

THE BATTLE OF NEIGHBORHOODS

FIND THE SIMILAR CITIES

By - Jayesh Khandelwal

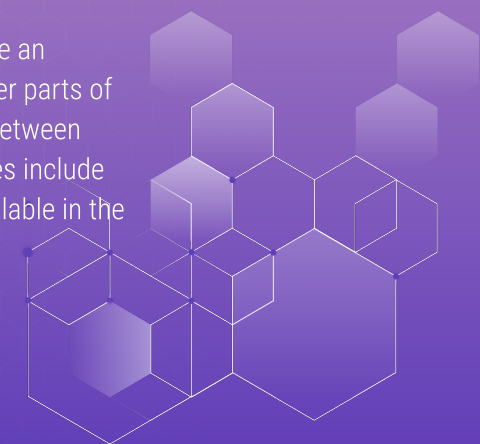


INTRODUCTION

The purpose of this Capstone project is to find similar cities around the world by exploring the venues in its neighborhood. It will help people make smart and efficient decisions on selecting places when they move around the globe. Here we find similar cities between London , New York and Toronto.

These three are the most densely populated cities around their region. Lots of people move around these cities. This project is for people to move around the world and still find the similar places to those which they are used to living.

This Capstone project aims to create an analysis for people migrating to other parts of the world by comparative analysis between neighborhoods of cities. The features include summations of kinds of venues available in the neighborhood.



The background is a gradient of purple and blue, overlaid with a pattern of white-outlined hexagons and 3D cubes. Some hexagons and cubes are filled with a lighter shade of purple, while others are just outlines. Small dark blue dots are scattered throughout the geometric shapes.

PROBLEM STATEMENT

Compare the neighborhood of different cities and find out which on them are more similar.
Which among London , New York City and Toronto are more similar?

DATA

Data link for London -

<https://www.geonames.org/postalcode-search.html?q=london&country=GB>

Data link for New York City -

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Data link for Toronto -

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





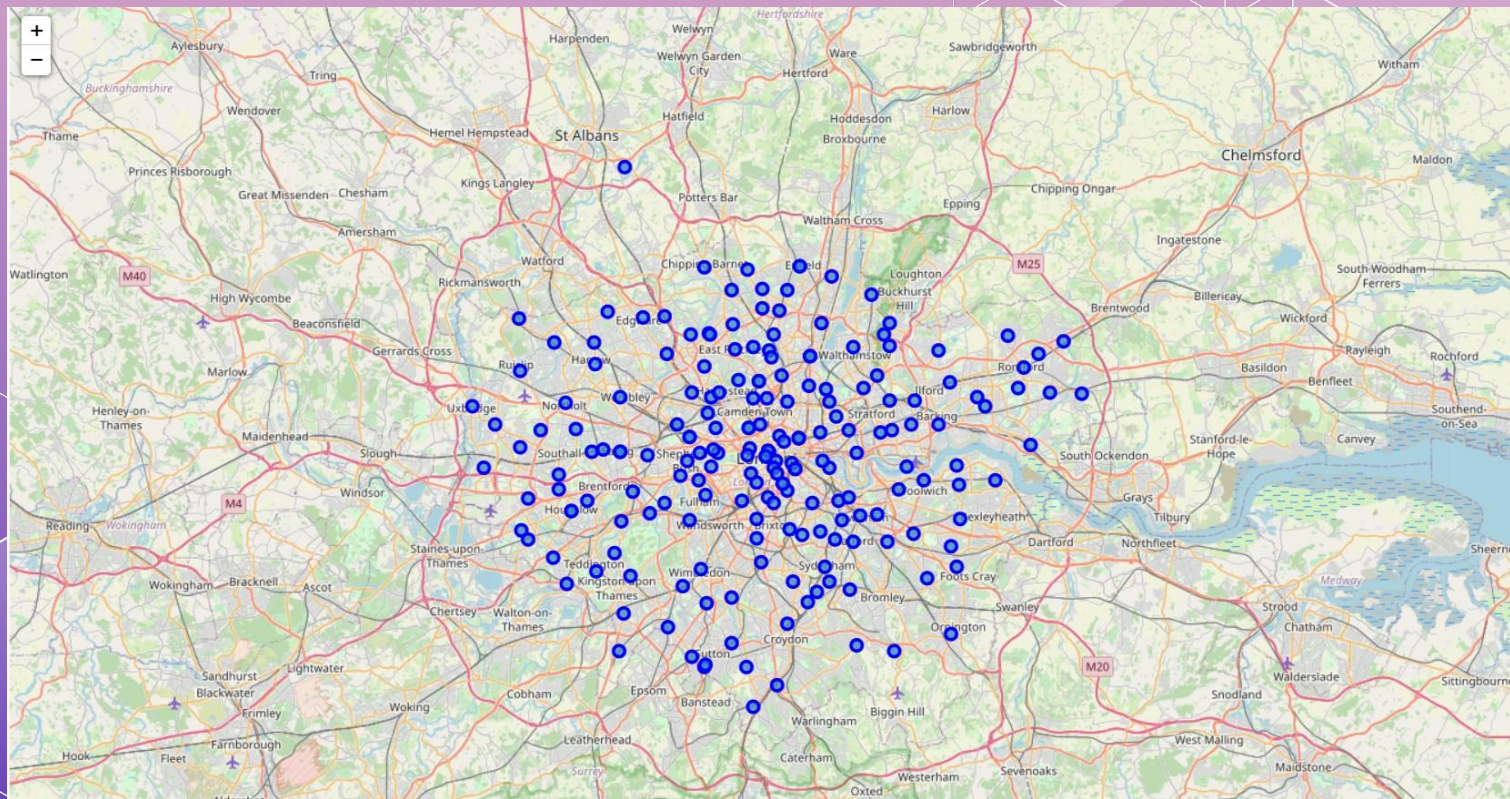
FourSquare API

FourSquare API is used to gather the data required. It has all data about venue , location , longitude , name , rating , reviews , photos and lot more.

