CAPSTONE PROJECT

THE BATTLE OF NEIGHBORHOODS – A COMPARATIVE STUDY OF THE EARLIER SIX LOWER-TIER CONSTITUENT MUNICIPALITIES OF TORONTO

DATA

It's the extraction and efficient processing of datasets, collected from various sources, that makes up the bulk of a 'data scientist'. What good is a data scientist without data? Collecting and wrangling data requires a lot of practice, patience and dedication. All datasets are unique in their own way, and each one requires a new approach.

In this capstone project, I used the data from the following sources.

1. Wikipedia: Wikipedia is a repository of a huge volume of data. The data for the different boroughs of Canada is readily available by clicking on the link given below.
   https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
   The data was extracted as a pandas dataframe, in the form as shown.

	Postal code	Borough	Neighborhood
0	M1A	Not assigned	NaN
1	M2A	Not assigned	NaN
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park / Harbourfront
5	M6A	North York	Lawrence Manor / Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government
7	M8A	Not assigned	NaN
8	M9A	Etobicoke	Islington Avenue
9	M1B	Scarborough	Malvern / Rouge
10	M2B	Not assigned	NaN

   We can see that many of the boroughs have not been assigned. Hence, the next thing to be done was to drop all the rows that had no boroughs assigned. Also, any neighborhood, that didn't have a value, was assigned its corresponding borough. The slash (/) for more than one neighborhood for a given borough was replaced by a comma (,). After all this, the dataframe looked like:

	Postal code	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park , Harbourfront
3	M6A	North York	Lawrence Manor , Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park , Ontario Provincial Government
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Malvern , Rouge
7	M3B	North York	Don Mills
8	M4B	East York	Parkview Hill , Woodbine Gardens
9	M5B	Downtown Toronto	Garden District, Ryerson
10	M6B	North York	Glencairn

2. https://cocl.us/Geospatial_data

   This link provided the geographical coordinates, i.e., the latitude and longitude of each Postal Code in Canada. A snapshot of the data collected from this site is given below.

	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
5	M1J	43.744734	-79.239476
6	M1K	43.727929	-79.262029
7	M1L	43.711112	-79.284577
8	M1M	43.716316	-79.239476
9	M1N	43.692657	-79.264848
10	M1P	43.757410	-79.273304

   The data obtained from Wikipedia was merged with the dataset shown above, to get the following dataframe.

	Postal Code	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park , Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor , Lawrence Heights	43.718518	-79.464763
4	M7A	Downtown Toronto	Queen's Park , Ontario Provincial Government	43.662301	-79.389494
5	M9A	Etobicoke	Islington Avenue	43.667856	-79.532242
6	M1B	Scarborough	Malvern , Rouge	43.806686	-79.194353
7	M3B	North York	Don Mills	43.745906	-79.352188
8	M4B	East York	Parkview Hill , Woodbine Gardens	43.706397	-79.309937
9	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
10	M6B	North York	Glencairn	43.709577	-79.445073

The above dataframe was used for classifying the earlier six lower-tier constituent municipalities of Toronto.

3. Foursquare API: Foursquare is a social location service that allows users to explore the world around them. The Foursquare API allows application developers to interact with the Foursquare platform. The API itself is a RESTful set of addresses to which one can send requests, so there's really nothing to download onto the server.

   Foursquare API was used in this project to get the common venues around each of the six places, by passing in the required parameters.

   The dataset was of the following form:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Malvern , Rouge	43.806686	-79.194353	Wendy's	43.807448	-79.199056	Fast Food Restaurant
1	Rouge Hill , Port Union , Highland Creek	43.784535	-79.160497	Royal Canadian Legion	43.782533	-79.163085	Bar
2	Guildwood , Morningside , West Hill	43.763573	-79.188711	RBC Royal Bank	43.766790	-79.191151	Bank
3	Guildwood , Morningside , West Hill	43.763573	-79.188711	G & G Electronics	43.765309	-79.191537	Electronics Store
4	Guildwood , Morningside , West Hill	43.763573	-79.188711	Big Bite Burrito	43.766299	-79.190720	Mexican Restaurant
5	Guildwood , Morningside , West Hill	43.763573	-79.188711	Enterprise Rent-A-Car	43.764076	-79.193406	Rental Car Location
6	Guildwood , Morningside , West Hill	43.763573	-79.188711	Woburn Medical Centre	43.766631	-79.192286	Medical Center
7	Guildwood , Morningside , West Hill	43.763573	-79.188711	Lawrence Ave E & Kingston Rd	43.767704	-79.189490	Intersection
8	Guildwood , Morningside , West Hill	43.763573	-79.188711	Eggsmart	43.767800	-79.190466	Breakfast Spot
9	Woburn	43.770992	-79.216917	Starbucks	43.770037	-79.221156	Coffee Shop
10	Woburn	43.770992	-79.216917	Tim Hortons	43.770827	-79.223078	Coffee Shop