**Data Requirements and Collection technique**

1. The primary datasource that will be used for getting Toronto neighbourhoods is

https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M

We will extract the details by web scrape technique from the wikipedia page

2. The latitude and longitude data will be used from IBM assignment csv.

<http://cocl.us/Geospatial_data>

3. We will be using four square client to get all the near by venues for the updated list of neighbourhoods of Toronto that we get after data cleaning .

4. We are planning to use only few categories of venues from four square venue category list to keep the data limited to our specific business query only. Below are those:

- 4d4b7105d754a06372d81259 :college and university

- 4bf58dd8d48988d13b941735 : school

- 4bf58dd8d48988d1ed931735 : airport

- 52f2ab2ebcbc57f1066b8b4f : bustop

- 4bf58dd8d48988d129951735 : Train station

- 52f2ab2ebcbc57f1066b8b51 : Tram station

Note: These are main categories. This also includes several subcategories which will be filtered out later based on our requirement.