



Relocation Recommendation from Delhi to Toronto Using Neighbourhood Clustering

Prepared for: Coursera Capstone

Prepared by: Naman Sharma

28 April 2020

DATA SELECTION AND PREPARATION

Based on definition of our problem, we will gather data for both the cities, the city of Delhi and the city of Toronto. We will consider following factors before we begin are search for data:

- Borough nearby Office/University location in the city of Toronto.
 - **FOR OUR PROJECT PURPOSE WE ASSUME THAT THESE BOROUGHS ARE NORTH YORK AND SCARBOROUGH**
- Similar neighbourhood in Toronto as in Delhi.
- Picking up data pertinent to following factors:
 1. Neighbourhood: Name of the neighbourhood in the Borough.
 2. Name of the Borough.
 3. Latitude of the Borough.
 4. Longitude of the Borough.
 5. Venue nearby our neighbourhood.

In our project we will gather and process data with following techniques:

- Acquire the names and boroughs of the neighbourhoods by scrapping a wikipedia pages and assigning them in "csv" coded data frames.
 - After we have the names of all the neighbourhoods, we will geocode them using the geopy.geocoder (Nominatim). And thus will add longitude and latitude data corresponding to borough location and neighbourhood data.
 - Next, we use the foursquare API to find all types of venues within a 500 meter radius for every neighbourhood. Then these are merged with the existing data set
 - Normalise the data using one hot coding and finally preparing it for cluster formation.
 - Merging the One hot Coded data and the Most Frequent venue data for proper visualisation of problem.
-