Data

Data collection is an important part for any Data science project. We will not be getting any readymade data to work with. Hence, with the limited data that we have obtained from Wikipedia, we will be proceeding with our analysis.

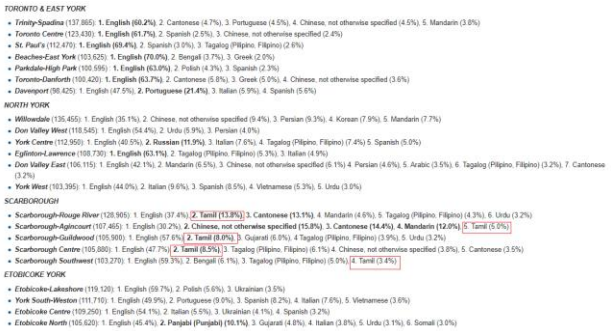
Wikipedia: *List of Postal Code for Toronto*

Below is the sample of the postal codes that we will be using for this analysis. We will be scrapping the below table data from Wikipedia and will be doing the necessary data wrangling and data cleansing activity and will use the final data for our further analysis. One such example is that, we will be removing the Borough's that has the value 'Not assigned' and will work with the remaining data. From the formatted data, we will be taking the data of Scarborough alone to apply the Data science methodology.

Postcode	Borough	Neighbourhood
M1A	Not assigned	Not assigned
M2A	Not assigned	Not assigned
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Harbourfront
M5A	Downtown Toronto	Regent Park
M6A	North York	Lawrence Heights

Wikipedia: *Demography of Toronto*

Complete demography of Toronto has been provided in the above link. We have taken only the below section to select a Borough – Scarborough, upon which we will be applying the Battle of Neighbourhood analysis. We have taken Scarborough for analysis, because, as we can see from the below image that majority of Tamil populations are in Scarborough. Whereas, the Tamil population is very less in other boroughs which is evident from the below image. Apart from this, we will not be using the below data for any other purpose.



TORONTO & EAST YORK

- *Trinity-Spadina* (137,865): **1. English (60.2%)**, 2. Cantonese (4.7%), 3. Portuguese (4.5%), 4. Chinese, not otherwise specified (4.5%), 5. Mandarin (3.6%)
- *Toronto Centre* (123,430): **1. English (61.7%)**, 2. Spanish (2.5%), 3. Chinese, not otherwise specified (2.4%)
- *St. Paul's* (112,470): **1. English (69.4%)**, 2. Spanish (3.0%), 3. Tagalog (Pilipino, Filipino) (2.6%)
- *Beaches-East York* (103,625): **1. English (70.0%)**, 2. Bengali (3.7%), 3. Greek (2.0%)
- *Parkdale-High Park* (100,595): **1. English (63.0%)**, 2. Polish (4.3%), 3. Spanish (2.3%)
- *Toronto-Danforth* (100,420): **1. English (63.7%)**, 2. Cantonese (5.8%), 3. Greek (5.0%), 4. Chinese, not otherwise specified (3.6%)
- *Davenport* (98,425): 1. English (47.5%), **2. Portuguese (21.4%)**, 3. Italian (5.9%), 4. Spanish (5.6%)

NORTH YORK

- *Willowdale* (136,465): 1. English (35.1%), 2. Chinese, not otherwise specified (9.4%), 3. Persian (9.3%), 4. Korean (7.9%), 5. Mandarin (7.7%)
- *Don Valley West* (118,545): 1. English (54.4%), 2. Urdu (5.9%), 3. Persian (4.0%)
- *York Centre* (112,950): 1. English (40.5%), **2. Russian (11.9%)**, 3. Italian (7.6%), 4. Tagalog (Pilipino, Filipino) (7.4%) 5. Spanish (5.0%)
- *Eglinton-Lawrence* (108,730): **1. English (63.1%)**, 2. Tagalog (Pilipino, Filipino) (5.3%), 3. Italian (4.9%)
- *Don Valley East* (106,115): 1. English (42.1%), 2. Mandarin (6.5%), 3. Chinese, not otherwise specified (6.1%) 4. Persian (4.6%), 5. Arabic (3.5%), 6. Tagalog (Pilipino, Filipino) (3.2%), 7. Cantonese (3.2%)
- *York West* (103,395): 1. English (44.0%), 2. Italian (9.6%), 3. Spanish (8.5%), 4. Vietnamese (5.3%), 5. Urdu (3.0%)

SCARBOROUGH

- *Scarborough-Rouge River* (128,905): 1. English (37.4%), **2. Tamil (13.8%)**, **3. Cantonese (13.1%)**, 4. Mandarin (4.6%), 5. Tagalog (Pilipino, Filipino) (4.3%), 6. Urdu (3.2%)
- *Scarborough-Agincourt* (107,465): 1. English (30.2%), **2. Chinese, not otherwise specified (15.8%)**, **3. Cantonese (14.4%)**, **4. Mandarin (12.0%)**, 5. Tamil (5.0%)
- *Scarborough-Guildwood* (105,900): 1. English (57.6%), **2. Tamil (8.0%)**, 3. Gujarati (6.0%), 4 Tagalog (Pilipino, Filipino) (3.9%), 5. Urdu (3.2%)
- *Scarborough Centre* (105,880): 1. English (47.7%), **2. Tamil (8.5%)**, 3. Tagalog (Pilipino, Filipino) (6.1%) 4. Chinese, not otherwise specified (3.8%), 5. Cantonese (3.5%)
- *Scarborough Southwest* (103,270): 1. English (59.3%), 2. Bengali (6.1%), 3. Tagalog (Pilipino, Filipino) (5.0%), 4. Tamil (3.4%)

ETOBICOKE YORK

- *Etobicoke-Lakeshore* (119,120): 1. English (59.7%), 2. Polish (5.6%), 3. Ukrainian (3.5%)
- *York South-Weston* (111,710): 1. English (49.9%), 2. Portuguese (9.0%), 3. Spanish (8.2%), 4. Italian (7.6%), 5. Vietnamese (3.6%)
- *Etobicoke Centre* (109,250): 1. English (54.1%), 2. Italian (5.5%), 3. Ukrainian (4.1%), 4. Spanish (3.2%)
- *Etobicoke North* (105,620): 1. English (46.4%), **2. Panjabi (Punjabi) (10.1%)**, 3. Gujarati (4.0%), 4. Italian (3.8%), 5. Urdu (3.1%), 6. Somali (3.0%)

Coursera: *Geospatial Data*

We will be using the Geospatial data - Latitude and Longitude details of all the postal code of Toronto that we have received through the capstone project as well. A sample of the data is provided in the below image

	Postal Code	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476