

## Data:

The “Snacks On Wheels” project need location data and trending venues nearby information for its analysis.

In this project, we will be required to explore, segment, and cluster the neighborhoods in the city of Toronto. For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. We will be required to scrape the Wikipedia page and wrangle the data, clean it, and then read it into a *pandas* dataframe so that it is will be in a structured format for further analysis.

1. Source: [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M),
2. This data contain Postcode ,Borough & Neighbourhood field .
3. Adjust Column header , Ignore cells with a borough "Not assigned"
4. If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough
5. If more than one neighborhood exist in one postal code area , combine them with comma separated

Once the data is in a structured format, we need to add Geospatial Data ( latitude and longitude) based on postal code. Even though geospatial data is available with google location service , due to frequent changes of google API , this project used the location data which is available on [https://cocl.us/Geospatial\\_data](https://cocl.us/Geospatial_data)

Once the current location information of “Snacks on wheel trailer” is available (postal code or borough – User will input this data or use GPS) , which is used to get the “trending venues nearby” data.

Foursquare API is used to get the trending venues nearby information. These are venues that have the highest foot traffic when the call to the database is made. Therefore the results vary depending on when the call is made. So in the morning we might find that trending venues are coffee shops or office spaces whereas in the evening trending venues can be malls, museums or parks. And to get the trending venues, we simply use the trending endpoint along with the credentials and the latitude and the longitude coordinates of the place of interest. When we make the call to the foursquare database, we get a JSON file of the trending venues that are nearby. In the JSON file, for each trending venue, we get mostly its name, unique ID, location, and category.

Request Base URL : <https://api.foursquare.com/v2/venues/trending>

The Number of results to return is limited by top 10 , with in the radius of 2000 meters. The distance between current location and trending location is calculated for further analysis. If distance is not available in Foursquare response we uses the ‘haversine’ formula to calculate the great-circle distance between two points – that is, the shortest distance over the earth’s surface – giving an ‘as-the-crow-flies’ distance between the points (ignoring any hills they fly over, of course!).

Haversine formula: 
$$a = \sin^2(\Delta\phi/2) + \cos \phi_1 \cdot \cos \phi_2 \cdot \sin^2(\Delta\lambda/2)$$

$$c = 2 \cdot \text{atan2}(\sqrt{a}, \sqrt{1-a})$$

$$d = R \cdot c$$

Where  $\phi$  is latitude,  $\lambda$  is longitude, R is earth’s radius (mean radius = 6,371km);

Note that angles need to be in radians to pass to trig functions!