IBM Applied Data Science Capstone Project

The Battle of the Neighbourhoods:

Toronto and New York

By: Ngai Cai Hua Kevin

Introduction/Business Problem

- Toronto and New York are big vibrant metropolitan cities with large population of diverse backgrounds

- Many tourists visit the cities; foreigners who also work and study in the cities.

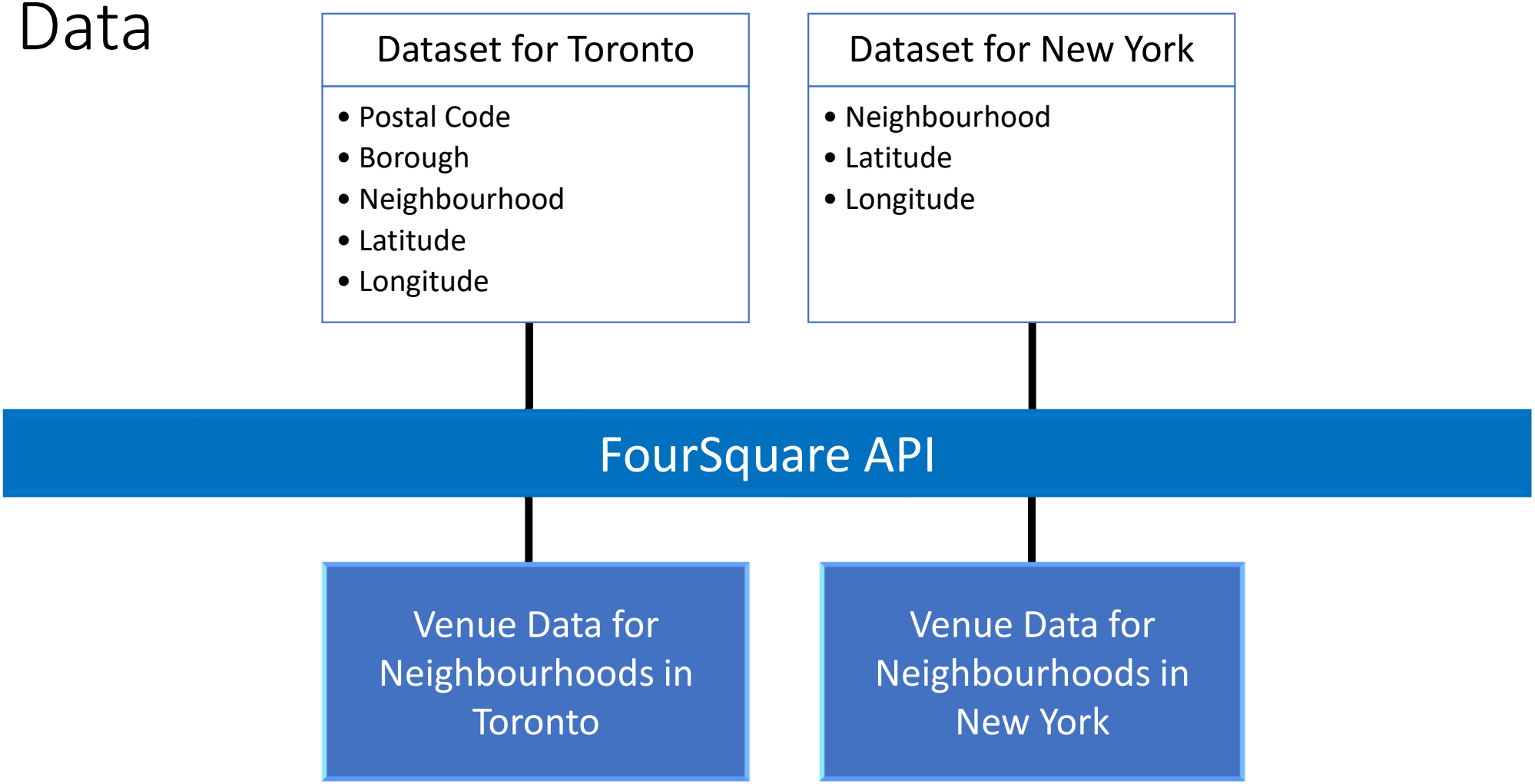
Business Problem

Identify and determine similarities and differences based on their venues/nearby attractions in their neighbourhoods.

