Data Description

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

We will use Scarborough dataset which we scrapped from wikipedia on Week 3. Dataset consisting of latitude and longitude, zip codes. To compare the similarities of two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods i. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm.

We will need the information retrieved from Foursquare as follow: Neighborhood, Neighborhood Latitude, Neighborhood Longitude, Venue, Name of the venue, Venue Latitude, Venue Longitude, Venue Category. While analysis, we will combine these data with which we scrap from wikipedia.

Libraries and API for project:

Foursquare API: perform location search, sharing and obtain details about business based on its database of millions of places.

Numpy and Pandas: basic libraries to manipulate series and dataframe

Matplotlib: data visualization

BeautifulSoup and requests: scrap data from webpages

Sklearn: perform machine learning, concretely, the k mean cluster

Folium: construct an interactive map