Introduction

Moving can be a stressful and tedious task. Packing & unpacking, arranging a moving company, finding the right time, cost etc. are all part of this challenging task. However, these hurdles to overcome do not have long lasting effects. For many people, the decision of "where" to move is the hardest one among all as it will have an enormous effect on their well-being during their stay in the place they moved in. One way to ease their pain on making this decision would be to cluster neighborhoods based on their similarity to each other and select from the ones that are similar to your place or to a place that you consider to be a decent place. This way, a very long list of places would significantly be narrowed down.

Business Problem & Target Audience

Say that you are a freelancer that helps customers to make a decision on where to move. In this particular case, you have a customer who currently lives in a neighborhood in Manhattan, New York and wants to move to a neighborhood in Toronto, Canada as he found a job there. The customer is very happy about his current place and wants to find a neighborhood in Toronto that is similar to his current place in terms of activities he can enjoy such as restaurants, cafes, museums etc. He asks our help to narrow down a list of potential places to move in. We will make clusters of neighborhoods including his current neighborhood and neighborhoods in Toronto. And he will select from places that are in the same cluster as his current place.

Data & Source

The following data will be required for this problem:

- List of neighborhoods in Toronto
- Coordinates of those neighborhoods as well as coordinates of his current neighborhood
- Top venue data of each neighborhood for clustering

The list of neighborhoods in Toronto will be extracted from the following Wikipedia page:

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

The data will be transformed into pandas data frame and will be processed into a proper version (data cleaning, wrangling etc.)

In order to get the venue data, we will use Foursquare API based on the coordinates of the neighborhoods, select the top venues, and make a few transformations on the data to be used for clustering.