

Introduction: Business Problem

In this project we will **compare the Neighbourhoods of the two cities** and determine how **similar** or **dissimilar** they are.

We will retrieve the top 100 venues that are in the Neighbourhoods of a city within a radius of 500 meters. We will explore, analyse and cluster the Neighbourhoods of the two cities based on their latitude and longitude. The top ten venues around the Neighbourhood would be used to compare the cities.

New York and Toronto are very diverse and are the financial capitals of their respective countries. We will use data science to determine how similar or dissimilar the Neighbourhoods are of the two cities New York and Toronto and come up with findings.

Data

Based on definition of our problem, factors that will influence our decision are:

- Venues of any type in the Neighbourhood
- Top ten most common Venues

Following data sources will be needed to extract/generate the required information:

- The dataset containing the Borough, Neighbourhood with Latitude and Longitude coordinates for each city
- The dataset containing the Borough, Neighbourhood with Latitude and Longitude for Toronto is not readily available, but can be obtained by scrapping the wiki page https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M,
- The dataset containing the Borough, Neighbourhood with Latitude and Longitude coordinates for New York city is available for free from the website https://geo.nyu.edu/catalog/nyu_2451_34572
- The venues in every Neighbourhood will be obtained using **Foursquare API**

Methodology

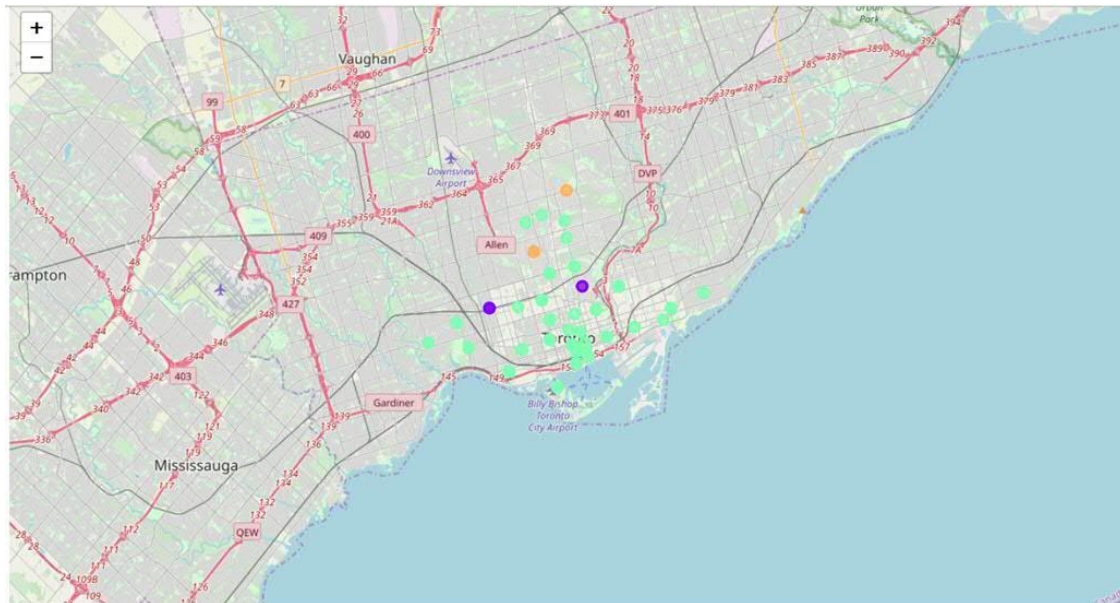
In this project we will retrieve the data of 100 venues near a Neighborhood using the Foursquare API. Then we will gather data for the Top ten most common venues for a Neighborhood.

In first step we have created the dataset for the two cities that has the Borough and the Neighborhoods. Then we have appended the data with the **Latitude and Longitude** values retrieved from the geo dataset.

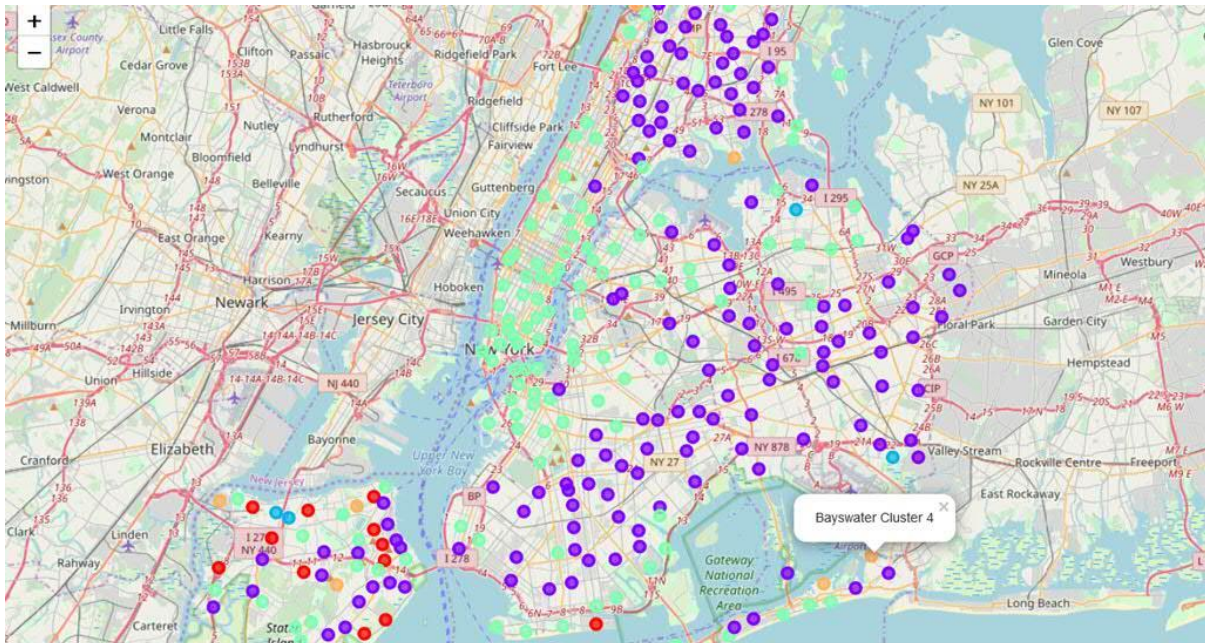
Second step in our analysis we have used the **Foursquare API** to retrieve the venues near the Neighborhoods, the top 10 most common venues for a Neighborhood. We have applied **one-hot encoding and normalized of data** of the venues.

In third and final step we have merged the data for the New York and Toronto cities applied **K-means clustering algorithm**. Then we have examined the map and the data generated with cluster labels.

Toronto Map with Clusters:



New York Map with Clusters:



Results and Discussion

Our analysis shows that Toronto and New York are similar in many ways.

Similarities: Both the cities are on waterfronts. Neighbourhoods have proximity to Parks, Playground. Very ethnically diverse, as there are restaurants catering to all types of cuisine. Both the cities are tourist destinations, as there are lot of tourist attractions.

Dissimilarities: New York neighbourhoods have proximity to Bus stops. New York is a large city far more venues compared to Toronto.

Conclusion

Purpose of this project was to compare the Neighbourhoods of the two cities and determine how similar or dissimilar they are. By using Foursquare API we were able to leverage the venues data to compare Neighbourhoods. The K-Means algorithm was very useful for Clustering similar data points.

The stakeholders can use this approach to compare Neighbourhoods effectively.