



Capstone Project

The Battle of The Neighborhood Report

Rose Condon

October 22, 2019

Table Content

- 1 Project Introductions
- 2 Data Descriptions
- 3 Methodology
- 4 Tools and Process
- 5 Conclusion and Improvement

Background

Toronto is the biggest city in Canada. It is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 in 2016. Toronto is a truly international business center in North America and financial capital of Canada.

Toronto is also the largest centre of education, research and innovation in Canada. The education system combines the Public and Private schools in Toronto, both elementary and secondary, All schools take their curricular mandate from the Ontario Ministry of Education.

There are four types of school boards in Ontario. Depends on individual student language, religious background or choice, students can attend English Public, English Catholic, French Public, or French Catholic schools.

Publicly funded education is divided into three stages: early childhood education, for children from birth to age 6; elementary school, for students from kindergarten to grade 8; and secondary school, for students from grade 9 to 12.

Business Problem

What is EQAO

- The Education Quality and Accountability Office (EQAO) is an independent government agency of the Government of Ontario. The purpose of EQAO is develops and oversees reading, writing and mathematics tests that Ontario students must take in Grades 3, 6, 9, and 10.
- The EQAO test results give parents, teachers, principals and school boards information about how well students have learned what the province expects them to learn in reading, writing and mathematics.

EQAO results

- Only half of Ontario's Grade 6 students met the provincial standards for math in the 2016-2017 academic year, down seven percentage points from 2013. Meanwhile, 62 per cent of Grade 3 students met the provincial math standards, a decrease of five percentage points from 2014.
- For Grade 9 students, only 44 per cent met the standard in the applied math in 2017-2018. That number experienced a decline compare with years in 2013-2014.

Impacts

- After the release of results from 2017-2018 year's EQAO standardized testing, The Ontario Government announced a four-year math strategy.
- Ontario will spend more than \$55 million this year hiring math learning leads for school boards, providing "extensive" training in elementary and secondary schools, and expanding other programs like tutoring.
- A public concern has been raised regarding the 2017-2018 EQAO results. There is growing number of students using or searching for private tutor services. <https://www.cbc.ca/news/canada/toronto/ontario-math-curriculum-private-services-1.4445472>

Project Initiative

-
- The diagram features two overlapping circles, one on the left labeled "Background" and one on the right labeled "Problem". Arrows point from each circle to a list of bullet points.
- This project is the final assignment for Applied Data Science Capstone course on Coursera.
 - The purpose of this project is to discuss a business problem; leverage what the data science certificate courses cover and discover the solution.
 - This is a good exercise for the learners to explore data to find out the opportunity in business use case study.
- The business problem address here is about recent EQAO results for 2017-2018 in Toronto Jr elementary schools and Sr secondary schools.
 - The EQAO results rise public immediate attentions regarding the academic improvement.

Toronto – International Business Center



About Toronto

- Toronto is the biggest city in Canada, with a population of 2,731,571.
- Toronto is a truly international business center in North America and financial capital of Canada
- Toronto is also the largest centre of education, research and innovation in Canada
- All schools in Toronto take their curricular mandate from the Ontario Ministry of Education
- There are four types of school boards in Ontario, English Public, English Catholic, French Public, or French Catholic schools.

Business Opportunity

Consider to open an after-school tutor service in Toronto? According to the Wall Street Journal's Smart Money Magazine, now, it could be the perfect time for you to get into the education business.

Let's go to explore data we collect from multiple data sources and arrange them as a data frame for the analysis; so that we can target the recommended locations across different areas according to what we discover from the collected data.



