

### 3. Data Overview

The data that will be required will be a combination of CSV files that have been prepared for the purposes of the analysis from multiple sources which will provide the list of neighbourhoods in Toronto (via Wikipedia), the Geographical location of the neighbourhoods (via Geocoder package) and Venue data pertaining to Korean restaurants (via Foursquare). The Venue data will help find which neighbourhood is best suitable to open an Korean restaurant.

Source: Toronto Neighborhoods via

Wikipedia [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)

For the Toronto neighborhood data, a Wikipedia page exists that has all the information we need to explore and cluster the neighborhoods in Toronto. It included the postal code, borough and the name of the neighbourhoods present in Toronto.

Using the Geocoder package, here is a link to a csv file that has the geographical coordinates of each postal code: [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)

Using Venue Data using Foursquare, grouped by the name of the neighbourhood with venue category. <https://developer.foursquare.com/docs/api-reference/venues/links/>