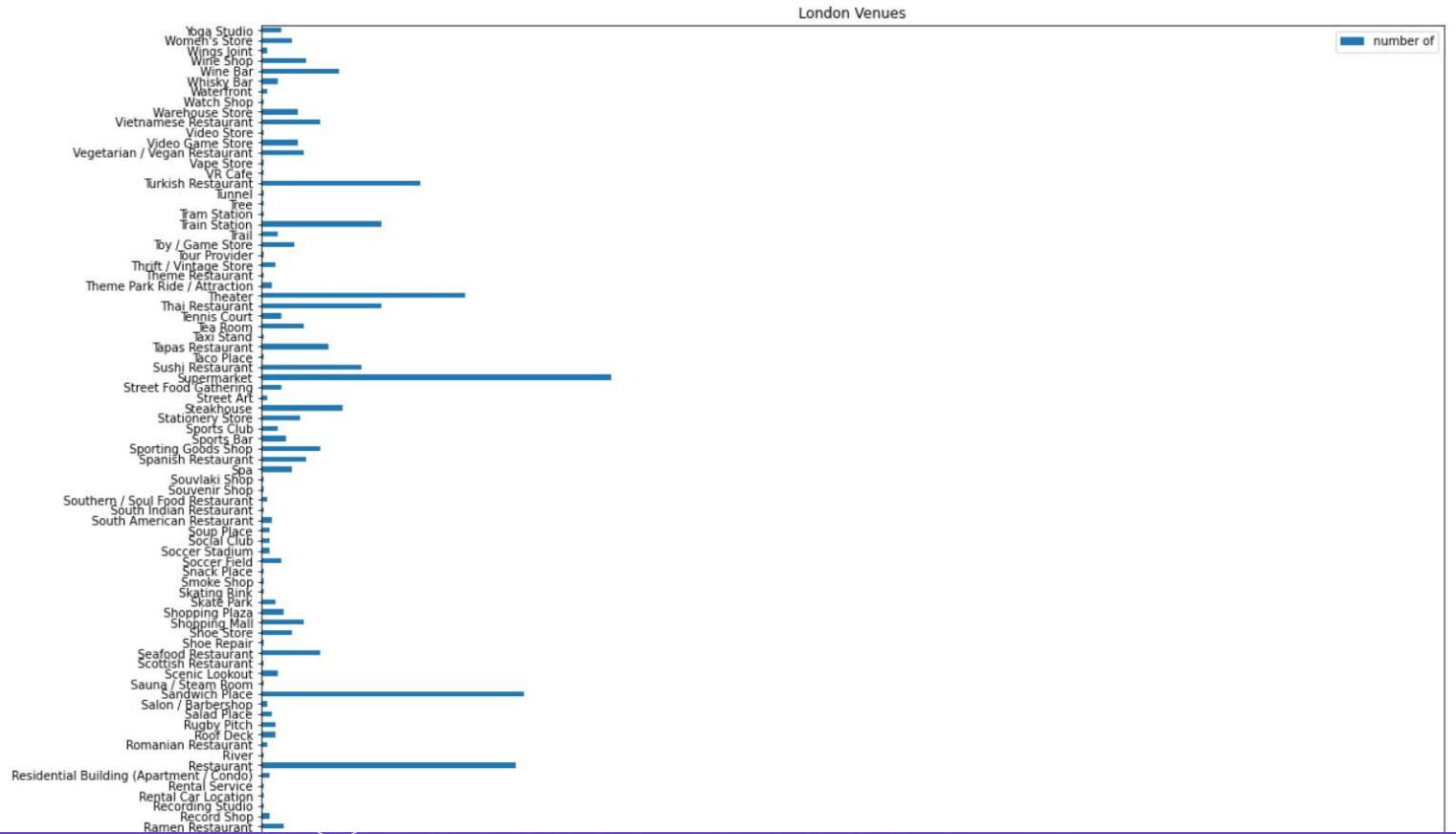
We connect our neighborhood data set we generated to create a data set :

1. Neighborhood
2. Neighborhood Latitude
3. Neighborhood Longitude
4. Venue
5. Venue Name
6. Venue Latitude
7. Venue Longitude
8. Venue Category