Data Descriptions

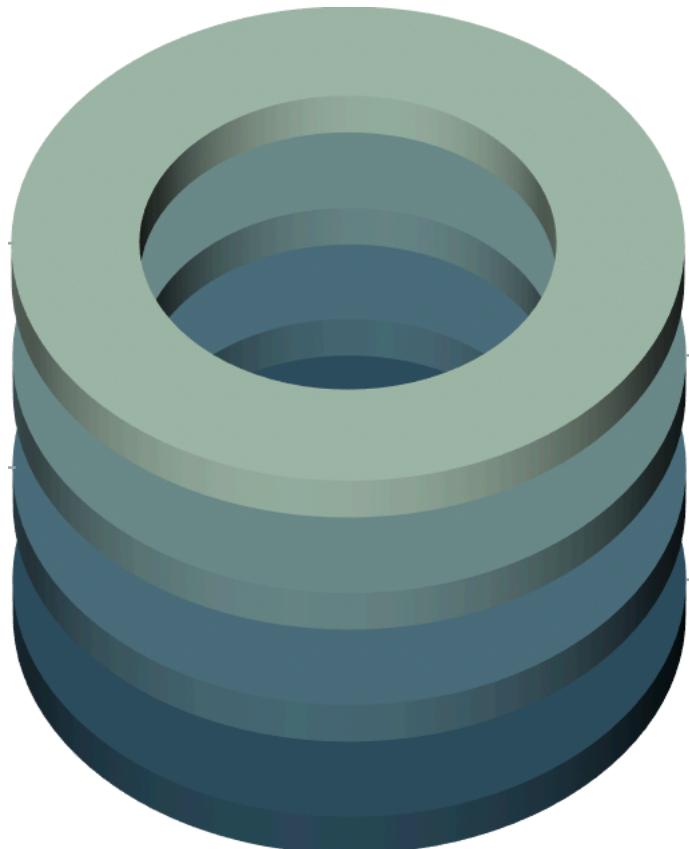
Data collection and process in most cases require up to 80% time in the whole Data Science project. How data is gathered and analyzed depends on many factors. These factors are including the content, the problems or issues can be identified with some indicators, the data source integrity, and the size of data.

There are some aspects should be considered in the data collection for this project.

- The schools number in a neighborhood: If the area has numbers of opening schools, particularly those are public schools, the higher demanding needs for tutoring services.
- The school ranking: if the school has lower ranking, then the number of students are looking for tutors services for academy improvement is higher.
- The number of tutor services: to avoid the competition and towards to more successful in business, the area has no or few tutor services business opening could have an opportunity to open one.

Data Preparation

Data preparation in most cases require up to 80% time in the Data Science project. How data is gathered and analyzed depends on many factors.



Aspects should be considered in the data collection

- Toronto nationhood Data : collected from WiKI
- Toronto schools Data : collected from Toronto city open dataset
- Geo data : collected from Wiki and others
- School Ranking data : collected Fraser Institute web site
- Correlation data : existing tutor services; and borough income range

Methodology in Data Science Project

According to one of the course in Professional Data Science Certificate, the methodology in Data Science contains 5 main aspects.

1. From Problem to Approach
2. From Requirements to Collection
3. From Understanding to Preparation
4. From Modelling to Evaluation
5. From Deployment to Feedback

