Finding the best Neighborhoods for Foreign Students in Toronto

1. Description of The Problem and a Discussion of the Background

Toronto being the capital city of the Canadian province of Ontario has been recorded as one of the most populous city with a population of 6,254,571 as of 2021 in Canada and the fourth most populous city in North America. The city is surrounding the western end of Lake Ontario is an international center of business, finance, arts, and culture, education, and is recognized as one of the most multicultural and cosmopolitan cities in the world. There are 140 neighborhoods officially recognized by the City of Toronto and upwards of 240 official and unofficial neighborhoods within the city's boundaries. These are Chinatown, Corso Italia, Little India, Kensington Market, Little Italy, Koreatown, and many more.

The immigrants during their initial few years and the foreign students always need to find a convenient locality to live which is close to the educational institutions, restaurants of their choices, and shopping facilities. The main objective of this Capstone Project is to identify the neighborhoods of Toronto for the foreign students from India and Pakistan, particularly, to find a reasonable cost effective residence which is near to their universities, Indian or Pakistani dining facilities, and the shopping malls. There are many cuisines which are difficult to distinguish as Indian or Pakistani and therefore it will mainly serve for both types of foreign students.

We would like answer the question *"What is the best convenient neighborhood for a Pakistani or Indian foreign students in Toronto?"*.

2. Description of the Data and how it will be used to Solve the Problem.

The data of Toronto available from Wikipedia page that contain Borough, Neighborhoods, and that has all the information we need to explore and identify the neighborhoods in Toronto. However, this data is not in a format that could be directly used for my Capstone Project. The data must be preprocessed before using it in any project of extracting the useful and required information and knowledge. The data of Toronto on Wikipedia page will be scrapped, wrangled, clean it, and then read it into a pandas dataframe so that it is in a structured format. The additional information of latitudes and longitudes of each neighborhoods will combined with the data from https://cocl.us/new_york_dataset.

The Foursquare API will be used to find and filter the neighborhoods for convenient residences for the foreign students from Pakistan and India. The Machine Learning algorithm i.e. Clustering will be used to generate the clusters of interests in this project including universities, Indian and/or Pakistani restaurants, and shopping malls.