Target Audience

- Toronto and New York Citizens
- Potential Tourists
- Potential Immigrants
- Potential Business Owners

Data


Dataset for Toronto
• Postal Code
• Borough
• Neighbourhood
• Latitude
• Longitude
Dataset for New York
• Neighbourhood
• Latitude
• Longitude
FourSquare API
Venue Data for Neighbourhoods in Toronto
Venue Data for Neighbourhoods in New York

Methodology

K-Means Clustering

Enables unlabelled data to be grouped based on similarity

Number of clusters = 5

Data Cleaning

Filter data and input into pandas dataframe

Data Transformation

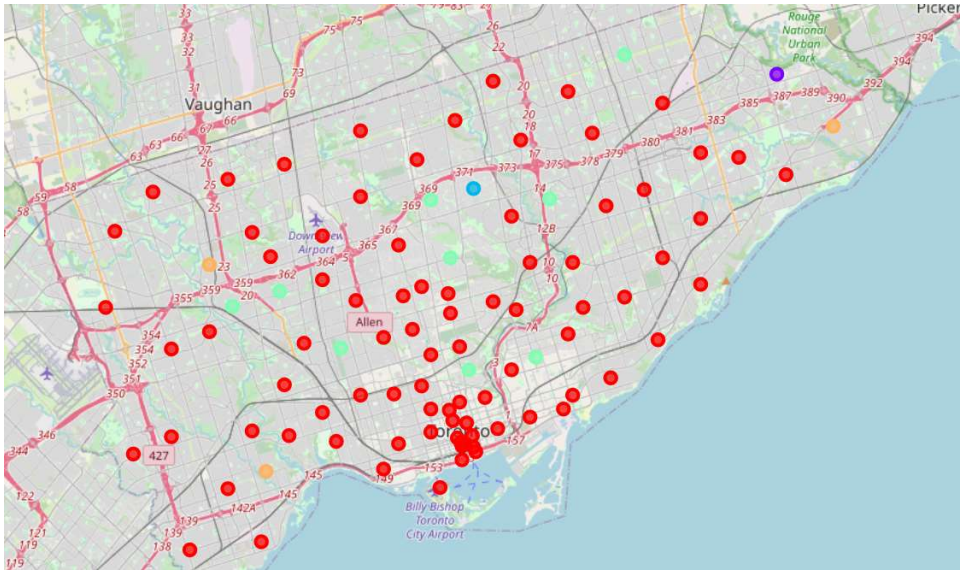
Establish connection with FourSquare API

Retrieve data from FourSquare API by providing the parameters (Toronto and New York location coordinates)

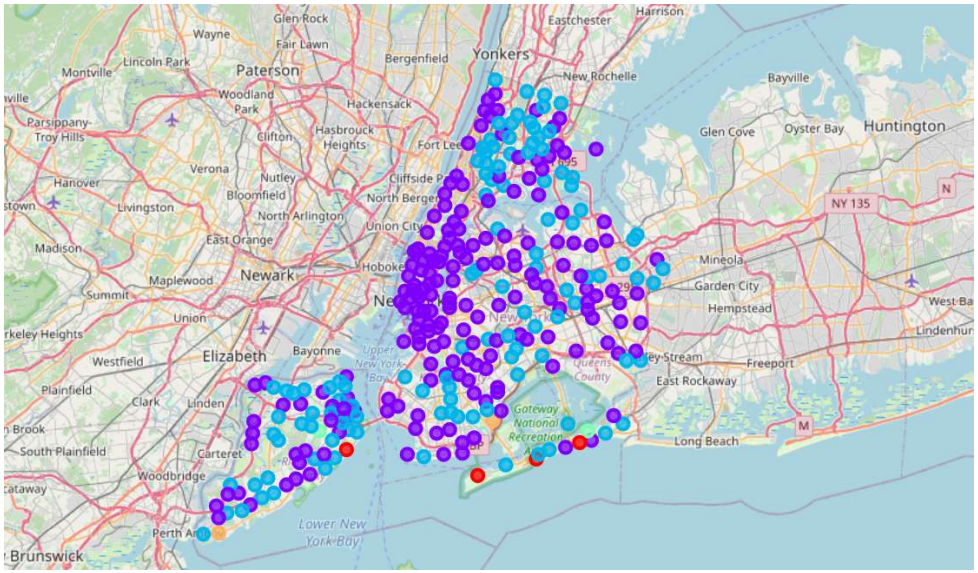
One Hot Encoding

Results/Observations

Map of Toronto Clusters



Map of New York Clusters



Total of 5 Clusters for each City

Results/Observations (Cont)

Similarities

- Food /Drink Outlets (Pizza Place is popular among both cities)
- Leisure and Recreation Venues (Parks, Pools, Beach)
- Fitness Venues (Yoga Studios and Gyms)

Differences

- More venues in New York than Toronto
  - New York is a larger city
- Location (Venue) points are more spread out in Toronto
  - Significantly less clustering in Toronto
  - Imply that Toronto is a less busier city than New York

Recommendations

Tourists:

- Both locations provide many good locations to visit
- Sight-seeing, visit different restaurants (wide variety)

Potential Business Owners:

- Setup business in either city, depending on the type of business
- New York offer more business opportunities
  - Large population and strong economy
- Very competitive in New York
  - Many competitors in the city

Conclusion

- Performed a comparison to determine the similarities and differences of both cities
- Methodology used
  - K-Means Clustering
  - Data Cleaning
  - Data Transformation
- Make recommendations for Target Audience