Five Aspects in Methodology

Deployment to Feedback

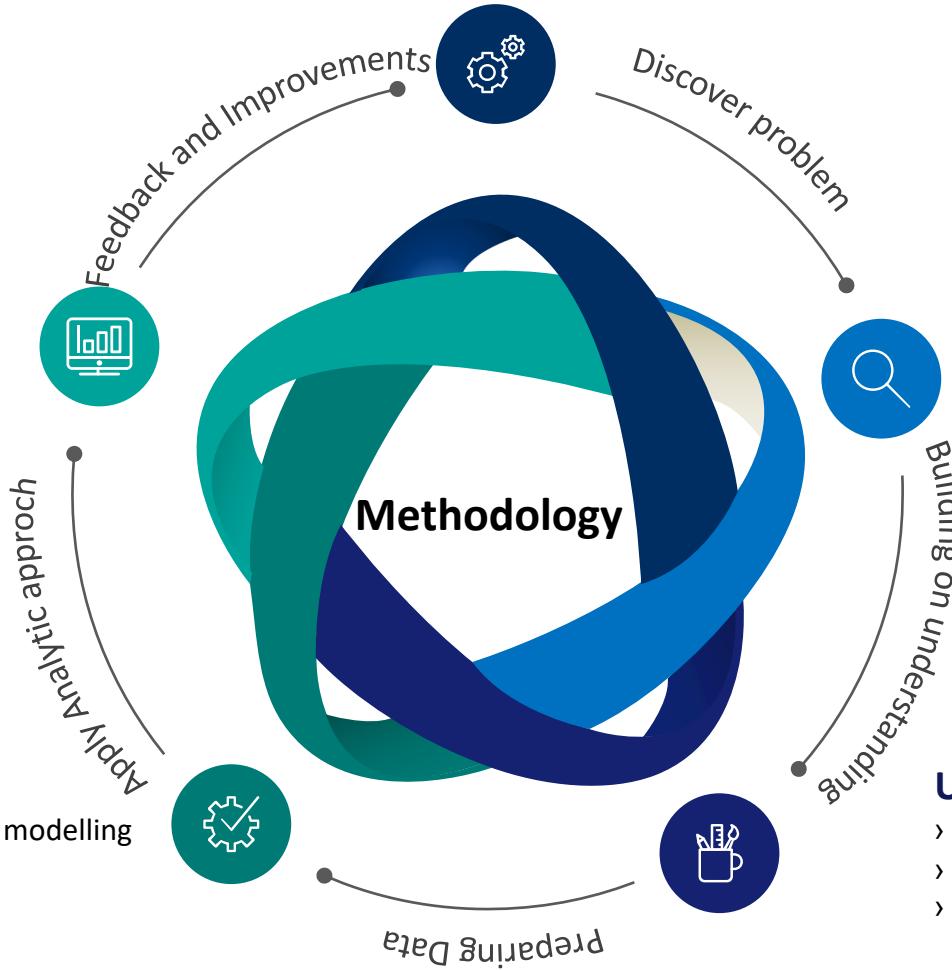
- › Can model identify high-risk patients
- › Can we use model design suitable intervention actions
- › Have received constructive feedbacks

Modelling to Evaluation

- › What training set for predictive modelling
- › Which algorithms will be used
- › Which way to visualize data

Problem to Approach

- › What problem trying to resolve
- › Can we get the data
- › How to use data



Requirement to Collection

- › What data do I need.
- › Where is data source.
- › How to get the data.

Understanding to Preparation

- › Has seeking clarification been done
- › Is collected data representative
- › Which analytic approaches will be applied.

Why is Important To Apply Methodology

A methodology can be used within data science, to ensure that we understand the question at hand; to obtain, understand, prepare, and model the data; so as to the data used in problem solving is relevant and properly manipulated to address the question.

For example

- Data requirement can identify the required data fulfills the data requirements
- Data collection requires data scientists and analytics team members can discuss various ways to better manage their data
- Model evaluation is performed during model development and before the model is deployed. Evaluation allows the quality of the model to be assessed but it's also an opportunity to see if it meets the initial request.
- The end goal is to move the data scientist to a point where a data model can be built to answer the question

Tools and Process

- This project uses the Jupyter Notebook for development IDE using Python programming as well as data science libraries, ie, pandas, matplotlib, seaborn, etc
- The Data sources are collected from data providers on public data repositories
- Identifying the necessary data contents, format from initial data collection
- Leverage Foursquare API to access location data
- Conduct statistical analysis and descriptive model
- Apply linear regression machine learning algorithm
- Using visualization libraries for better insights about the data

Toronto neighborhood Data: including borough, geo location and address

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge, Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek, Rouge Hill, Port Union	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



