COMPARE THE NEIGHBOURHOODS OF TWO CITIES

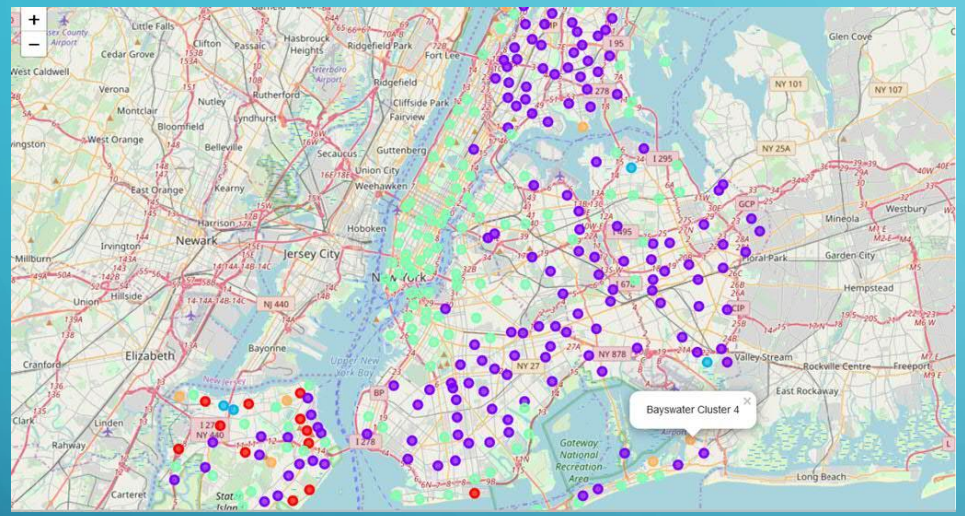
COMPARING NEIGHBOURHOODS VALUE

- Comparing cities is useful to understand the similarities and dissimilarities.
- The stakeholders would be interested in comparing the 2 cities for investment, Financial interests, setting up new Business etc
- The Tourists or visitors to the city would be interested in gathering insights on the city.
- New York and Toronto are very diverse and are the financial capitals of their respective countries.

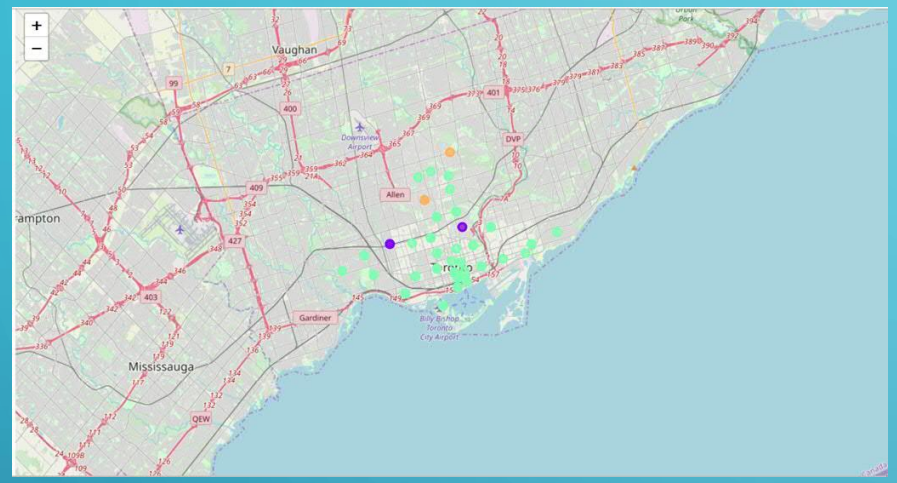
DATA ACQUISITION AND CLEANING

- 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**
- The handling of missing values and conversion to whole numbers as needed

NEW YORK CITY CLUSTERS WITH K-MEANS



TORONTO CLUSTERS WITH K-MEANS



FINDINGS

- 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.