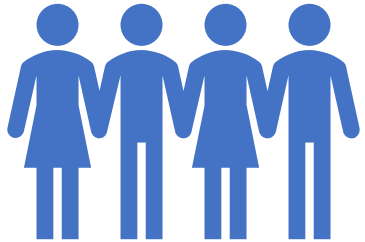
Predicting an optimal location of the Coffee Shop

"IBM Data Science“ Capstone Project

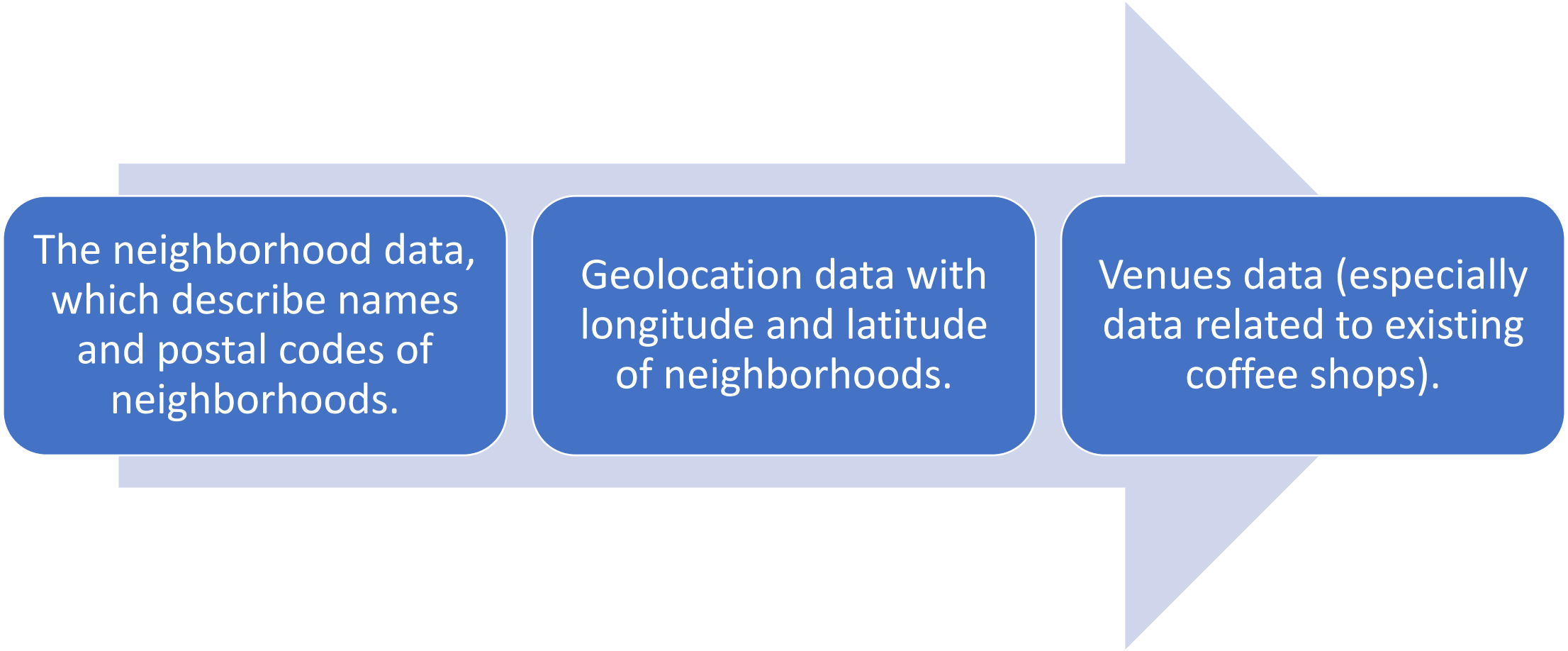
The problem

The objective of this project is to determine an optimal location for opening a new coffee shop in Toronto. The location significantly affects the profit of any company. Thus we could select the "right" location and get much more money. Using clustering approach allows us to solve this business problem.



The target audience of this project is any person who cares about establishing new business connected with coffee shops.

Data


The neighborhood data, which describe names and postal codes of neighborhoods.
Geolocation data with longitude and latitude of neighborhoods.
Venues data (especially data related to existing coffee shops).

