Moving to Toronto: A Study in the Livability Factor of the City

Data:

As discussed in the previous section on what the problem is, we will need to analyze what makes a neighborhood a livable neighborhood by categorizing them into different clusters. The initial data we will need will be a table of neighborhoods and their postal codes along with the borough that they belong to within Toronto. This data can be easily found on Wikipedia[1].

This data only contains the postal codes, neighborhoods and boroughs though, so to do a proper analysis we will need to somehow grab a list of venues that are nearby to the postal code's locations. Luckily, Foursquare has this data readily available, but before we can grab this data, we need the latitude and longitude co-ordinates, which we can grab from the python geocoder module. Alternatively, if the geocoder module does not work, a csv with the latitude and longitude locations of all the postal codes in Toronto was provided to us by the IBM team at coursera, so we elected to use that data since, in our case, the geocoder module did not work properly.

Once all the latitude and longitude coordinates were in the database, we could do a call to the Foursquare API to collect a list of attractions that were within a certain set radius (in our case, 250 meters). Once this data was collected, we stuck it all into a pandas data frame and started our analysis.

1