The Battle of the Neighbourhoods

Researching Toronto neighbourhoods to start a new private school

This report serves as a culmination of the IBM Data Science Professional Certificate course on Coursera. In this document, concepts and methodologies related to the entire course will be employed to solve a business problem.

1 Introduction

Toronto is the largest city in the most populous province of Canada. It is home to a large population of immigrants who come from all the four corners of the world, in addition to a large indigenous population. This leads to many differing demands for food choices, retail experiences and schooling needs.

The education system in Canada is one of the world's best but even the public education system cannot keep up with the myriad tastes, standards and expectations of parents and caregivers. This creates an opportunity to fill a demand to provide quality education catered to the students needs by way of private education.

2 Problem Statement

The large multicultural population in Toronto requires a more flexible and varied education system that goes above and beyond the one-size fits all public education system. This creates a business need for schools that offer more flexibilities in terms of subject choices and curriculum.

This proposal addresses this by proposing the setting up of a new private school. As many private schools already exist in the city, we shall use data science and machine learning to help us locate candidates among Toronto neighbourhoods for the new school.

3 Selection Criteria

Several factors will be considered when exploring possible neighbourhoods to place this new school.

1) Population Density

A new school will require a certain minimum number of students to make it a viable investment. We will obtain the population data from Statistics Canada to help us decide on which areas to include or exclude.

2) Number of existing private schools

Existing private schools will provide competition and affect the number of students and revenue to a new school. Thus, neighbourhoods with many private schools will be excluded.

4 Assumptions

The following assumptions are taken when considering the data available and for reasons of expediency.

1) Religious and non-vocational schools will be considered public schools.
2) The number of public schools is assumed to have negligible impact (as they cater to different segments of the population).
3) The makeup of the population in the neighbourhood (race, religion, social standing etc.) will not be a factor.

4) The population breakdown across ages will be assumed to be fairly consistent across all neighbourhoods

5 Target Audience

This project is aimed towards educationists and entrepreneurs who want to start a new private school to cater to the diverse needs of the population of Toronto.

6 Data Sources

Data from various sources will be used in our analysis.

1) Wikipedia page of Toronto Neighborhoods.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This gives us information on the boroughs, neighbourhoods and the postal codes. The postal codes, via their geolocation data, will allow us to search Foursquare for information on private schools in a particular neighbourhood.

2) Geolocation data on Toronto neighbourhoods.

http://cocl.us/Geospatial_data

This link contains a CSV file with the latitude and longitude information of Toronto neighbourhoods. The data is combined with the postal code information from the Wikipedia page to get the geolocation of each neighbourhood;

3) Foursquare location data to obtain information on private schools in Toronto neighbourhoods.

https://developer.foursquare.com/docs/build-with-foursquare/categories/

The Foursquare API search will use the venue categories for Private School "**52e81612bcbc57f1066b7a46."** This will return a list of all the private schools within a given location.

As an example, searching the Foursquare database using a neighbourhood's latitude and longitude and specifying the venue category for private schools will return a list of private schools within a certain radius.

4) Information on the population of each Toronto neighbourhood.

Population of Metropolitan Toronto by postal codes

Statistics Canada has an abundance of data on their official website based on the latest population census carried out in 2016. The dataset published in the link above shows the population breakdown based on the first three characters of the postal code. For purposes of this report, we will focus only on postal codes starting with 'M' as they represent Metropolitan Toronto neighbourhoods.

7 Methodology

To decide on the best Toronto neighbourhoods to launch a new private school, we need to build a profile of the private schools that are already in operation in the city as their numbers and distribution will determine the competitiveness and hence the profitability and survivability of the new school.

I will walk through the process of obtaining the datasets, cleaning them, normalising the different datasets and summarising the results to help us make an informed decision. A map of the prospective locations will also be shown on a city-wide map of Toronto together with other private schools in the vicinity.

7.1 Data on Toronto postal codes

The first step is to gather information on Toronto postal codes. The Wikipedia page on Metropolitan Toronto explains that there are **103** forward sorting areas (FSA) or postal codes starting with “M”.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

The data on this page was sorted to give us a snapshot of the postal codes, boroughs and neighbourhoods.

	Postal Code	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae
...	...	...	...
98	M9N	York	Weston
99	M9P	Etobicoke	Westmount
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
102	M9W	Etobicoke	Northwest, West Humber - Clairville

The postal code ranges from “M1B" to “M9W”.

Missing from this dataset are the latitude and longitude for each postal code. Geocoder, a geocoding library for Python, provides this information. The CSV file at the URL below also contains this information and will be used.

http://cocl.us/Geospatial_data

Combining the geolocation data with the original dataset gives us the result below.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476
...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188
99	M9P	Etobicoke	Westmount	43.696319	-79.532242
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054

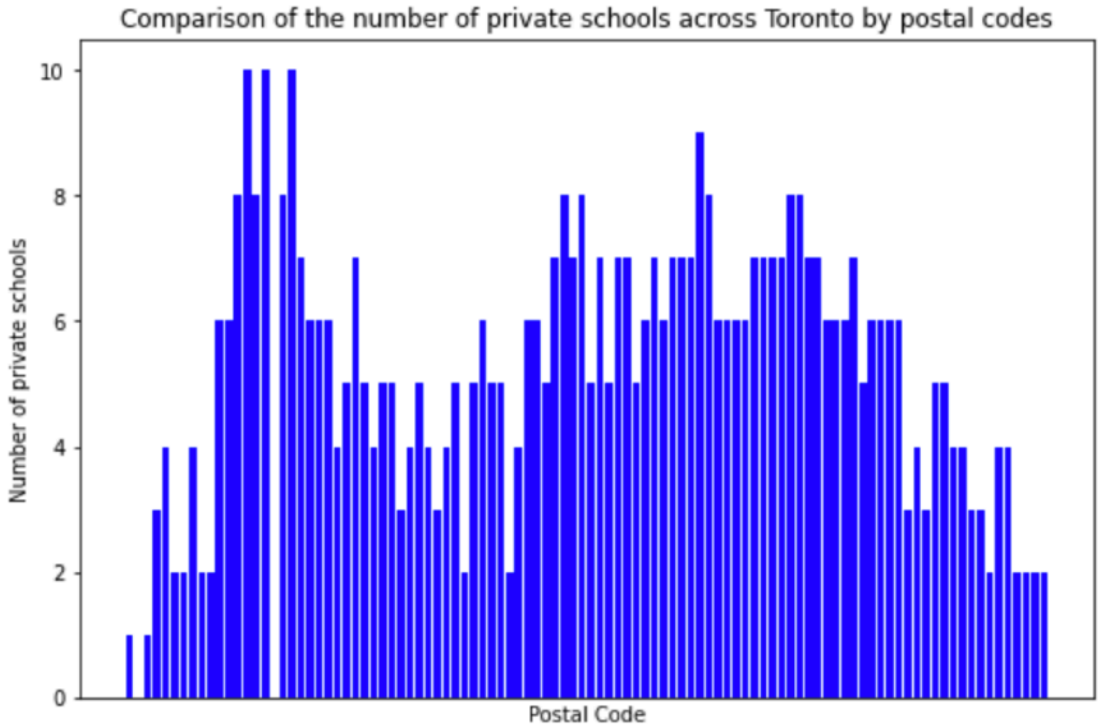
With the geolocation data, we can now harness the information in Foursquare's location database.

7.2 Foursquare API

Foursquare provides us with information on points of interest in a particular city or area and the data is available for free to developers accessing through its API. By doing a venue category search using the code for private schools "**52e81612bcbc57f1066b7a46"** and passing the latitude and longitude of the location to be searched together with the radius (6 km was chosen), the API returns information on all the private schools in the area within 6 km from the location. The number of private schools for each location is then counted and the results are displayed below.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Number of Schools
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	1
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1
3	M1G	Scarborough	Woburn	43.770992	-79.216917	3
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	4
...	...	...	...	...	...	...
98	M9N	York	Weston	43.706876	-79.518188	4
99	M9P	Etobicoke	Westmount	43.696319	-79.532242	2
100	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724	2
101	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437	2
102	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054	2

The bar chart below shows the same results starting with postal codes M1B and progressing to M9W on the x-axis. We can see that the number ranges from 0 to 10 private schools in each area.



With a larger radius, it is most likely that the number of schools will increase. However, it will also increase the chances of overlap between the areas i.e. the same school might be counted twice or three times due to it falling within the radius of several locations.

7.3 Population statistics

This next step involves obtaining data on the population in metropolitan Toronto. Statistics Canada maintains a population database with the latest data obtained from the census taken in 2016 (see link below). The complete dataset is quite detailed and so we will focus only on the population of each area code in Toronto.

Population of Metropolitan Toronto by postal codes

Data wrangling gives us the population based on each Toronto district.

	Postal Code	Population, 2016
895	M1B	66108.0
896	M1C	35626.0
897	M1E	46943.0
898	M1G	29690.0
899	M1H	24383.0
...	...	...
992	M9N	25074.0
993	M9P	20874.0
994	M9R	33743.0
995	M9V	55959.0
996	M9W	40684.0

Analysing the datasets showed that the population data for postal code “M7R” was missing. Upon searching for more details, it became clear that this area represented the “Canada Post Gateway Processing Centre”, likely a special district for administrative purposes. This area is dropped from the dataset. Working with the remaining data, we arrive at the following combined dataset with information on the postal code, borough, neighbourhood, geolocation, number of private schools and population.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Number of Schools	Population, 2016
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	1	66108.0
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0	35626.0
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	1	46943.0
3	M1G	Scarborough	Woburn	43.770992	-79.216917	3	29690.0
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476	4	24383.0
...	...	...	...	...	...	...	...
97	M9N	York	Weston	43.706876	-79.518188	4	25074.0
98	M9P	Etobicoke	Westmount	43.696319	-79.532242	2	20874.0
99	M9R	Etobicoke	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724	2	33743.0
100	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437	2	55959.0
101	M9W	Etobicoke	Northwest, West Humber - Clairville	43.706748	-79.594054	2	40684.0

8 Results

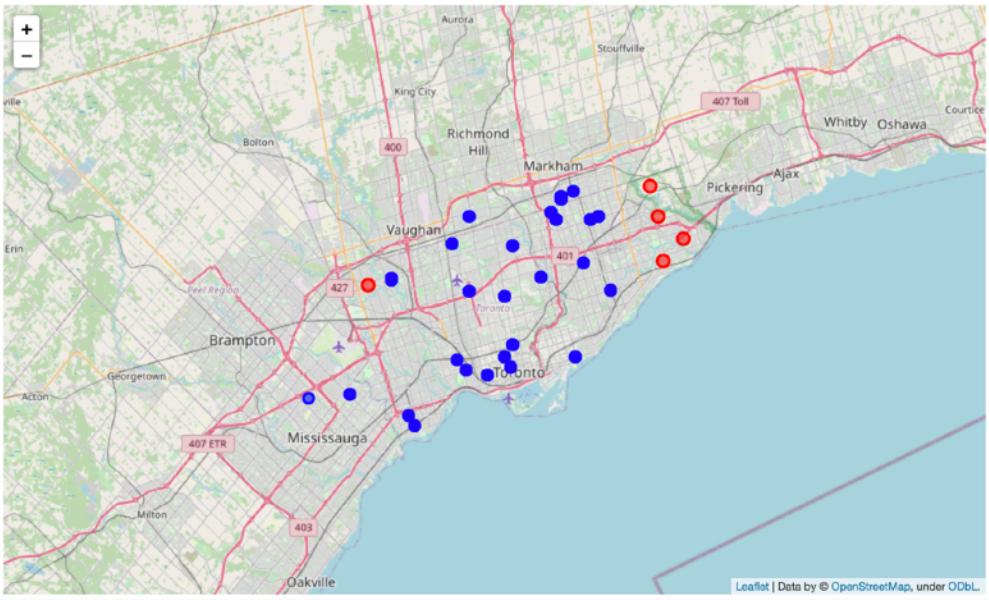
Before we can come to any useful conclusion, we should first normalise the population data and number of schools as the difference between the two is quite large.

A ratio of number of schools divided by the population is then taken and sorted in ascending order.

	Postal Code	Borough	Neighborhood	Latitude	Longitude	Number of Schools	Population, 2016	Ratio
0	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497	0.0	0.469399	0.000000
1	M1X	Scarborough	Upper Rouge	43.836125	-79.205636	0.0	0.198914	0.000000
2	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353	0.1	0.871023	0.114808
3	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711	0.1	0.618509	0.161679
4	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437	0.2	0.737302	0.271259

The first five rows show good candidate locations to setup a new private school. “M1C” and “M1X” both have no private schools and so a new school in either district will yield good results, especially for “M1C” as it has a fairly large population. Same for “M1B” with only one private school and an even larger population.

Finally, we show the locations of this five districts (in red) on a map of Toronto together with the locations of private schools (in blue).



9 Discussion

The results are presented based on the assumption that a large population in any given area with a small number of competitors will be an ideal location to base a new private school. There are other factors that can influence enrolments in a new school and this includes socio-economic conditions including poverty levels and the number of public and vocational schools.

Foursquare is unable to return results based on districts or postal codes and instead offers search criteria based on latitude, longitude and the distance from that location (radius). This skews the results in that we are unable to get a precise number of private schools in a particular area code. Searching using distance from a location or “near” a location does not take into account that districts are not all the same size. Choosing a small radius will mean we might miss out on schools in a large district. On the other hand, choosing a large radius might cause the same schools to be counted several times, thereby increasing the error and skewing the results by making the number of schools seem larger than they really are.

Some of the population data for Metropolitan Toronto was missing (in the case of area code “M7R”) and others could have been omitted without affecting the results. An example is a population reading of zero for “M5K” and “M5L” and a population of less than twenty people for four other

areas. There is no explanation from Statistics Canada on the reasons behind this.

10 Conclusion

It is hoped that this project will be beneficial to educationists and entrepreneurs hoping to start a new private school in Metropolitan Toronto. The success of any new school depends on many factors and this document has explored one of the main ones i.e. location. We hope other data scientists will explore other factors including socio-economic conditions, education, poverty levels and the ethnic make-up of the area to offer a more complete picture.

Thanks for reading.

Darren