The Battle of Neighborhoods

Applied Data Science Capstone Project

Wei Wang

1. Introduction

1.1 Background

According to UN Department of Economic and Social Affairs (DESA), there are nearly 272 million international migrants worldwide in 2019. Toronto and Shanghai, two major well known cities in Canada and China, are the economic centers and financial capitals of their countries. In 2019, Toronto welcomed 117,000 immigrates while Shanghai's population is growing at the rate of 700,000 people a year. Despite the fact that Toronto and Shanghai are far away from each other, it is still possible to find the similarities between these two cities.

1.2 Business problem

Relocation can be considered a big decision for a person. In addition to job opportunities, living conditions, environment and culture shock are also big concerns for migrants when moving to another city. Therefore, it is advantageous for them to find a similar neighborhood in the new city as the one they live in before. In this project, we will adopt machine learning tools to cluster Toronto and Shanghai neighborhoods in order to recommend the neighborhoods which are the best choices for migrants based on surrounded essential facilities such as school, hospital, stores and parks, etc.

As a result, the audiences of this report will be anyone who is planning to move from one city to another. This can also benefit stakeholders who are interested in determining how similar or dissimilar two cities are.

2. Data

2.1 Neighborhoods

We will be using two datasets for this project. The first dataset consists of Toronto's boroughs, neighborhoods and their respective postcodes. The dataset is scraped from Wikipedia page. The second dataset consists of Shanghai's city name, districts and subdistrict. The dataset is scraped from Wikipedia page.

PostalCode	Borough	Neighbourhood
M3A	North York	Parkwoods
M4A	North York	Victoria Village
M5A	Downtown Toronto	Regent Park / Harbourfront
M6A	North York	Lawrence Manor / Lawrence Heights
M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government

City	District	Subdistrict
Shanghai	Baoshan	Wusong
Shanghai	Baoshan	Youyi
Shanghai	Baoshan	Zhangmiao
Shanghai	Baoshan	Dachang

Toronto Dataset   *Shanghai Dataset*

2.2 Coordinates

We will use a csv file that has the geographical coordinates of each postal code in Toronto to get the latitude and the longitude coordinates of each neighborhood of Toronto.

Postal Code	Latitude	Longitude
M1B	43.8066863	-79.1943534
M1C	43.7845351	-79.1604971
M1E	43.7635726	-79.1887115
M1G	43.7709921	-79.2169174

Coordinates in Toronto

For coordinates of each neighborhood in Shanghai, we will use Nominatim API.

2.3 Venues

Foursquare API search feature will be used to collect neighborhood venues information, which will be used to explore and compare geographical locations of Toronto and Shanghai.