Description of Data

The location data will be collected from 2 separate data sources. The Toronto data will be scraped from the following webpage:
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

This data will then be combined with the Location Data for NYC, which will be taken from the following:
https://cocl.us/new_york_dataset

Next the Foursquare API will be used to explore all venues within a 500-metre radius of each borough. The top 10 most common venues in each borough will then be used to assign clusters to each borough using K-means clustering. Finally, the application will be able to suggest similar boroughs between NYC and Toronto based on the results of the clustering.