Map of London



London Venues



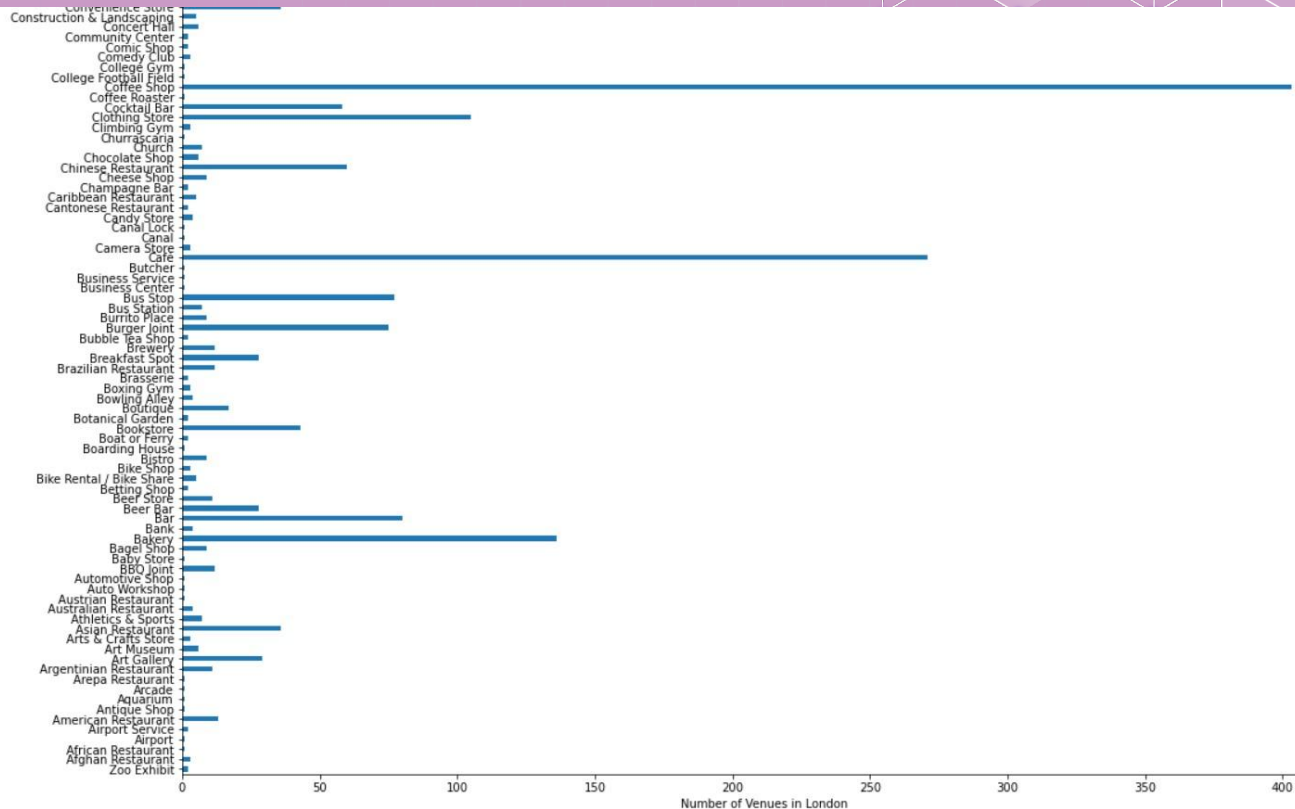
London Venues



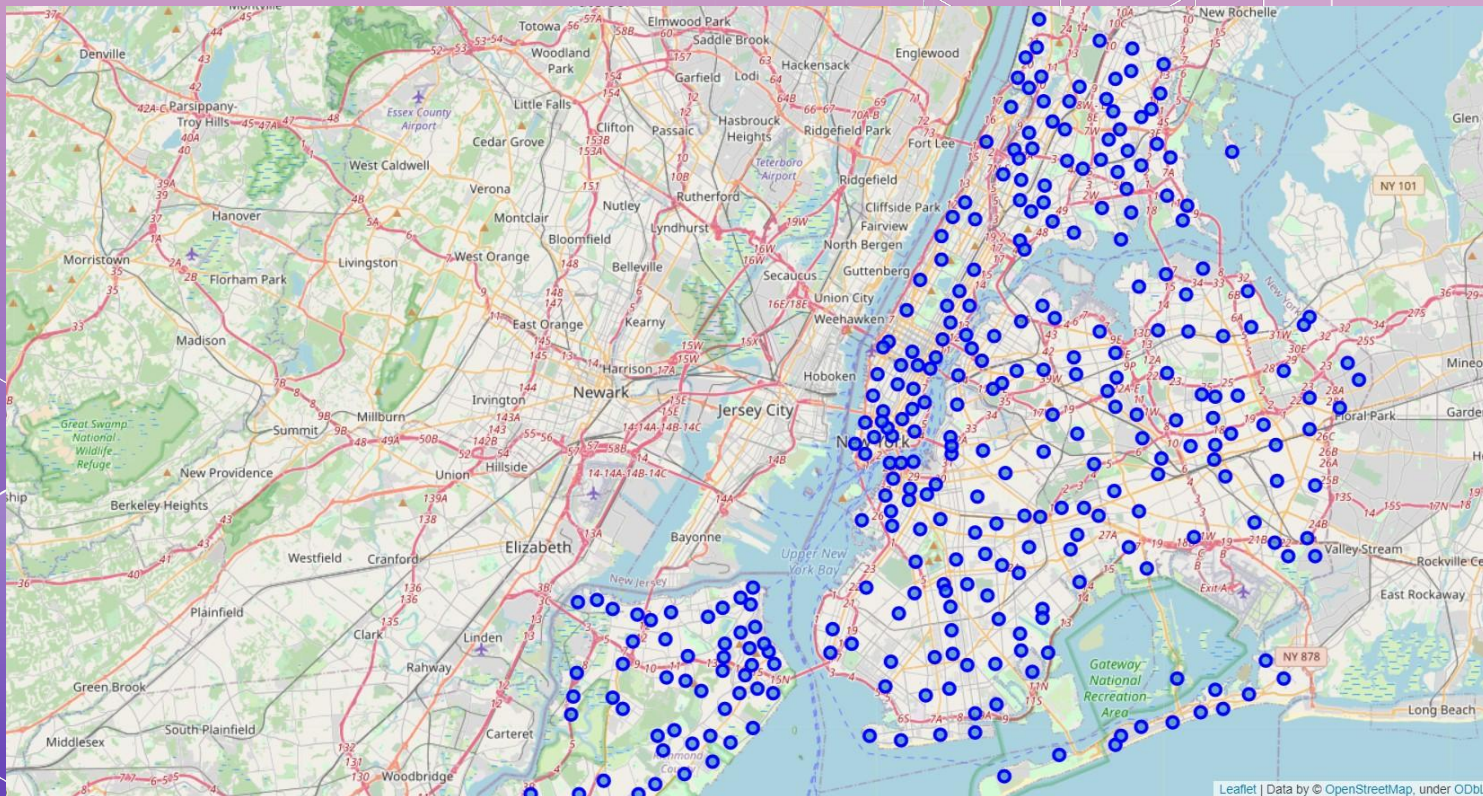
London Venues



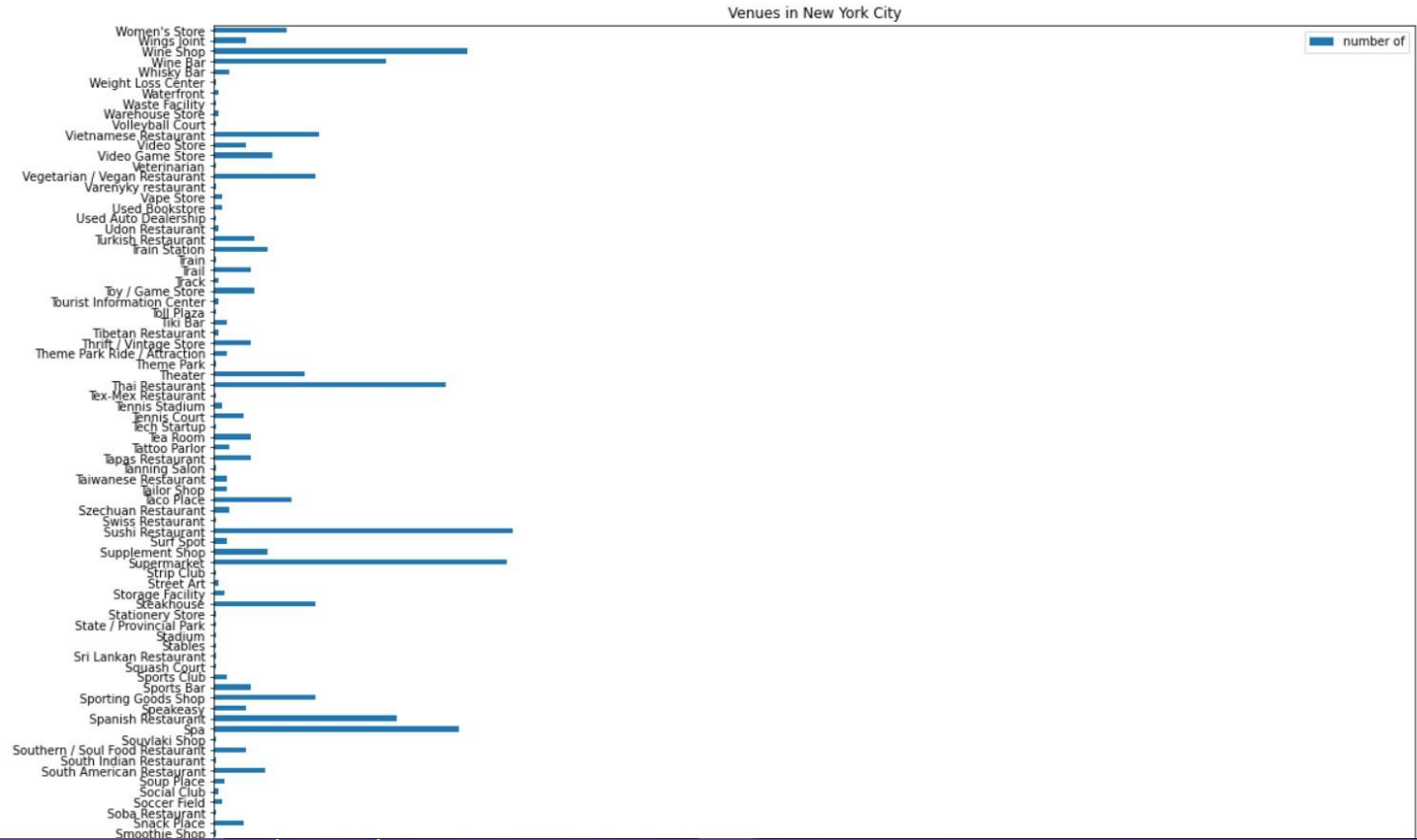
London Venues



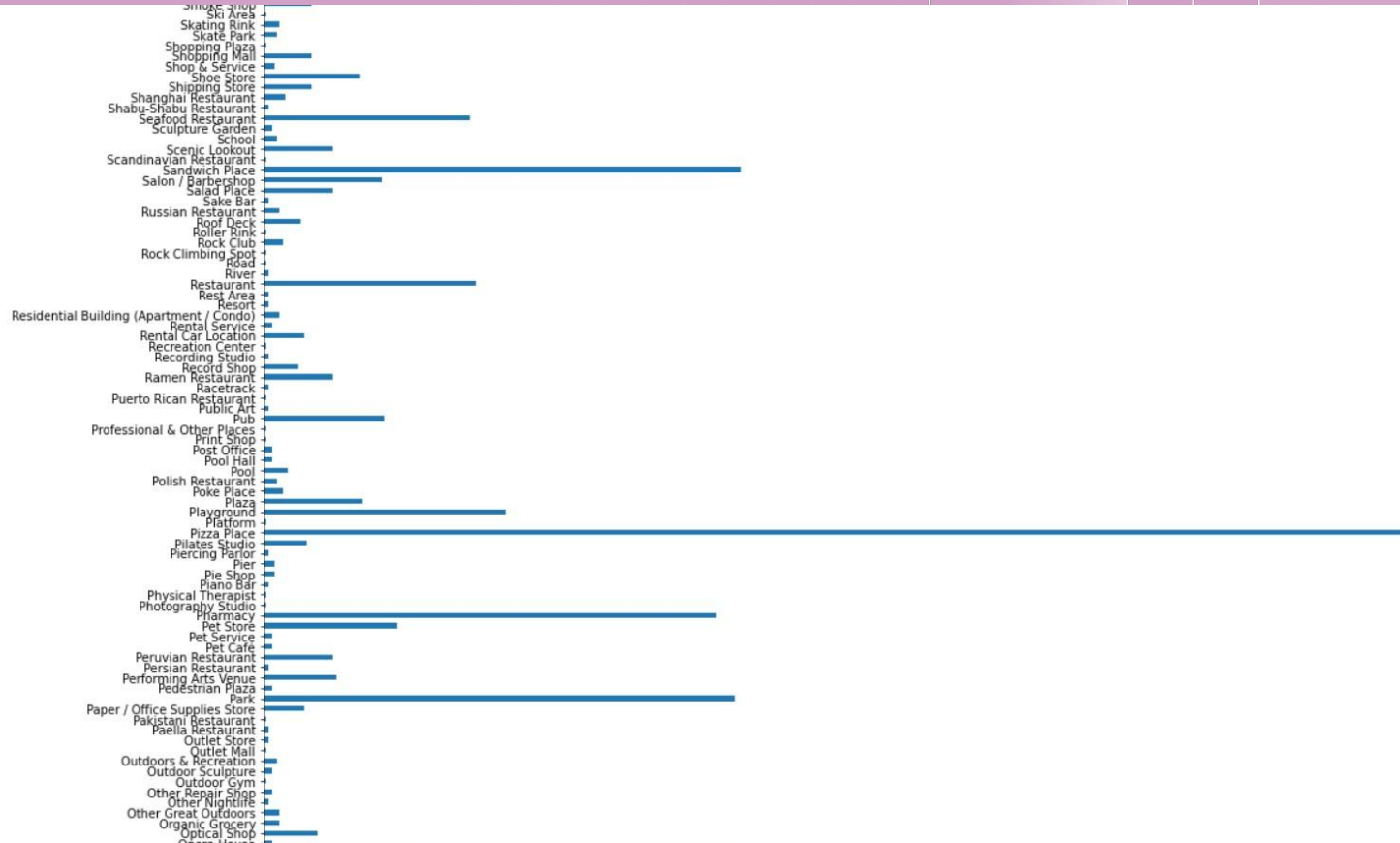
Map of New York City



