The Battle of Neighborhoods

APPLIED DATA SCIENCE CAPSTONE

Introduction

New York and Toronto are both cities with global recognition and attract a diverse set of people from different parts of the world. Being ethnically diverse cities requires them to cater to a vast variety of people hailing from different countries. New York having an immigrant population of 3.1 million as of 2018 and Toronto having a population of roughly 1.5 million as of 2017, both cities host people from a multitude of cultures and ethnicities.

With a diverse culture comes diversity in cuisine. This project looks into the availability of restaurants of varying cuisines in both cities. Both have restaurant types ranging from Italian, Chinese, Indian to American, French and German. Through this project, the magnitude and availability of these different types of restaurants is presented which can be useful for any person looking to move to either of the cities or wanting to open a new restaurant in either city.

Problem:

This project aims to achieve the following:

Compare the top 10 cuisines in both New York and Toronto

Provide a view of similar neighborhoods in terms of cuisine in both cities.

Data

For this project, the following data has been used:

New York City Data containing information on boroughs, neighborhoods and their geographical coordinates

Source: https://cocl.us/new_york_dataset

Description: The data is used in conjugation with Foursquare API to zero in on each venue of New York City and identify the same on a map.

Toronto City Data containing information on boroughs, neighborhoods and their geographical coordinates

Source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M for neighborhoods and boroughs.
https://cocl.us/Geospatial_data for geospatial coordinates

Description: The data is used in conjugation with Foursquare API to zero in on each venue of Toronto City and identify the same on a map.

Methodology

We begin by extracting data for New York City using the following link:

https://cocl.us/new_york_dataset

Using the geopy library, we obtain latitude and longitude values of New York City and print a map of New York using folium

After that, the venues of each neighborhood are extracted using Foursquare API.

Following this, a one-hot representation of different venues in New York is created. This is grouped by mean of occurrence and narrowed down to various types of restaurants in New York.

This facilitates printing each neighborhood along with 5 most common cuisines in that area.

Top 10 types of restaurants in each neighborhood are displayed after sorting the 'Total' column.

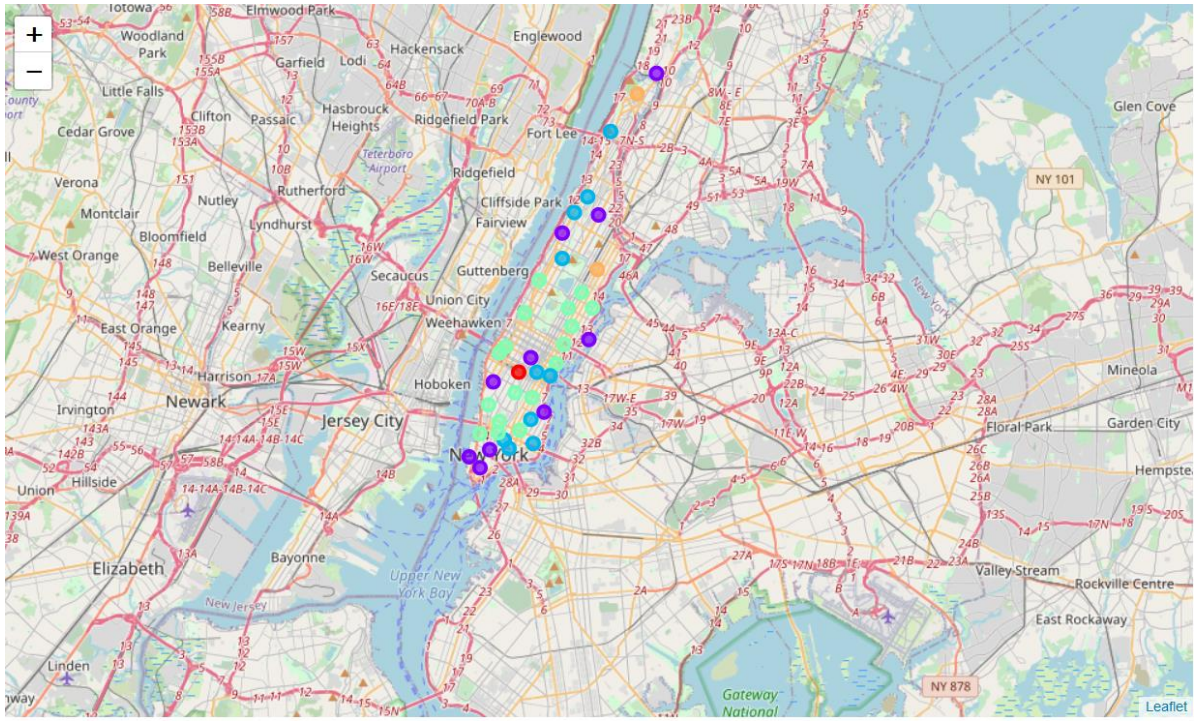
Using K-means clustering, the neighborhoods are sorted into clusters and similar neighborhoods are displayed on a map using folium.

