9W	Northwest	43.708748	-79.594054	0.0	0.0	0.333333
49	Mississauga	M7R	Canada Post Gateway Processing Centre	43.636986	-79.615819	0.0	0.0	0.000000
50	North York	M2H	Hillcrest Village	43.803782	-79.363452	0.0	0.0	0.000000
51	North York	M2J	Fairview, Henry Farm, Oriole	43.778517	-79.346556	0.0	0.0	0.000000

60	North York	M3C	Flemington Park, Don Mills South	43.725900	-79.340923	0.0	0.0	0.000000
62	North York	M3J	Northwood Park, York University	43.767980	-79.487282	0.0	0.0	0.250000
63	North York	M3K	CFB Toronto, Downsview East	43.737473	-79.464763	0.0	0.0	0.000000
64	North York	M3L	Downsview West	43.739015	-79.506944	0.0	0.0	0.000000
65	North York	M3M	Downsview Central	43.728498	-79.495997	0.0	0.0	0.000000
66	North York	M3N	Downsview Northwest	43.761631	-79.520999	0.0	0.0	0.000000
69	North York	M6A	Lawrence Heights, Lawrence Manor	43.718518	-79.464763	0.0	0.0	0.000000
71	North York	M6L	Downsview, North Park, Upwood Park	43.713756	-79.490074	0.0	0.0	0.000000
73	North York	M9M	Emery, Humberlea	43.724766	-79.532242	0.0	0.0	0.000000
74	Queen's Park	M9A	Queen's Park	43.667856	-79.532242	0.0	0.0	0.025841
75	Scarborough	M1B	Rouge, Malvern	43.806686	-79.194353	0.0	0.0	0.000000
76	Scarborough	M1C	Highland Creek, Rouge Hill, Port Union	43.784535	-79.160497	0.0	0.0	0.333333
78	Scarborough	M1G	Woburn	43.770992	-79.216917	0.0	0.0	0.000000
79	Scarborough	M1H	Cedarbrae	43.773136	-79.239476	0.0	0.0	0.000000
80	Scarborough	M1J	Scarborough Village	43.744734	-79.239476	0.0	0.0	0.000000
81	Scarborough	M1K	East Birchmount Park, Ionview, Kennedy Park	43.727929	-79.262029	0.0	0.0	0.000000
82	Scarborough	M1L	Clairlea, Golden Mile, Oakridge	43.711112	-79.284577	0.0	0.0	0.000000
83	Scarborough	M1M	Cliffcrest, Cliffside, Scarborough Village West	43.718316	-79.239476	0.0	0.0	0.000000
84	Scarborough	M1N	Birch Cliff, Cliffside West	43.692657	-79.264848	0.0	0.0	0.000000
85	Scarborough	M1P	Dorset Park, Scarborough Town Centre, Wexford ...	43.757410	-79.273304	0.0	0.0	0.000000
86	Scarborough	M1R	Maryvale, Wexford	43.750072	-79.295849	0.0	0.0	0.000000
87	Scarborough	M1S	Agincourt	43.794200	-79.262029	0.0	0.0	0.000000
89	Scarborough	M1V	Agincourt North, L'Amoreaux East, Milliken, St...	43.815252	-79.284577	0.0	0.0	0.000000
93	West Toronto	M6K	Brookton, Exhibition Place, Parkdale Village	43.636847	-79.428191	0.0	0.0	0.041667
94	West Toronto	M6P	High Park, The Junction South	43.661608	-79.464763	0.0	0.0	0.080000
95	West Toronto	M6R	Parkdale, Roncesvalles	43.648980	-79.456325	0.0	0.0	0.071429
97	York	M6C	Humewood-Cedarvale	43.693781	-79.428191	0.0	0.0	0.000000
98	York	M6E	Caledonia-Fairbanks	43.689026	-79.453512	0.0	0.0	0.000000
99	York	M6M	Del Ray, Keeleleale, Mount Dennis, Silverthorn	43.691116	-79.476013	0.0	0.0	0.250000
101	York	M9N	Weston	43.706876	-79.518188	0.0	0.0	0.000000

Figure 11 Cluster 0 for bars in Toronto neighbourhoods

The cluster 1 and the other clusters did not show the match of zero distribution of both bar and pizza places e.g. the cluster one or cluster two as shown below.