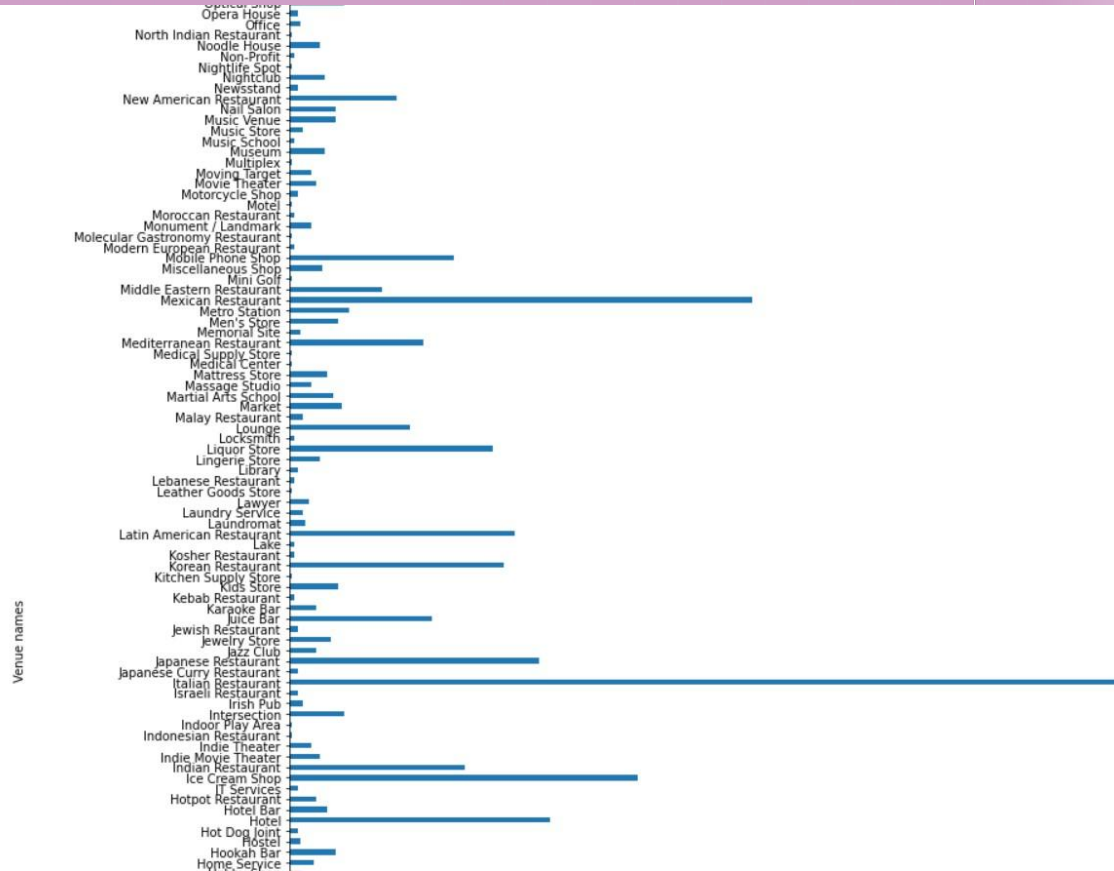
New York Cities Venues



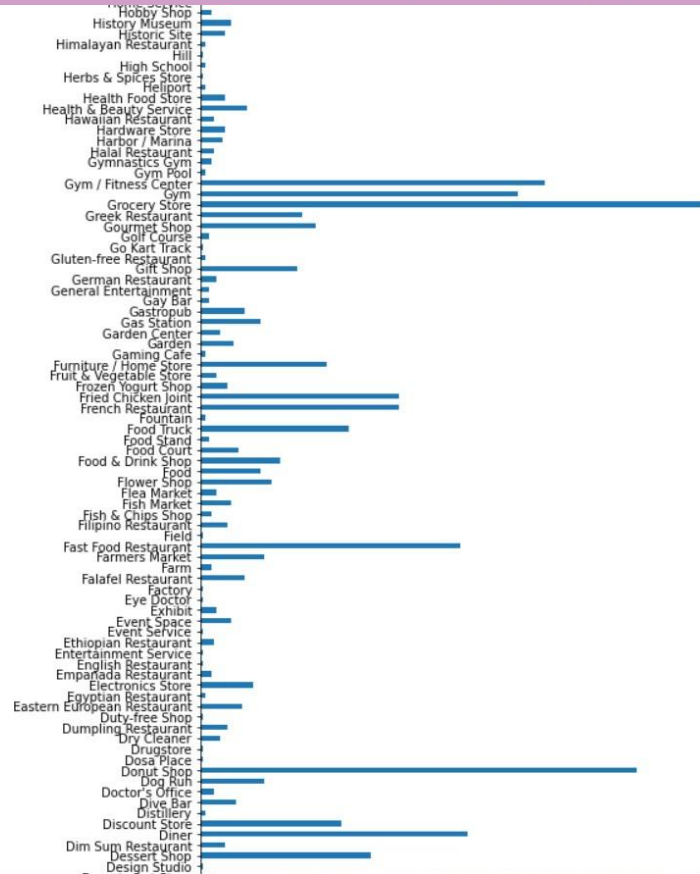
New York Cities Venues



New York Cities Venues



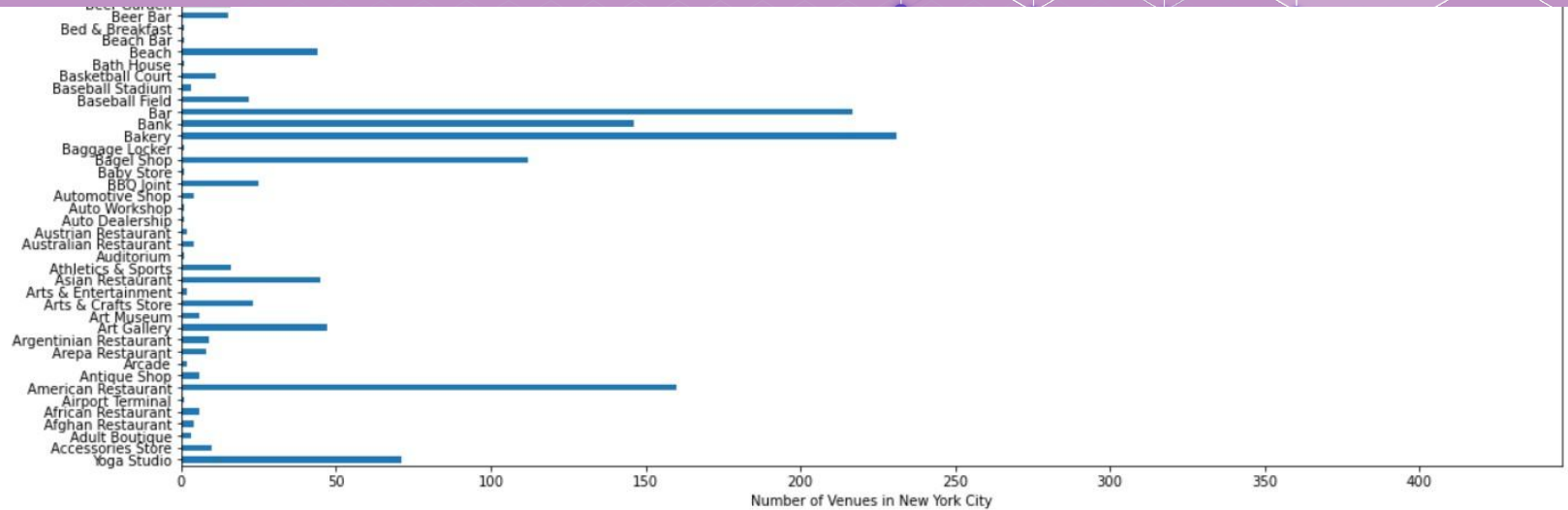
New York Cities Venues



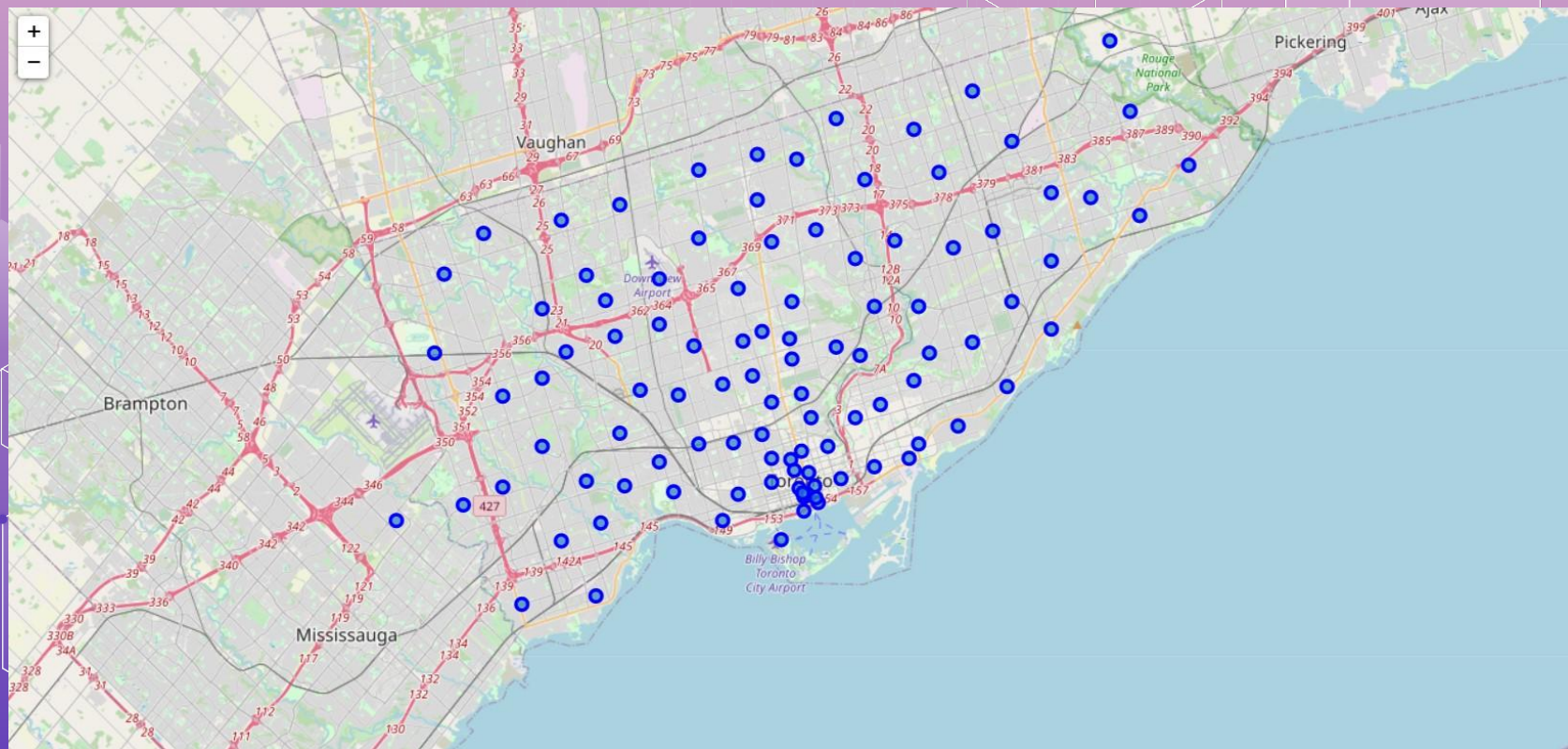
New York Cities Venues



