**Determining Similarity Between Cities and Recommending locations for Business Owners**

**1.Introduction**

**1.1 Background:**

To compare the similarity between two biggest financial cities of the world Toronto and New York. This similarity well be compare based on the neighborhoods and venues nearby like hospital, entertainment, school, cafes, outdoor activities etc. This comparison will help people in migrating to either countries. People always have questions before migrating to any place like good schools for kids, hospitals, Jobs, entertainment, safety, public transport etc. This model will help give them all the answers and help them in determining whether to move or not. Target audience could be financial analysts or people from financial business who constantly move from one place to another. Second audience could be general public who is thinking to move to different country Canada especially in Toronto.

**2nd problem statement** is for business owners/entrepreneurs who wants to start a business in either cities. Model will help in determining which location is more appropriate for their business. This is very critical for business owners if opened in wrong place can led to a disaster results and can loss million of dollars. This model will not save their millions of dollars but provide them all the strategic information in quick time which help them in determining appropriate location. Target audience is business owners, industrialist or entrepreneurs, small business.

**1.2 Objectives:**

**1.2.1 Similarity between cities**

This project aims to predict similarity between two financial cities include schools, entertainment, universities, hospitals, restaurants etc. which helps people to make strategic decision whether to migrate or not.

**1.2.2 Determining Location**

Second aim of this project to provide appropriate location to start their business which helps business owners, industrialist etc. to take strategic decision.

**2. Data Sources**

For this project we will be using below data sources:

* Four square data location data
* Extracting Toronto data from Wikipedia.
* Geospatial Data

**2.1 To get the New York neighborhoods and their longitude and latitude** [**https://cocl.us/new\_york\_dataset**](https://cocl.us/new_york_dataset)

Data from this link will be in JSON format will be converted into pandas data frame. This data has 4 attributes namely Borough, neighborhoods, latitude and longitude and 306 rows. 306 unique neighborhoods which will be classified in one of the boroughs.

**Borough Neighborhood Latitude Longitude**

Manhattan Marble Hill 40.876551 -73.9106601

Manhattan Chinatown 40.715618 -73.9942792

Manhattan Washington Heights40.851903 -73.9369003

Manhattan Inwood 40.867684 -73.9212104

Manhattan Hamilton Heights40.823604 -73.949688

* 1. **Four square New York location Data**

Four square location API will help in getting different venues like hospitals, education, restaurant data etc. to explore each individual borough and neighborhoods.

* 1. **Toronto data extracted from Wikipedia**

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>'

|  |  |  |  |
| --- | --- | --- | --- |
|  | Post Code | Borough | Neighbourhood |
| 0 | M3A | North York | Parkwoods |
| 1 | M4A | North York | Victoria Village |
| 2 | M5A | Downtown Toronto | Harbour Front |
| 3 | M5A | Downtown Toronto | Regent Park |
| 4 | M6A | North york | Lawrence Park |

* 1. **Geospatial\_Coordinates CSV file** provided by this course itself.

There are three attributes Postal code, Latitude, Longitude.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Post Code | Latitiude | Longitude |
| 0 | M3A | 43.806686 | -79.194353 |
| 1 | M4A | 42.806686 | -79.209876 |
| 2 | M5A | 43.780987 | -79.194345 |
| 3 | M5A | 43.809686 | -79.198452 |
| 4 | M6A | 43.807645 | -79.197612 |

* 1. **Four square Toronto location Data**

Four square location API will help in getting different venues like hospitals, education, restaurant data etc. to explore each individual borough and neighborhoods.