

# DATA

## Data Acquisition

The data containing the postal codes of Canada was obtained from [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M). This dataframe consisted of three columns - Postal Code, Borough and Neighbourhood.

The next set of data containing the latitudes and longitudes of each neighbourhood was obtained from Google Maps Geocoding API. As Google was charging for the same, the Geocoder Python package was used instead (<https://geocoder.readthedocs.io/index.html>). The one drawback with this package is that it can be unreliable; to bypass this unreliability, alternatively, the data could also be collected from [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data), where the data was available in the csv format.

The dataset that was finally created, consists of the following columns:

- Neighbourhood – Name of the neighbourhood in the borough
- Borough – Name of the borough
- Latitude – Latitude of the borough
- Longitude – Longitude of the borough

## Data Cleaning

From the first data set, only the cells that have an assigned borough are selected. The others are ignored. However, if a borough does not have an assigned neighbourhood, then it will be considered the same as the borough.

	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

The second dataset containing the co-ordinates of all the boroughs is merged with the above dataset at the postal code.

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

This merged dataframe is used to explore the boroughs of Toronto using Foursquare.