**Tutoring Service Location Search**

**Final Capstone Project [SA]: Week 2**

## 1. Introduction/Business Problem:

Our goal is to find the best location (areas/neighborhood) to start a tutoring service business in Ontario (Brampton), Canada. We need to find areas with the schools with the largest enrollment as well as the lowest number of existing tutoring services (offered nearby). The presence of schools determines existence of clients (students), the lack of existing tutoring services means a lack of competition for our prospective business owner. We also consider schools with a lower percentage of low-income families or a higher percentage of parents with university education which indicates that parents can afford the tutorial services offered by the business owner. The process is to specialize, filtering for Elementary or Secondary School, cities, and such.

## 2. Data:

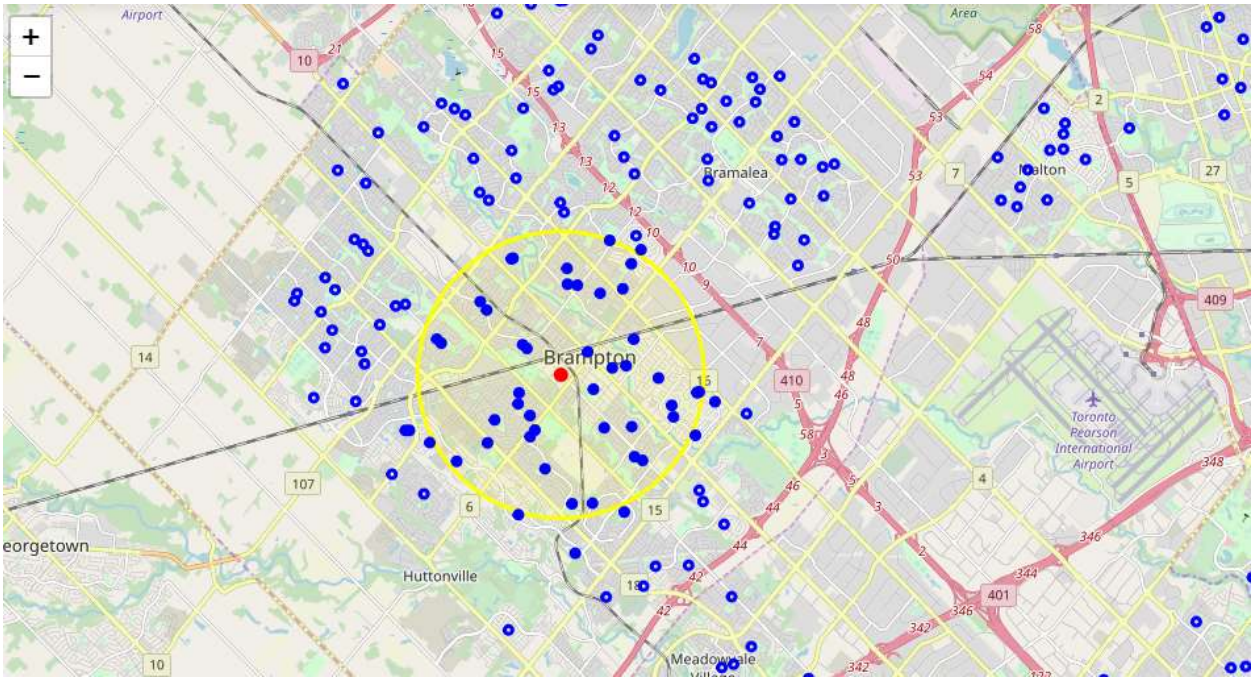
Source data is publicly available from the government of Ontario at the following link:
https://www.ontario.ca/data/schoolinformation-and-student-demographics
This data-set provides info about almost 5,000 schools all over Ontario, with the location coordinates and the number of students enrolled. Using the location data along with foursquare, a venue search engine, we may search for tutoring services near each of the schools. These will be categorized by the number of services nearby.

## 3. Methodology:

After getting the data, we will keep only the columns that will be utilized for our study purpose. Required fields are: school name, location coordinates, enrolment and the percentages of low-income families and parents with university education. Entries with null school names, coordinates or enrolment values are ignored. Null values in the remaining columns are replaced with the average of the non-null values. We will use the schools in the city of Brampton to show the process applied. Using the Brampton school coordinates we search the Foursquare database for ‘tutor‘, ‘math‘, and ‘learning‘ venues at the location of the school. A search radius of 3 km has been chosen for this example. From the resulting venues we collect the number of venues returned for each school as well as a complete list of unique venues returned along with their coordinates for mapping purposes.

We will plot markers for both the schools and the tutoring services found:

- Blue markers are schools.
- Red markers are tutoring services each with a yellow circle for the 3km search radius used.




A naive measure of available students per school is defined. We calculate for each school by the following equation, the 1 in the denominator represents including our projected tutoring service, the measure gives each nearby tutoring service an equal share of the enrolled students for the school:

Available_students = (enrolled_students / (nearby_tutors + 1))

This measure is important because it emphasizes the priority of schools with few nearby tutoring services by the nature of f(x) = (1 / (x+1))

This measure can (and will soon) be extended by instead acquiring the distance between a given school and each of the nearby services and assigning a portion of the enrolled students to each service depending on how close they are to the school.