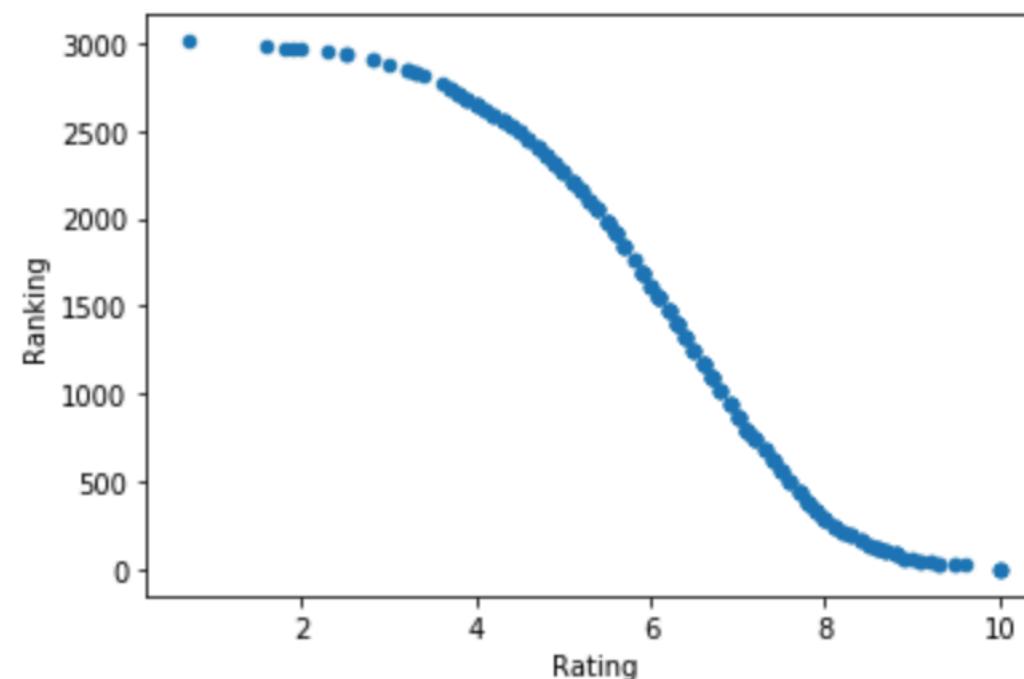
	NAME	SOURCE_ADDRESS	SCHOOL_TYPE_DESC	ADDRESS_FULL	POSTAL_CODE	MUNICIPALITY	LATITUDE	LONGITUDE
0	Avondale Public School	25 Bunty Lane	English Public	25 Bunty Lane	M2K 1W4	North York	43.776502	-79.385190
1	Avondale Secondary Alternative School	24 Silverview Dr	English Public	24 Silverview Dr	M2M 2B3	North York	43.786315	-79.411846
2	AYJ Global Academy	4 Lansing Sq	Priv	4 Lansing Sq	M2J 5A2	North York	43.774091	-79.325220
3	Bais Brucha School	3077 Bathurst St	Priv	3077 Bathurst St	M6A 1Z9	North York	43.719936	-79.429570
4	Bais Chaya Mushka Elementary School	4375 Chesswood Dr	Priv	4375 Chesswood Dr	M3J 2C2	North York	43.760717	-79.477445
5	Bais Chomesh High School for Girls	3600 Bathurst St	Priv	3600 Bathurst St	M6A 2C9	North York	43.731124	-79.433080
6	Bais Yaakov Elementary School	15 Saranac Blvd	Priv	15 Saranac Blvd	M6A 2G4	North York	43.721429	-79.431869
7	Baitul Mukarram Academy School	3334 Danforth Ave	Priv	3334 Danforth Ave	M1L 1C6	Scarborough	43.693739	-79.278141
8	Bala Avenue Community School	6 Bala Ave	English Public	6 Bala Ave	M6M 2E1	York	43.689471	-79.499016
9	Balmy Beach Community School	14 Pine Ave	English Public	14 Pine Ave	M4E 1L6	former Toronto	43.676436	-79.289827

After done some data clean up, here we have nice Toronto Schools Data which is including name, geo location and address

Data clean up and re-order need to be done for Toronto Schools Ranking Data for Junior Elementary Schools

	2017-18 Rank	School Name	Postal Code	City	Rating	Ranking
0	1/3046	Avondale Alternative	M2N 2V4	Toronto	10.0	1.0
1	1/3046	Havergal	M5N2H9	Toronto	10.0	1.0
2	1/3046	Islamic Institute of Toronto	M1X 1S3	Toronto	10.0	1.0
3	1/3046	Northmount	M3B 1S3	Toronto	10.0	1.0
4	1/3046	Sathya Sai	M1R 4E5	Toronto	10.0	1.0