New York Cities Venues



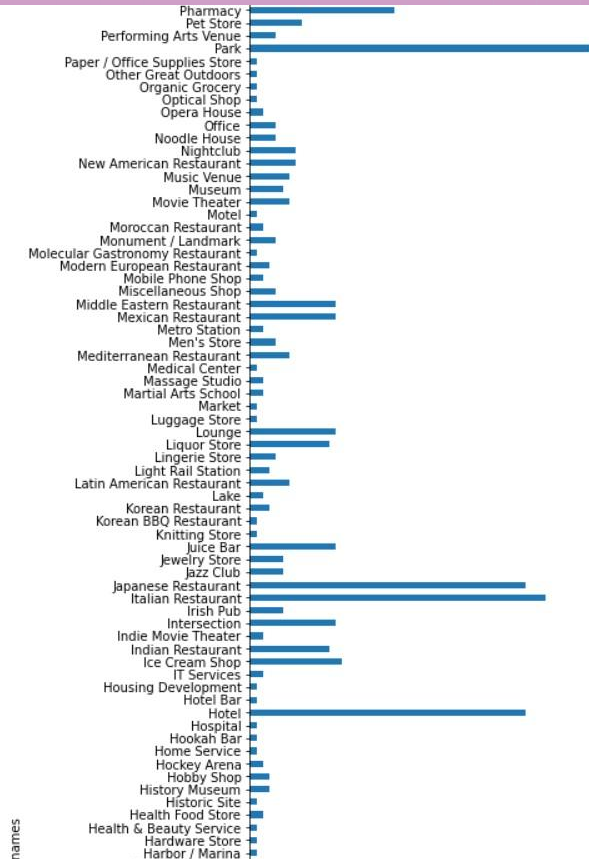
Map of Toronto



Toronto Venues

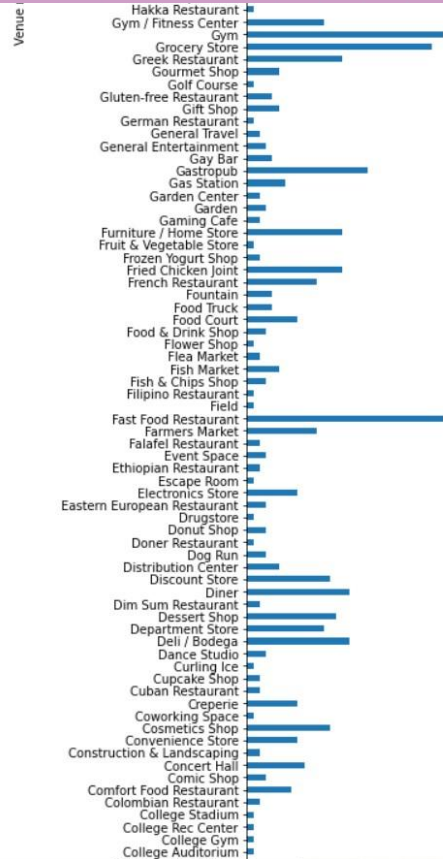


Toronto Venues

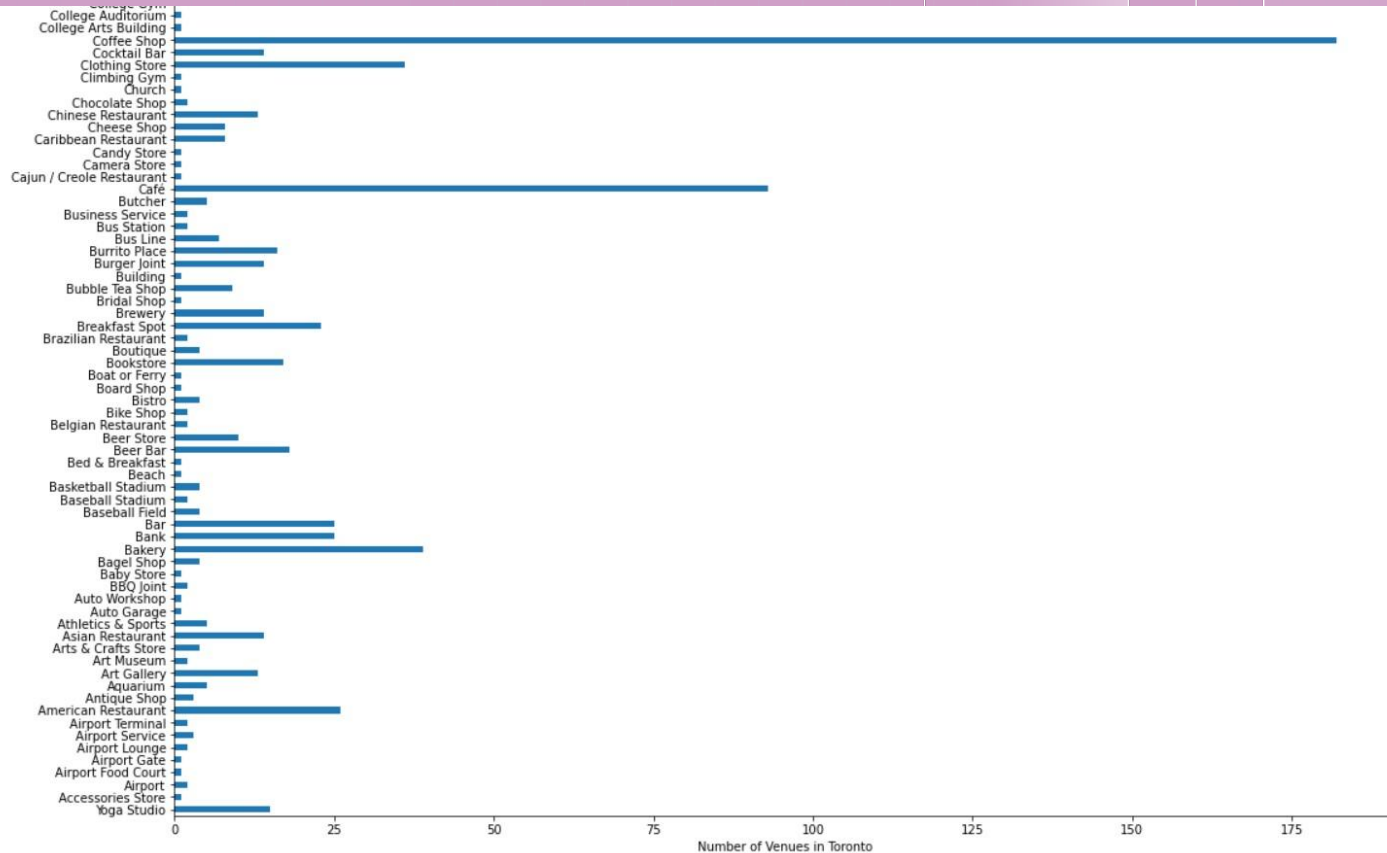


names

Toronto Venues



Toronto Venues





Methodology

