**Data description**

To solve the problem, the following data will be required:

1. ***Location of interest:*** It is mainly the address of the workplace or the center of interest based on what the whole analysis will be conducted. For this project we are considering **'301 Front St W, Toronto, ON' –** address of CN tower, as our workplace location.
2. ***Apartment data:***This the data that contains all the nearby Residential Building (Apartment / Condo) with in 5 Km distance from the location of interest. This data mainly contains the latitude and longitude coordinates of those apartments. This is required in order to plot the map and to get the nearby venue data.
3. ***Postal codes of Canada:*** ([*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 data contains the Postal codes, Borough and the Neighborhoods of Toronto city. This data was merged with the ‘**Apartment data**’ on the postal codes to check the Boroughs of each apartments within the radius of the location of interest.
4. ***Geospatial data:*** ([*http://cocl.us/Geospatial\_data*](https://cocl.us/Geospatial_data)) This data contains the latitude and the longitude coordinates of all the Postal codes of Toronto city. This data was merged with the ‘***Postal codes of Canada***’ data to get the neighborhoods with corresponding coordinates. This data is useful to find a residence based on the neighborhood amenities instead of the apartments.

Toronto city neighborhoods were chosen as the observation target due to the following reasons:

* The availability of higher job opportunity,
* The availability of geo data which can be used to visualize the dataset onto a map.

The process of collecting and clean data:

* Location of interest or the workplace data was chosen randomly at the heart of Toronto downtown by searching on google.
* Apartment data was found using foursquare API by searching for the nearby Residential Building (Apartment / Condo) with in 5 km distance. The cleaned apartments dataset is given below:



Figure 1 – Apartment dataset

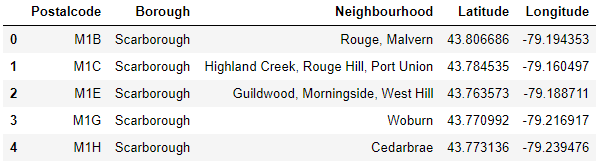
* Scraped [Wikipedia page](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) and transformed the data into a structured data frame.
* Geospatial data of Toronto neighborhoods were found from week 3 module of the IBM capstone project course and merged with the ‘Postal codes of Canada’ data. The Toronto neighborhoods dataset and the final data set for analysis are given below:

Figure 2 – Toronto neighborhoods dataset

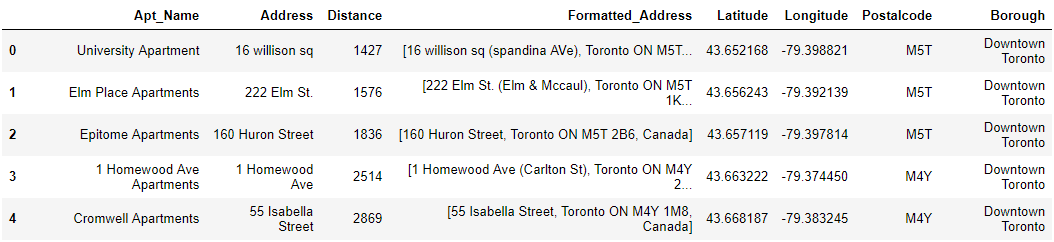


Figure 3 – Final apartment dataset in Toronto city