

## Capstone Project

---

### 1. Introduction to Business Problem

I live in Downtown Toronto (postal code: M5S) and work for a IT company there. My wife is running a Gyms, sport clothes & accessories shop at home. Due to company requirement, I have to move to work for company's branch in Scarborough next month, so all of my family have to move to Scarborough.

My wife also would like to keep to running her business in Scarborough. Hence, **it comes to the question that which neighborhood we should live there to support her business.**



To answer above question, my idea is that we will explore Scarborough neighborhoods to find out which neighborhood has highest number of Gym / Fitness Center, Sport clubs, parks (where people can do exercise), swimming pools and so on. Neighborhood which is met above condition is one of important condition to support my wife business running well there.

### 2. Data requirements:

Below is data that I plan to use for this project:

1. Canada's postal code, neighborhoods and geographic data
  - + [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M'](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M')
  - + [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)
2. Venues and venues category information in Scarborough (from Foursquare data)

### 3. Data Processing:

#### Step 1: Processing Scarborough geography data:

- Processing data to get geography data of Scarborough as below table. This is the result of collecting data from provided sources above.
- This data will help to explore about Scarborough neighborhood via Foursquare API in next step.

|   | PostalCode | Borough     | Neighbourhood                          | 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 | Morningside, West Hill                 | 43.763573 | -79.188711 |
| 3 | M1G        | Scarborough | Woburn                                 | 43.770992 | -79.216917 |
| 4 | M1H        | Scarborough | Cedarbrae                              | 43.773136 | -79.239476 |

#### Step 2: Explore Scarborough venues via Foursquare API

- Explore venues category from Foursquare and select categories which will be impacted to my wife business: {'Gym', 'Park', 'Yoga Studio', 'Gym / Fitness Center', 'Soccer Field', 'Beach', 'Gym Pool', 'Badminton Court', 'Pool', 'Pool Hall', 'Golf Course', 'Tennis Court'}

|   | Neighborhood   | Neighborhood Latitude | Neighborhood Longitude | Venue              | Venue Latitude | Venue Longitude | Venue Category       |
|---|----------------|-----------------------|------------------------|--------------------|----------------|-----------------|----------------------|
| 0 | Rouge, Malvern | 43.806686             | -79.194353             | Images Salon & Spa | 43.802283      | -79.198565      | Spa                  |
| 1 | Rouge, Malvern | 43.806686             | -79.194353             | Caribbean Wave     | 43.798558      | -79.195777      | Caribbean Restaurant |
| 2 | Rouge, Malvern | 43.806686             | -79.194353             | Harvey's           | 43.800106      | -79.198258      | Fast Food Restaurant |
| 3 | Rouge, Malvern | 43.806686             | -79.194353             | Wendy's            | 43.807448      | -79.199056      | Fast Food Restaurant |
| 4 | Rouge, Malvern | 43.806686             | -79.194353             | Wendy's            | 43.802008      | -79.198080      | Fast Food Restaurant |

- Using onehot technique to create below table which only contained category which I need to focus on

|                                 | Gym | Park | Yoga Studio | Gym / Fitness Center | Soccer Field | Beach | Gym Pool | Badminton Court | Pool | Pool Hall | Golf Course | Tennis Court |
|---------------------------------|-----|------|-------------|----------------------|--------------|-------|----------|-----------------|------|-----------|-------------|--------------|
| Neighborhood                    |     |      |             |                      |              |       |          |                 |      |           |             |              |
| Agincourt                       | 0   | 1    | 0           | 0                    | 0            | 0     | 0        | 1               | 1    | 1         | 0           | 0            |
| Agincourt North, Milliken       | 1   | 2    | 0           | 0                    | 0            | 0     | 0        | 0               | 0    | 0         | 0           | 0            |
| Birch Cliff                     | 1   | 2    | 0           | 0                    | 0            | 0     | 1        | 0               | 0    | 0         | 0           | 0            |
| Cedarbrae                       | 0   | 0    | 1           | 1                    | 0            | 0     | 0        | 0               | 0    | 0         | 0           | 0            |
| Clairlea, Golden Mile, Oakridge | 1   | 2    | 0           | 0                    | 1            | 0     | 0        | 0               | 0    | 0         | 0           | 0            |

- Using K-nearest algorithm to cluster our neighborhood into 5 different groups (k=5). Group with highest 'Total Sum' will be the group which is most fitted to my purpose.

|           | Gym | Park  | Yoga Studio | Gym / Fitness Center | Soccer Field | Beach | Gym Pool | Badminton Court | Pool | Pool Hall | Golf Course | Tennis Court | Total Sum |
|-----------|-----|-------|-------------|----------------------|--------------|-------|----------|-----------------|------|-----------|-------------|--------------|-----------|
| <b>G5</b> | 0.0 | 1.000 | 0.000       | 0.00                 | 0.000000     | 0.0   | 0.000000 | 1.000           | 1.0  | 1.0       | 0.000       | 0.000000     | 4.000000  |
| <b>G1</b> | 1.0 | 2.000 | 0.000       | 0.00                 | 0.333333     | 0.0   | 0.333333 | 0.000           | 0.0  | 0.0       | 0.000       | 0.000000     | 3.666667  |
| <b>G3</b> | 0.0 | 1.000 | 0.000       | 0.00                 | 0.000000     | 2.0   | 0.000000 | 0.000           | 0.0  | 0.0       | 0.000       | 0.000000     | 3.000000  |
| <b>G4</b> | 1.0 | 0.000 | 0.000       | 0.00                 | 0.000000     | 0.0   | 0.333333 | 0.000           | 0.0  | 0.0       | 0.000       | 0.333333     | 1.666667  |
| <b>G2</b> | 0.0 | 0.375 | 0.125       | 0.25                 | 0.125000     | 0.0   | 0.000000 | 0.125           | 0.0  | 0.0       | 0.125       | 0.000000     | 1.125000  |

#### 4. Result:

- With K-nearest algorithm, **Agincourt** is the best neighborhood which my family should move in and find a new house there.
- In case of, the cost to rent a house in Agincourt is too expensive, “Agincourt North, Milliken”, “Birch Cliff”, “Clairlea, Golden Mile, Oakridge” are also other options for us to consideration.

|           | Neighborhood                                      | Group |
|-----------|---|-------|
| <b>0</b>  | Agincourt   | 5     |
| <b>1</b>  | Agincourt North, Milliken                         | 1     |
| <b>2</b>  | Birch Cliff                                       | 1     |
| <b>3</b>  | Cedarbrae   | 2     |
| <b>4</b>  | Clairlea, Golden Mile, Oakridge                   | 1     |
| <b>5</b>  | Cliffcrest, Cliffside                             | 3     |
| <b>6</b>  | Dorset Park, Scarborough Town Centre, Wexford ... | 2     |
| <b>7</b>  | Highland Creek, Rouge Hill, Port Union            | 2     |
| <b>8</b>  | Ionview, Kennedy Park                             | 2     |
| <b>9</b>  | Maryvale, Wexford                                 | 2     |
| <b>10</b> | Morningside, West Hill                            | 4     |
| <b>11</b> | Rouge, Malvern                                    | 4     |
| <b>12</b> | Scarborough Village                               | 2     |
| <b>13</b> | Steeles West                                      | 4     |
| <b>14</b> | Tam O'Shanter                                     | 2     |
| <b>15</b> | Woburn  | 2     |

#### 5. Conclusion:

- In the limitation of data source, this report is only considered the around environments which have good impact to provided business. Others condition (like rental cost, populations, etc.) have not taken into account yet.