We get the data set by FourSquare API. Then we used OneHot Coding Feature to get the count of individual venues in a neighborhood.

After achieving that we sum up to get the total venues in a particular cities.

Then we use cosine similarity to find the most similar cities.

COSINE SIMILARITY

London & New York
City

0.702866909

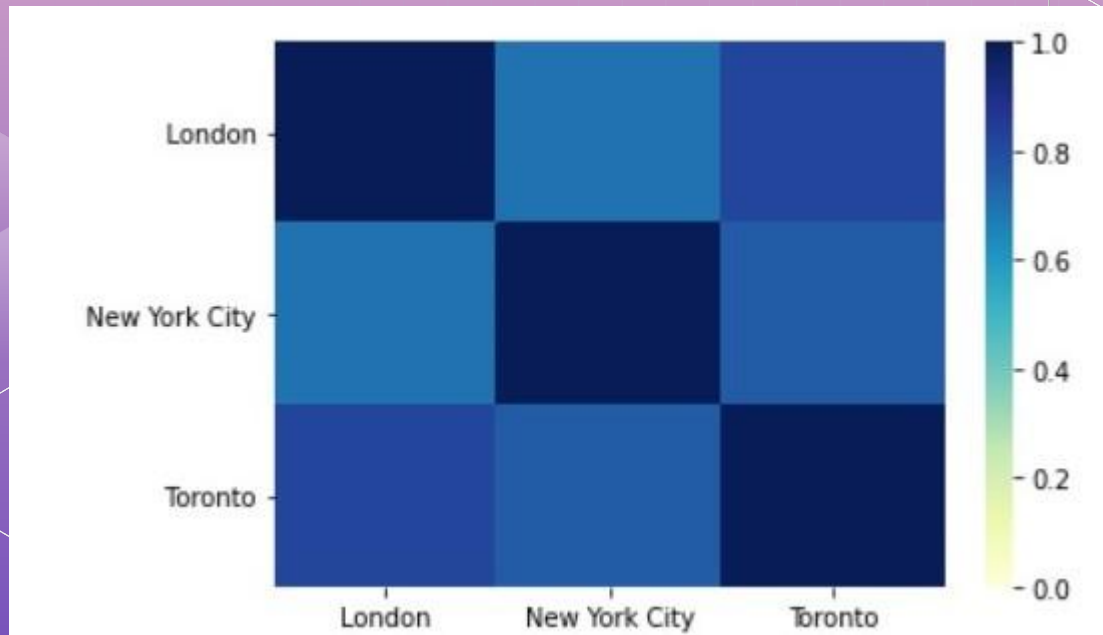
New York City &
Toronto

0.7560195483

Toronto & London

0.8203776641

Results



Thus we can analysis that London and Toronto are more similar.

Libraries Used

Numpy - to handle data in a vectorized manner

Pandas - For data Wrangling

