

Introduction

In this **capstone project** I will be creating a program to compare the neighbourhoods of the two cities namely **Brooklyn** from New York and **Downtown Toronto** from Toronto and determine how similar or dissimilar they are.

I will be comparing all the venues , hotspots to visit, number of residential friendly places, etc. As we know both Brooklyn and downtown Toronto are very huge and diverse containing people from all around the world. These places have a lot of **restaurants, cafes, pubs** and so on. So by comparing these two neighbourhoods we can find a lot of similarities and see how these places vary.

I will use **foursquare api** to collect all the details and explore these regions and will be using **k means** to cluster different categories of data and display it on the exact location using **folium maps**.

Later I will be using the two maps to differentiate between the neighbourhoods and see how one varies from the other. This analysis is mainly done when people are moving from one place to another for various purposes. *They would want to have the same facilities in the new neighbourhood as they have in their current neighbourhood.*

Thus by comparing various major cities based on its amenities and venues we make a proper analysis and visualize the data obtained using **Data Science**.

Data section

As I said in the introduction I will be creating a program to analyse Brooklyn and Downtown Toronto. I will be web scraping to obtain the table of **Canada's neighbourhoods** from Wikipedia and converted it into a Dataset using pandas and cleaned it so that it fits for my analysis.

Later I will make new data frame which comprises of **Downtown Toronto** data only.

Link had been obtained from previous labs from coursera:

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

I will also be needing the **New York dataset** so that I can extract all the information from **Brooklyn**. I obtained This data set from the lab as well:

https://cocl.us/new_york_dataset

The above json file will be cleaned and only a dataframe containing only Brooklyn data will be formed.

I will also require a **geospatial coordinate** dataset which contains all the latitude and longitude information of the cities. This link was also available in the labs.

http://cocl.us/Geospatial_data

I will be using **jupyter notebook** to create the program. **Foursquare api** will be used to explore the data.

All the modules which I will be using in the program are listed below:

- 1. Numpy**
- 2. Pandas**
- 3. Matplotlib**
- 4. Seaborn**
- 5. Json**
- 6. Geopy**
- 7. Requests**
- 8. Sklearn**
- 9. Folium**