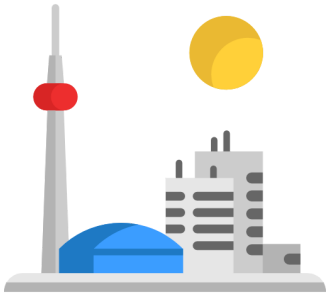
REPORT
Capstone Project - The Battle of Neighborhoods



New York **Toronto**

1. Introduction: Busines Problem

Introduction to the business problem and who would be interested in this project

I will compare the Neighbourhoods of New York and Toronto for the availability of sports classes (gyms)

The problem: to determine similarity and (or) dissimilarity.

I will:

- retrieve the top 100 venues that are in the Neighbourhoods of a city within a radius of 1000 meter;
- explore, analyse and cluster the Neighbourhoods of the two cities;
- use and compare the top ten venues around the Neighbourhood

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 Neighborhoods are of the two cities New York and Toronto and come up with findings.

Estate agencies would be interested in this project.

2. Data

Data with a description of the data that will be used to solve the problem and the source of the data

For Toronto:

https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, - dataset containing the Borough, Neighbourhood with Latitude and Longitude for Toronto

For New York:

https://geo.nyu.edu/catalog/nyu_2451_34572 - dataset containing the Borough, Neighbourhood with Latitude and Longitude coordinates for New York city

For both - Foursquare API

Based on the definition of our problem, the data that will be used to solve the problem are Venues of any type in the Neighbourhood

3. Methodology

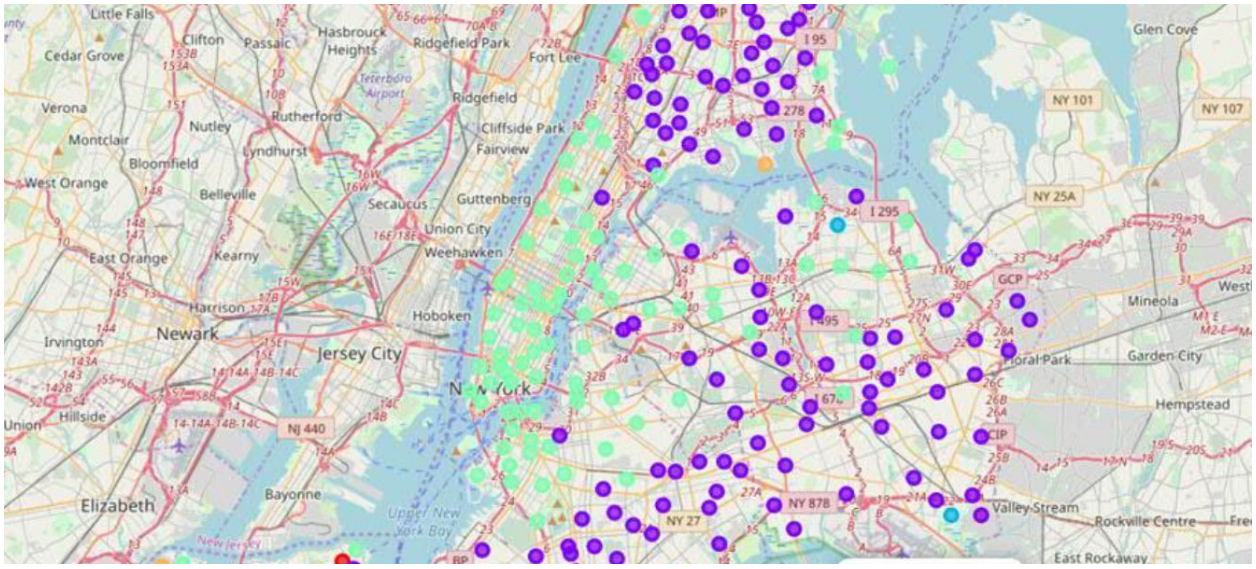
Methodology section which represents the main component of the report with exploratory data analysis and inferential statistical testing that you performed,
if any, and what machine learnings were used and why

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.

