

2. A description of the data and how it will be used to solve the problem.

The data that we have used in this project is obtained from Wikipedia and Foursquare. We have obtained the list of localities or neighbourhoods of Toronto and Melbourne from the following Wikipedia pages and website, as well as their geo-location data like latitude and longitude.

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

https://en.wikipedia.org/wiki/Postal_district_numbers_of_Melbourne

http://www.mapdevelopers.com/batch_geocode_tool.php

These data is then scraped using the Beautiful Soup library in python, and then formatted into a pandas dataframe for further processing.

Foursquare provides detailed data relating to each of these neighbourhoods. Foursquare is a local search-and-discovery service mobile app which provides search results for its users. Foursquare API will be used to explore the various types of venues and their categories available in each neighborhood. The information for venues, categories, hotels, etc is extracted using foursquare to solve the business problem. This data is to be used for segmenting and clustering the localities of Melbourne. This will help to detect the similarities and dissimilarities of neighborhoods. We find the cluster which is most similar to Rosedale and then obtain a list of some of the hotels which are located in this cluster.