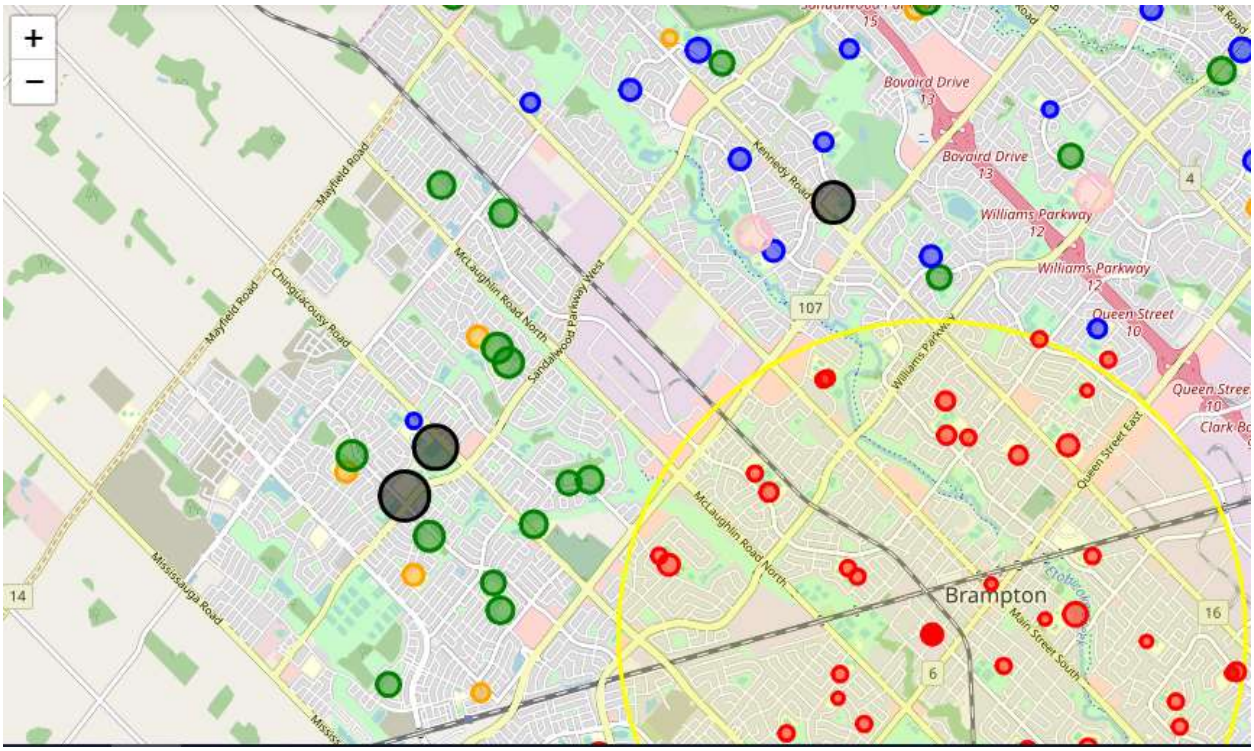
Take for example School X which has 100 students, Service A 1km away and Service B 2km away. We have a ‘total service distance‘ of 3km we want Service A to obtain twice as many students as it is half as far, we calculate (1− (1km / 3km)) = 0.67 and (1− (2km / 3km)) = 0.33. As desired Service A is assigned 67 students and Service B is assigned 33 students.

This can be further augmented by only assigning the portion of the students which are not low income or assigning more importance to the percentage of parents with university education. We have some data to categorize our schools by profitability now. We will group our schools into clusters through a KMeans algorithm. We would like to have classes for low/medium/high levels of our three features: enrolment, number of nearby services and economic status so we look to define 9 clusters.