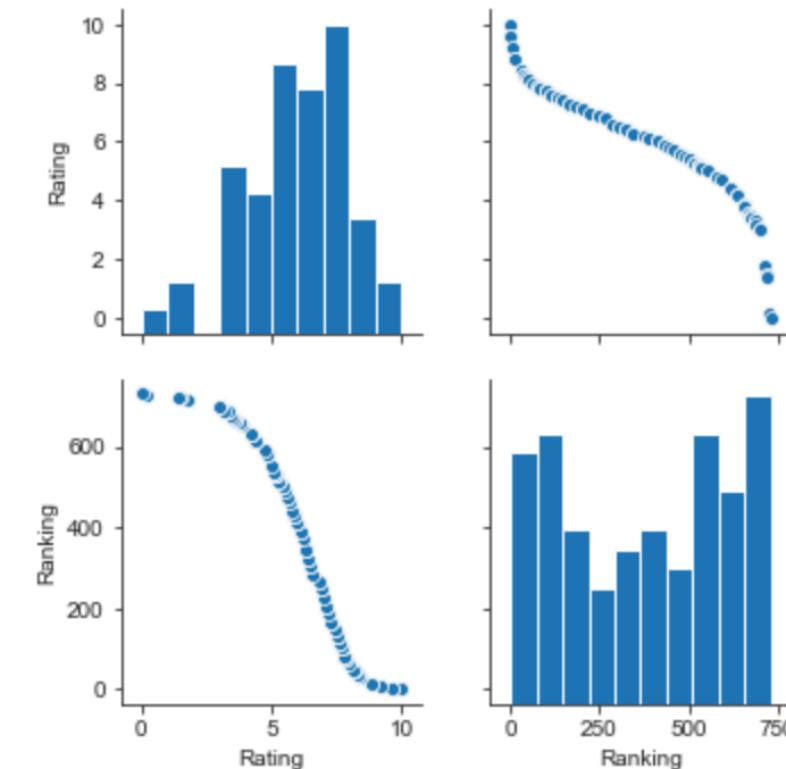
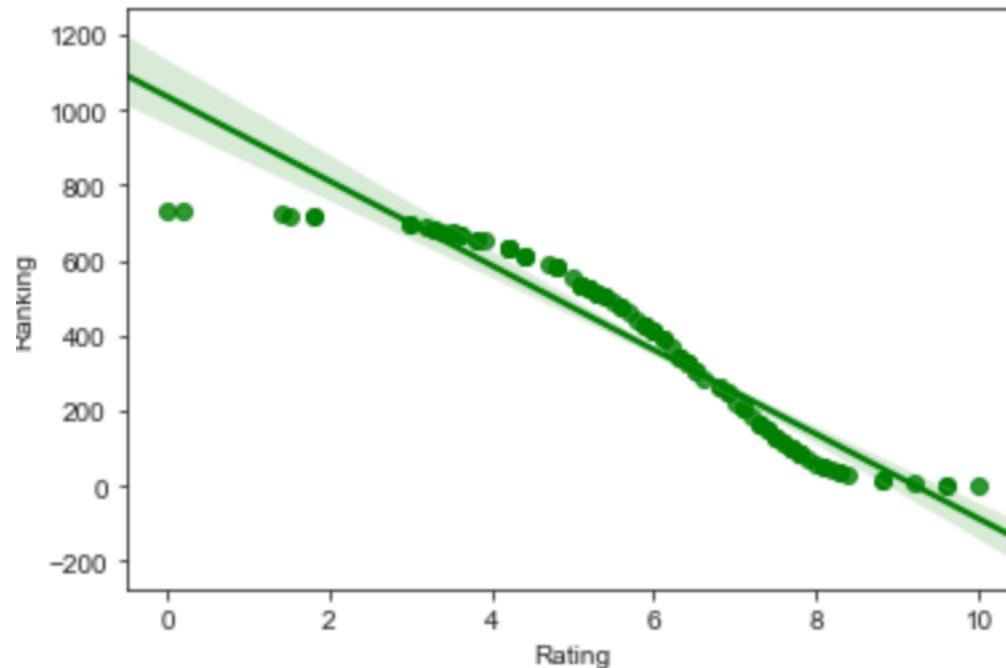
Use matplotlib for better insights of the average ranking



Same activates of Data clean up and re-order for Toronto Schools Ranking Data for Senior Secondary Schools