Using the one-hot representation of the neighborhood venues dataframe, the 10 most common cuisines in New York are displayed using a matplotlib bar chart.

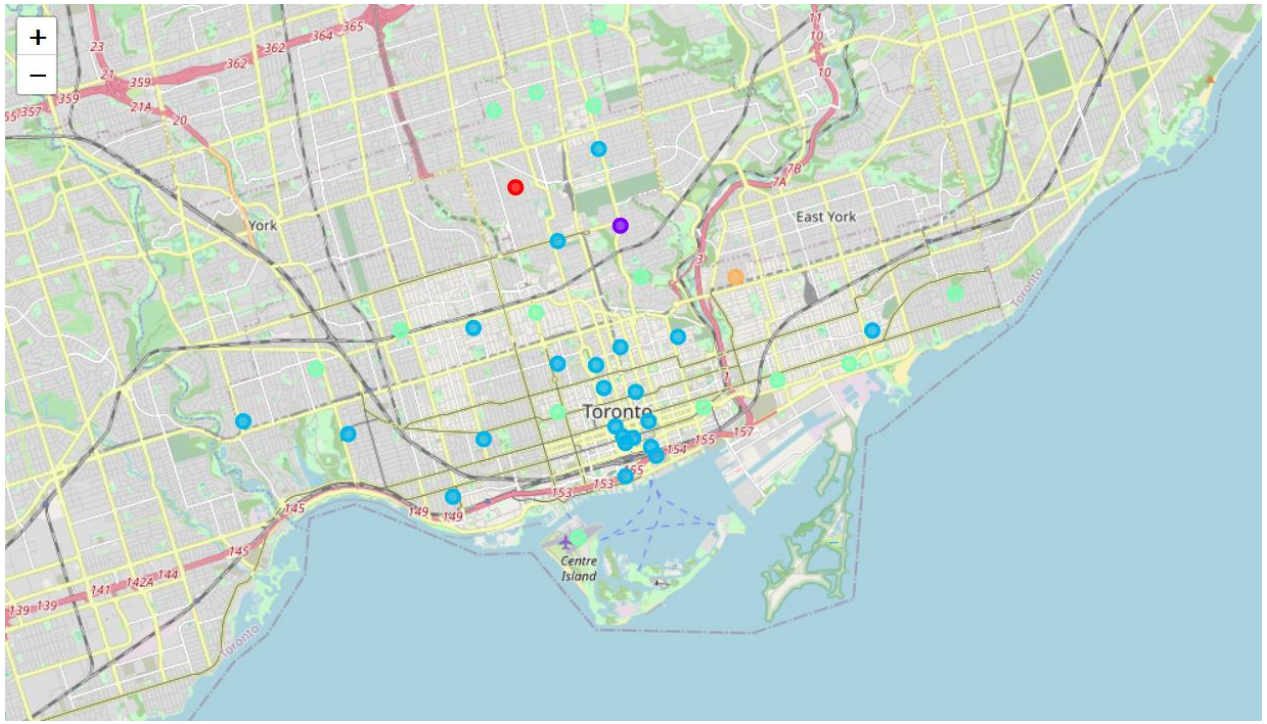
The same procedure is repeated for Toronto.