Folium - Visualization of Map

JSON - To handle JSON files

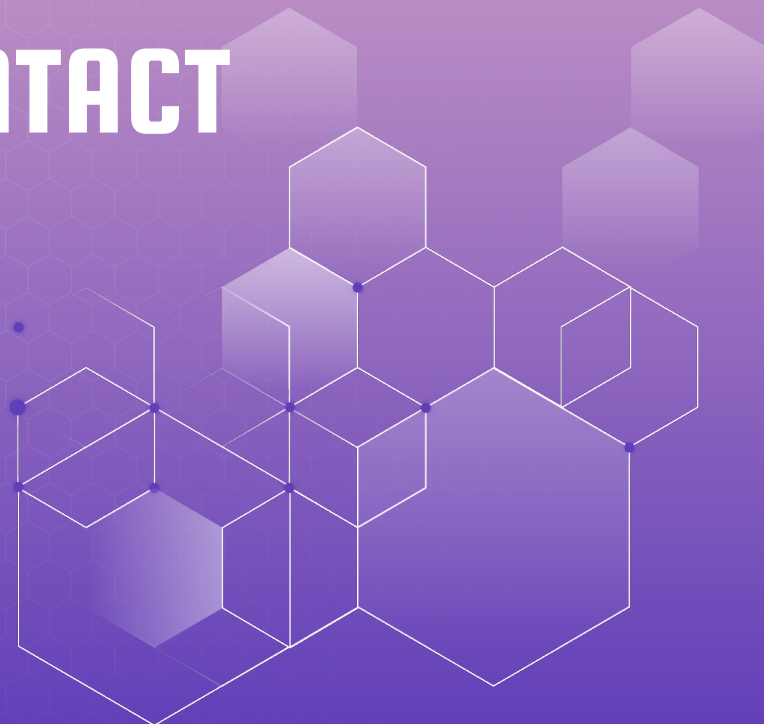
Geocoder - To retrieve Location Data

Matplotlib & Seaborn - Plotting Module



CONTACT

khandelwal.4@iitj.ac.in
+91 9429429446
<https://github.com/JayeshK10>





THANKS