Methodology

WEB-SCRAPPING DATA

ADD GEOSPATIAL DATA

USE FOURSQUARE API TO OBTAIN VENUES DATA

GROUP DATA BY NEIGHBORHOOD AND CALCULATE VENUE FREQUENCIES

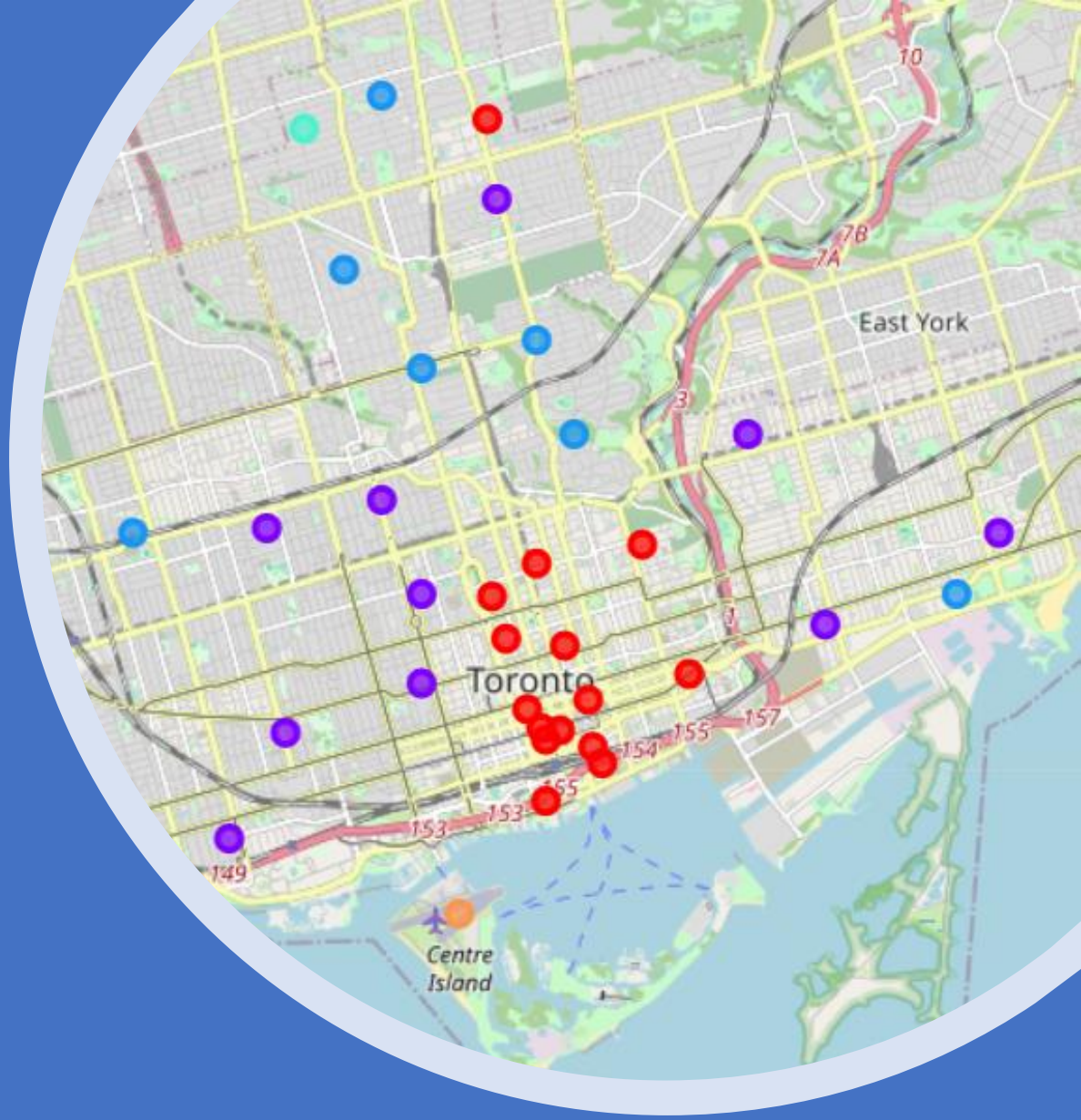
SELECT FEATURES

DEFINE CLUSTERS NUMBER

CLUSTERING

Results

- We can see that clusters 0, 4, 5 have a very high number of coffee shops in most neighborhoods, otherwise, clusters 1, 2, 3 has a low number of coffee shops. It provides a great opportunity for establishing a new business.
- Based on the map, most coffee shops are concentrated in the center of the city, meanwhile, the suburban area still has very few coffee shops.



Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 6 clusters based on their similarities, and lastly providing recommendations to the relevant to open a new coffee shop.