	2017-18 Rank	School Name	Postal Code	City	Rating	Ranking
0	1/738	Havergal	M5N2H9	Toronto	10.0	1
1	3/738	St Michael's Choir	M5B1X2	Toronto	9.6	3
2	3/738	Ursula Franklin	M6P3J7	Toronto	9.6	3
3	7/738	North Toronto	M4P1T7	Toronto	9.2	7
4	14/738	Cardinal Carter-Arts	M2N3C8	Toronto	8.8	14

Use matplotlib and seaborn to create plots for better insights of the average ranking



We have our target schools. Now we need to find out the correlation data
 - the neighborhood geo data for the conclusion

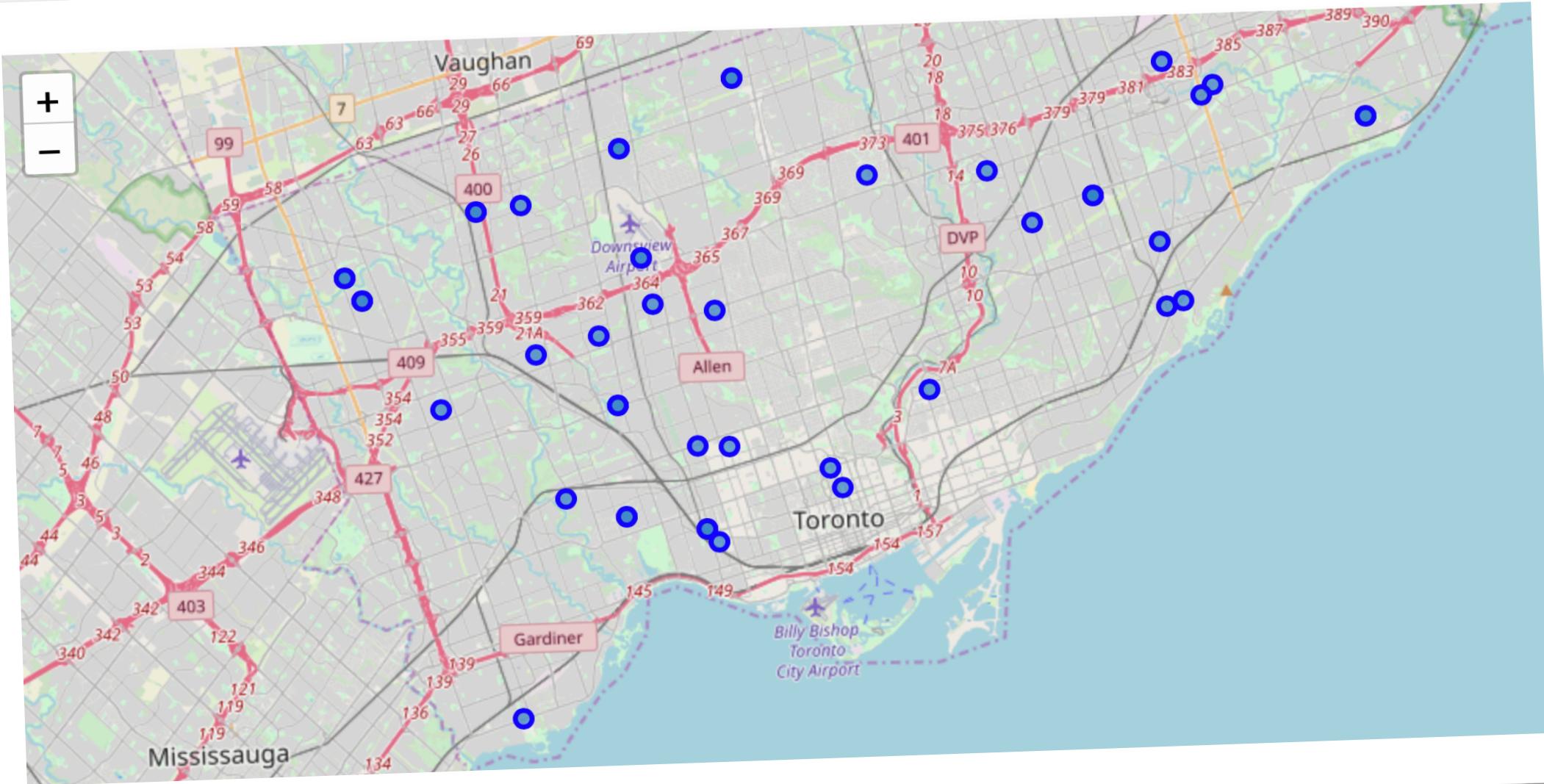
2017-18 Rank	School Name	Postal Code	City	Rating	Ranking	Neighborhood	Borough	Latitude	Longitude
0	2315/3046 Charles E Webster	M6M3X7	Toronto	4.9	2315.0	Del Ray, Keelesdale, Mount Dennis, Silverthorn	York	43.690287	-79.476243
1	2315/3046 Fairmount	M1M1C7	Toronto	4.9	2315.0	Cliffcrest, Cliffside, Scarborough Village West	Scarborough	43.715646	-79.242479
2	2315/3046 Our Lady of Guadalupe Catholic	M2J3C2	Toronto	4.9	2315.0	Fairview, Henry Farm, Oriole	North York	43.781941	-79.348741
3	2315/3046 St Bernard	M6M4W4	Toronto	4.9	2315.0	Del Ray, Keelesdale, Mount Dennis, Silverthorn	York	43.690287	-79.476243
4	2315/3046 St Charles	M6B2W1	Toronto	4.9	2315.0	Glencairn	North York	43.717388	-79.435480
5	2315/3046 Tam O'Shanter	M1T2V3	Toronto	4.9	2315.0	Clarks Corners, Sullivan, Tam O'Shanter	Scarborough	43.800624	-79.301345
6	2315/3046 Walter Perry	M1K4M7	Toronto	4.9	2315.0	East Birchmount Park, Ionview, Kennedy Park	Scarborough	43.733465	-79.251249
7	2361/3046 Carleton Village	M6N2Z4	Toronto	4.8	2361.0	The Junction North, Runnymede	York	43.673121	-79.499085
8	2361/3046 Elmbank	M9V3R4	Toronto	4.8	2361.0	Albion Gardens, Beaumont Heights, Humbergate, ...	Etobicoke	43.730931	-79.586709
9	2361/3046 Greenholme	M9V3M5	Toronto	4.8	2361.0	Albion Gardens, Beaumont Heights, Humbergate, ...	Etobicoke	43.730931	-79.586709