	Borough	Postcode	Neighborhood	Latitude	Longitude	Cluster Labels	Pizza Place	Bar
39	Etobicoke	M8W	Alderwood, Long Branch	43.602414	-79.543484	1.0	0.222222	0.0
46	Etobicoke	M9R	Kingsview Village, Martin Grove Gardens, Richv...	43.688905	-79.554724	1.0	0.250000	0.0
67	North York	M4A	Victoria Village	43.725882	-79.315572	1.0	0.200000	0.0
70	North York	M6B	Glencairn	43.709577	-79.445073	1.0	0.250000	0.0
100	York	M6N	The Junction North, Runnymede	43.673185	-79.487262	1.0	0.250000	0.0

Figure 12 Cluster 1 for bars in Toronto neighbourhoods

	Borough	Postcode	Neighborhood	Latitude	Longitude	Cluster Labels	Pizza Place	Bar
3	Central Toronto	M4S	Davisville	43.704324	-79.388790	2.0	0.057143	0.000000
8	Central Toronto	M5R	The Annex, North Midtown, Yorkville	43.672710	-79.405678	2.0	0.047619	0.000000
10	Downtown Toronto	M4X	Cabbagetown, St. James Town	43.667967	-79.367675	2.0	0.047619	0.000000
32	East Toronto	M7Y	Business Reply Mail Processing Centre 969 Eastern	43.662744	-79.321558	2.0	0.058824	0.000000
36	East York	M4H	Thorncliffe Park	43.705369	-79.349372	2.0	0.052632	0.000000
55	North York	M2N	Willowdale South	43.770120	-79.408493	2.0	0.057143	0.000000
61	North York	M3H	Bathurst Manor, Downsview North, Wilson Heights	43.754328	-79.442259	2.0	0.047619	0.000000
68	North York	M5M	Bedford Park, Lawrence Manor East	43.733283	-79.419750	2.0	0.040000	0.000000
91	West Toronto	M6H	Dovercourt Village, Dufferin	43.669005	-79.442259	2.0	0.052632	0.052632
92	West Toronto	M6J	Little Portugal, Trinity	43.647927	-79.419750	2.0	0.039216	0.117647

Figure 13 Cluster 2 for bars in Toronto neighbourhoods

## **5. Discussion**

Based on Figure 8 (K-means clusters map for pizza places) and Figure 9 (Cluster 0 for pizza places in Toronto neighbourhoods), it can be observed that there are a lot of neighbourhoods in Scarborough, North York, Downtown Toronto and Central Toronto that lack pizza places. Also, Figure 10 (K-means clusters map for bars) and Figure 11 (Cluster 0 for bars in Toronto neighbourhoods) indicate that Central Toronto, North York and Scarborough lack bars in a lot of neighbourhoods. Since, Central Toronto, North York and Scarborough lack both pizza places and bars, the neighbourhoods identified in Figure 10 will be most suitable for opening a chain of a gourmet pizza place with a bar.

## **Future Work**

In this project, existence of bars and pizza places was the only factor that was considered to make a decision to open a chain of gourmet pizza place with a bar at multiple locations. There are many other factors that can affect this decision such as population, income per capita etc. Due to limited time and scope of this project, these factors were not considered. However, future work should also consider these factors in the decision making of opening a chain of gourmet pizza places with bar.

## **6. Conclusion**

In this project, firstly, the business problem was identified. Then, required data to solve the business problem was specified. The data was extracted from the identified sources and prepared for analysis. Exploratory data analysis was performed to have rough idea of tackling the business problem. Lastly, Machine learning i.e. k-means clustering was used to make decision to solve the business problem i.e. the neighbourhoods ideal for opening chain of gourmet pizza place with bar.