Results

New York

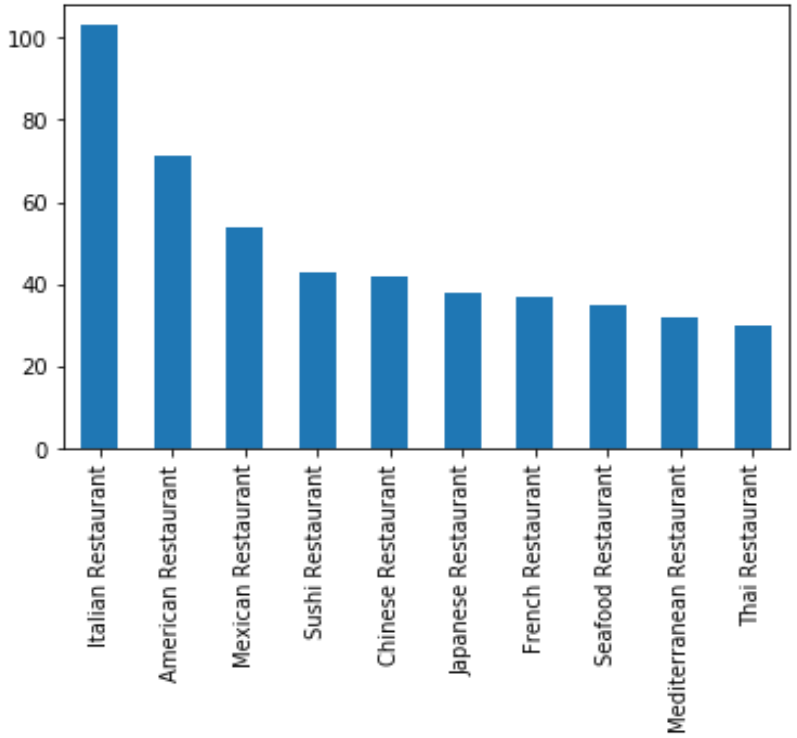


Toronto

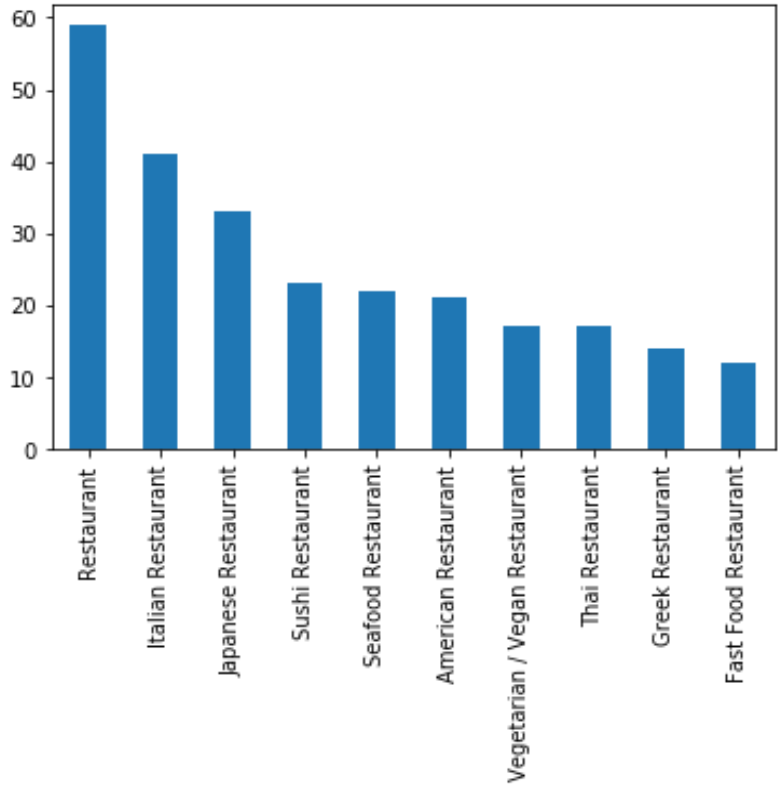


Results

New York



Toronto



Conclusion

With the available data, the above report has been generated but there is always some scope of improvement in the same. Most common cuisines have been identified and clusters of similar cuisines have been portrayed.