Conclusions

Now we can come out a list of name to answer the business problem question:

- **Which areas or neighborhoods in Toronto are ideal locations to open a after-school tutor service?**

Del Ray, Keelesdale, Mount Dennis, Silverthorn,
Cliffcrest, Cliffside, Scarborough Village West,
Glencairn,
East Birchmount Park, Ionview, Kennedy Park,
Albion Gardens, Beaumont Heights, Humbergate, Jamestown, Mount Olive, Silverstone, South Steeles,
Thistletown, Downsview, North Park, Upwood Park,
Emery, Humberlea,
Rouge, Malvern,
East Toronto,
Church and Wellesley,
Christie,
Caledonia-Fairbanks,
Downsview Northwest,
Dovercourt Village, Dufferin,
Dorset Park, Scarborough Town Centre, Wexford Heights,
Parkwoods,
Kingsview Village, Martin Grove Gardens, Richview Gardens, St. Phillips,
Harbord, University of Toronto,
Runnymede, Swansea,
Guildwood, Morningside, West Hill, Woburn,
Maryvale, Wexford,
Agincourt, Northwest,
"L'Amoreaux West",
Brockton, Exhibition Place, Parkdale Village,
Newtonbrook, Willowdale



Future Improvements

-
- ```
graph TD; A((Problem to Approach)) --> B((Requirements to Collection)); B --> C((Understanding to Preparation)); C --> D((Modelling to Evaluation)); D --> E((Deployment to Feedback)); E --> A
```
- Expand problem coverage
  - Add more channels of expertise for questionnaires
  - Add more ingredients to data collected from public repos
    - Collect more details for requirements
  - More efficient in data preparation
  - Add feature engineering
  - Consider Classification modeling
    - Reduce complexity
  - Improves test data quantity
  - Share results with peers with report