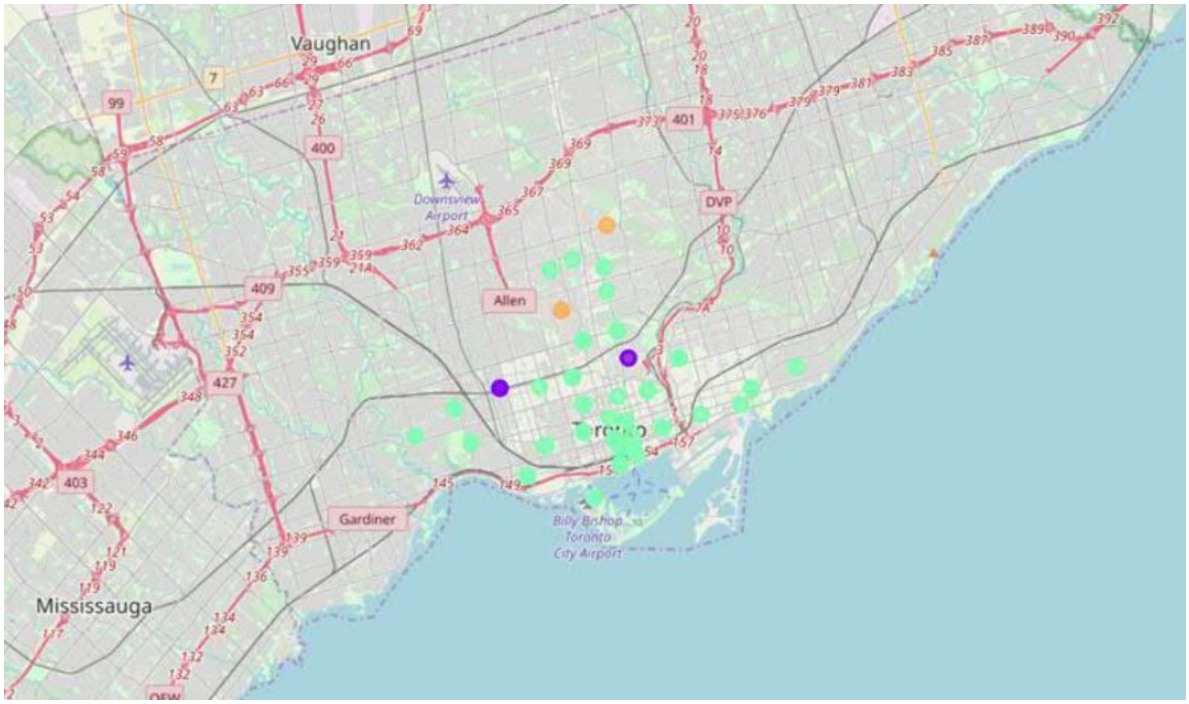
1. Created the dataset for the two cities that has the Borough and the Neighborhoods.
2. Appended the data with the **Latitude and Longitude** values retrieved from the geo dataset.
3. Used the **Foursquare API** to retrieve the venues near the Neighborhoods, the top 10 most common venues for a Neighborhood.
4. Applied **one-hot encoding and normalized of data** of the venues.
5. Merged the data for the New York and Toronto cities
6. Applied **K-means clustering algorithm**.
7. Examined the map and the data generated with cluster labels.

4. Results

New York



Toronto



5. Discussion

Discussion section with discussion any observations you noted and any recommendations you can make based on the results.

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

Similarities: Both the cities are on waterfronts. Neighbourhoods have proximity to Restaurants with all types of cuisine, Bars, Parks, Culture Centers. Very ethnically diverse. Both the cities propose a lot of opportunities for training and sport activities

Dissimilarities: New York neighbourhoods have more gyms for sports activities compared to Toronto.

6. 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.