**4. Results:**

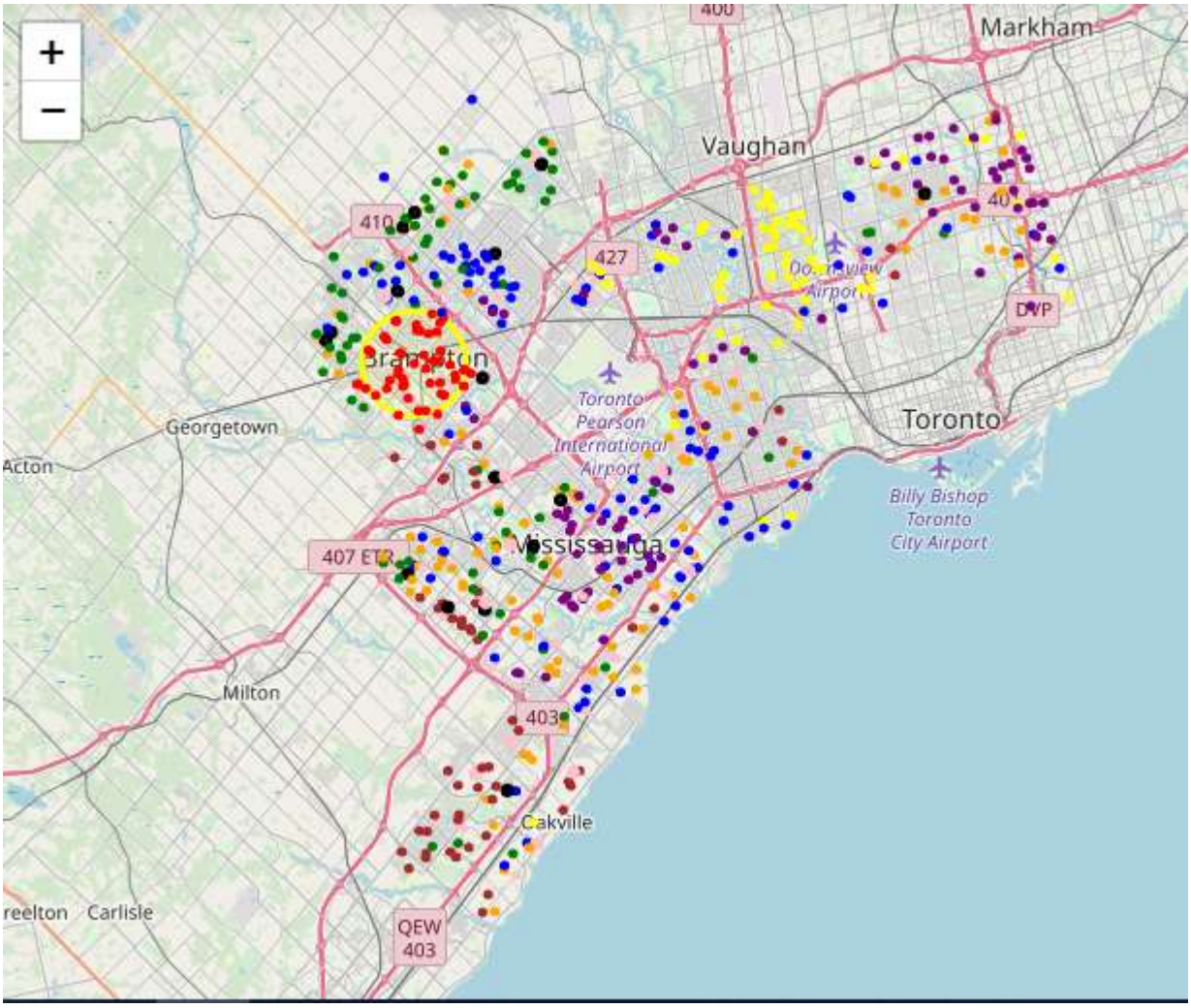
The clusters received reveal the schools with the least competition and highest available students. The areas interesting to us are the green and orange markers, green markers are high enrolment schools:

| | Count | Enrol Tutors Ratio | Enrol | Pct Low Income | Pct Uni Parents | Tutor Services |
|---|---|---|---|---|---|---|
| **Green** | 27.0 | 1754.06 | 1754.06 | 19.61 | 24.29 | 0.0 |
| **Orange** | 14.0 | 1089.96 | 1089.96 | 20.94 | 22.68 | 0.0 |
| **Brown** | 22.0 | 731.83 | 731.83 | 14.85 | 40.33 | 0.0 |
| **Red** | 20.0 | 590.04 | 590.04 | 12.08 | 61.43 | 0.0 |
| **Lime** | 17.0 | 418.96 | 418.96 | 28.12 | 38.87 | 0.0 |
| **Purple** | 20.0 | 364.92 | 364.92 | 15.34 | 41.34 | 0.0 |
| **Pink** | 6.0 | 349.53 | 349.53 | 32.42 | 15.25 | 0.0 |
| **Blue** | 28.0 | 317.44 | 317.44 | 18.41 | 21.38 | 0.0 |
| **Black** | 15.0 | 274.39 | 548.78 | 18.39 | 24.17 | 1.0 |




The best area within Brampton seems to be on the west, near Guardian Angels Catholic Elementary School and the surrounding schools (Worthington Public School, St. Bonaventure Catholic Elementary, McCrimmon Middle School, Brisdale Public School, St. Aidan Catholic Elementary, etc.).

Most of the code used to generate this report has been generalized for simple use in doing this process for cities in Ontario as well as combinations for cities. Example Map:




**5. Discussion:**

This study was kept simple. There exists more data that may be utilized, for example the percentage of students who pass standardized tests. And we are only considering three searches, for better success we may add more. We are also filtering said results manually and it could be automated. The biggest limitation in the study is our sole use of foursquare free service. More search engines could be applied to obtain a complete list of the nearby services.

**6. Conclusion:**

We may conclude that there is one clearly optimal area towards the west side of Brampton. Please see the map below for the best location for our Business Owner to start a Tutorial Service: