IBM Data Science Professional Certificate

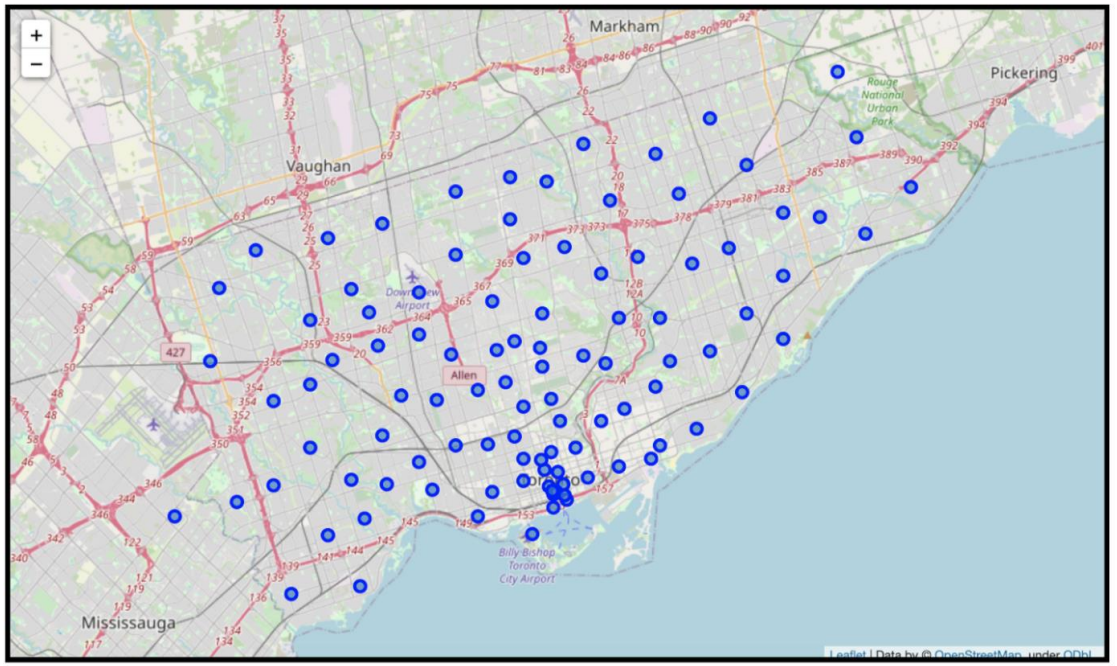
Battle of Neighborhood

By-ANCHALA BALARAJ

INTRODUCTION

- The purpose of this project is to help people in exploring better facilities around their neighborhood. It will help people making smart and efficient decisions on selecting great neighborhoods out of numbers of other neighborhoods in Scarborough, Toronto.

- We will be using data visualization techniques, Foursquare API to retrieve location data for the state of Toronto and use this to perform data analysis.



Battle of the Neighborhoods

BACKGROUND

- The goal is to inform potential business owners who are looking to expand in new areas. After pulling the data for these respective cities - we will have an ability to provide a recommendation. We will be using several data visualization techniques, in particular, we will be making use of Foursquare API to retrieve location data for the state of Toronto in Canada and use this data to perform data analysis.

INTEREST

- This analysis will be useful for those who want to start a new business in the state of Toronto .It also aims on helping those who want to travel to Toronto and want to visit global Cuisine restaurants in Toronto. Clusters tell us what neighborhoods are fairly similar to each other so the person can skip travelling to many of the same neighborhoods.

PROBLEM

- In growing regions - it is a challenge to know where it is appropriate to start a business, let alone what business to start. In this report we will be exploring the need for additional restaurants as in Toronto.

- Prior launching any restaurant, it's important to know if the business as a good opportunity. In order to do so, this report will try to gather data about other restaurant localization, competitors and best localization. also keeping in mind the nearby tourist spots they can visit which makes it easier for business person to choose the right neighborhood for their restaurant.

DATA ACQUISTION

- The foursquare location platform will be used as the sole data source all the stated required information can be obtained through the API.

- Cleaning Data:
  We can combine rows that have same postal code and make them into a single row and neighborhoods that are associated with that postal code would be put into the same row separated by commas. We sort data buy the postal codes.

- Feature Selection:
  We focus on the neighborhood in Toronto. So drop all rows that have Borough outside of Toronto. And our dataframe is ready to use.

	index	Postal Code	Borough	Neighborhood
0	6	M1B	Scarborough	Malvern, Rouge
1	12	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	18	M1E	Scarborough	Guildwood, Morningside, West Hill
3	22	M1G	Scarborough	Woburn
4	26	M1H	Scarborough	Cedarbrae

Sorted the data by Postal Code

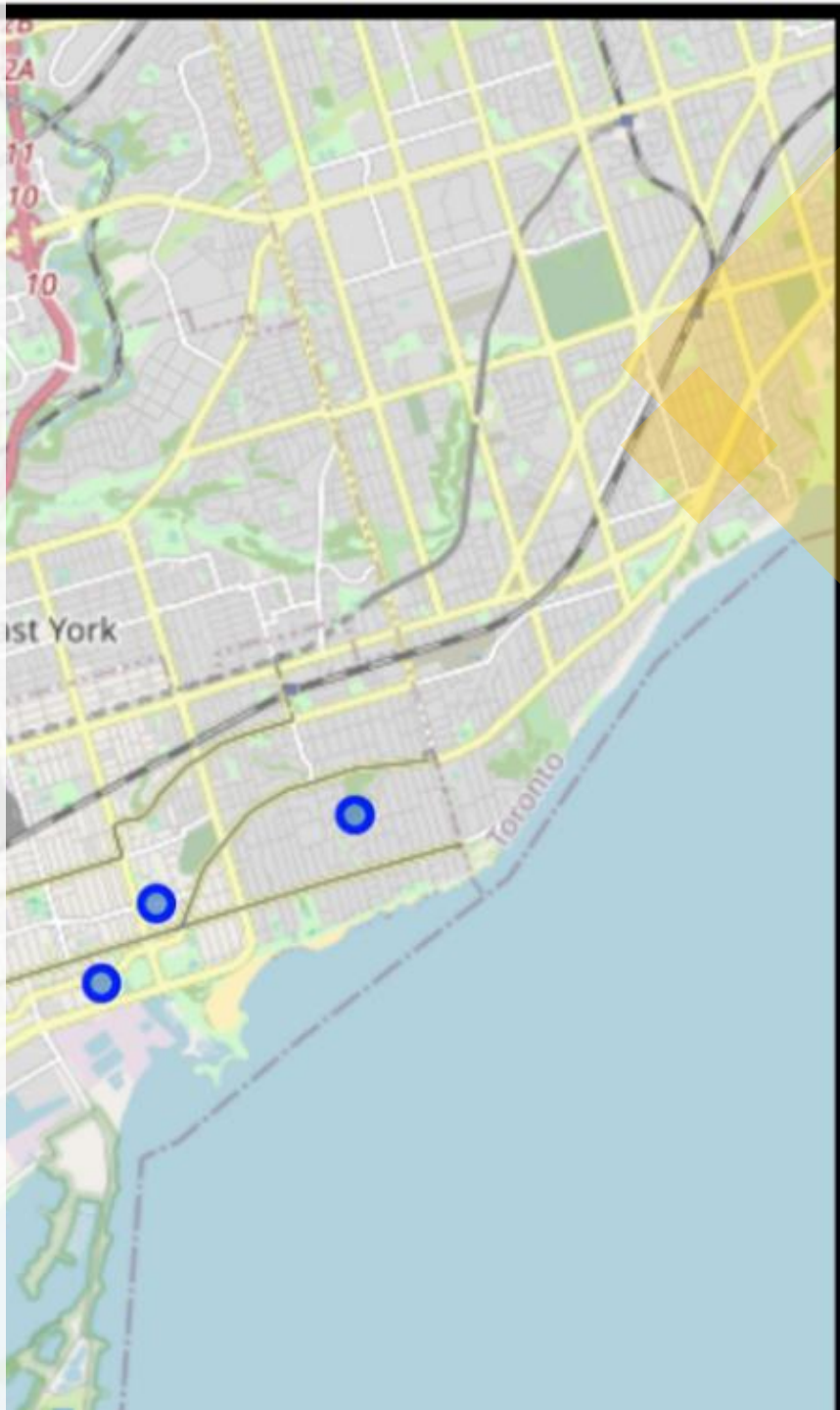
	Postal Code	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Regent Park, Harbourfront
5	M6A	North York	Lawrence Manor, Lawrence Heights
6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

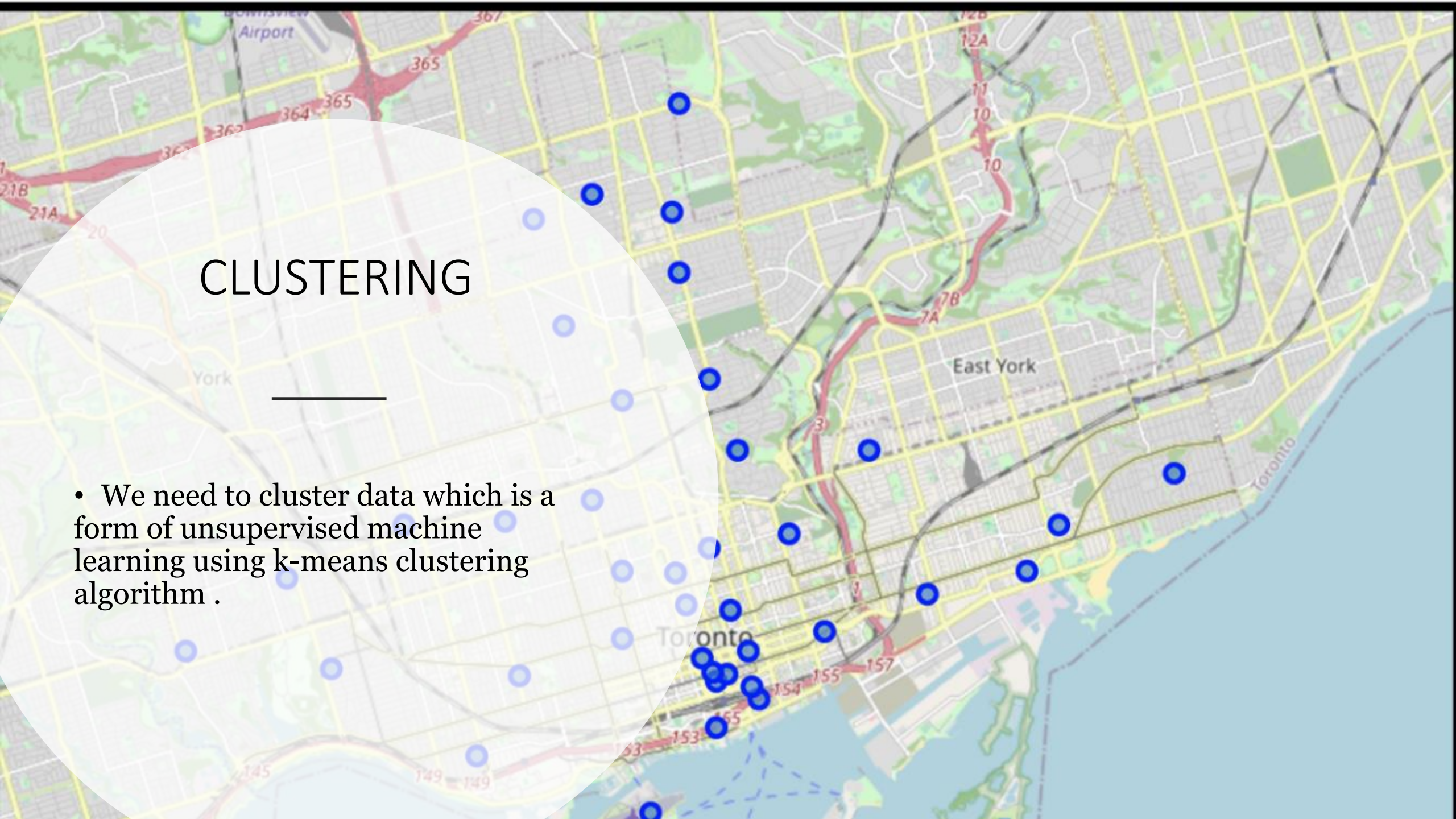
Data from Wikipedia

METHODLOGY

Presence of Vast data & Classification :

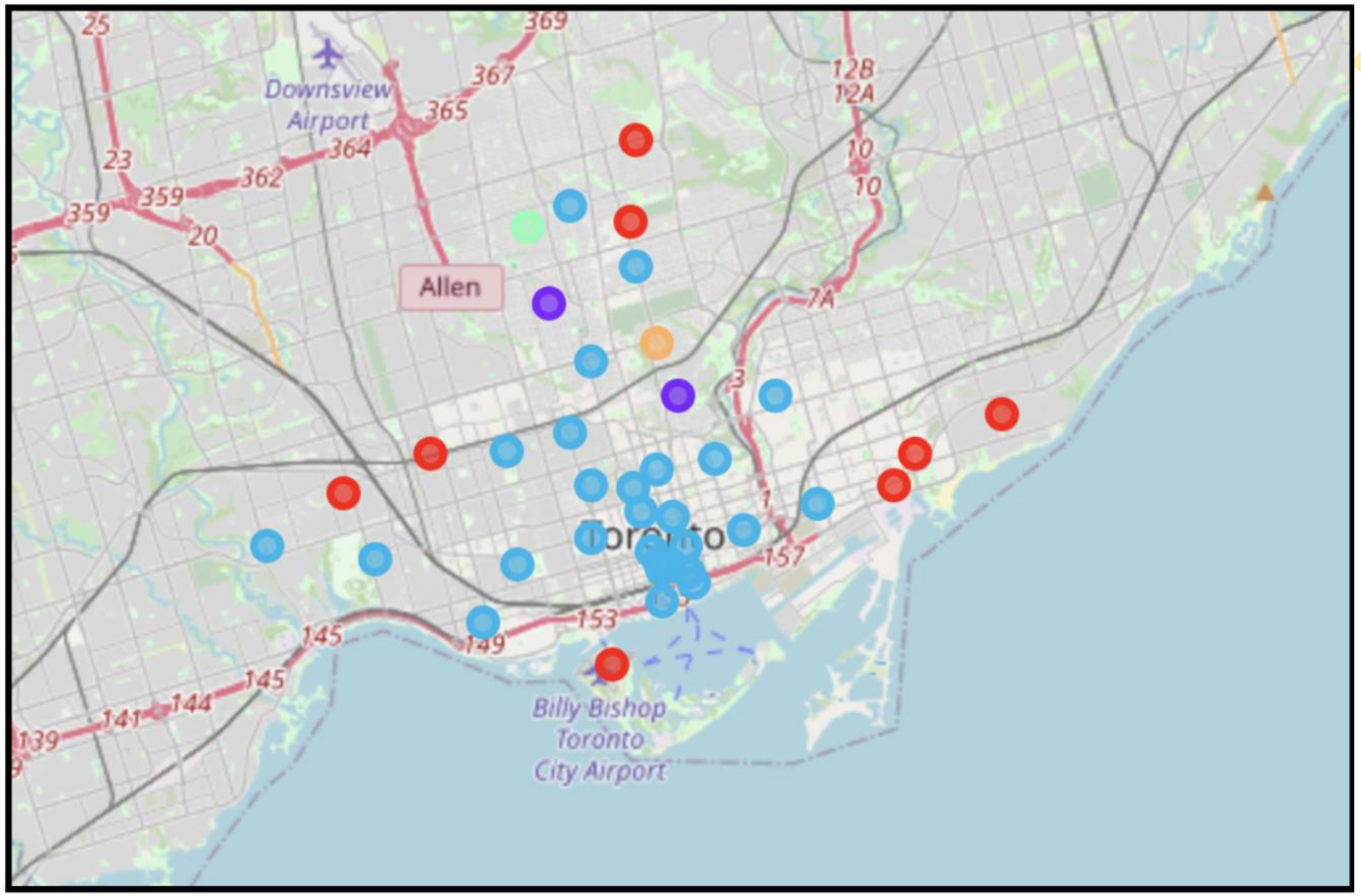
- The dataset will allow us to group the neighborhoods together according to the similarity in the type of venues in each neighborhood.
- We make use of the mean of the frequency of the occurrence of each category and combine the neighborhoods with similar venues.





CLUSTERING

- We need to cluster data which is a form of unsupervised machine learning using k-means clustering algorithm .



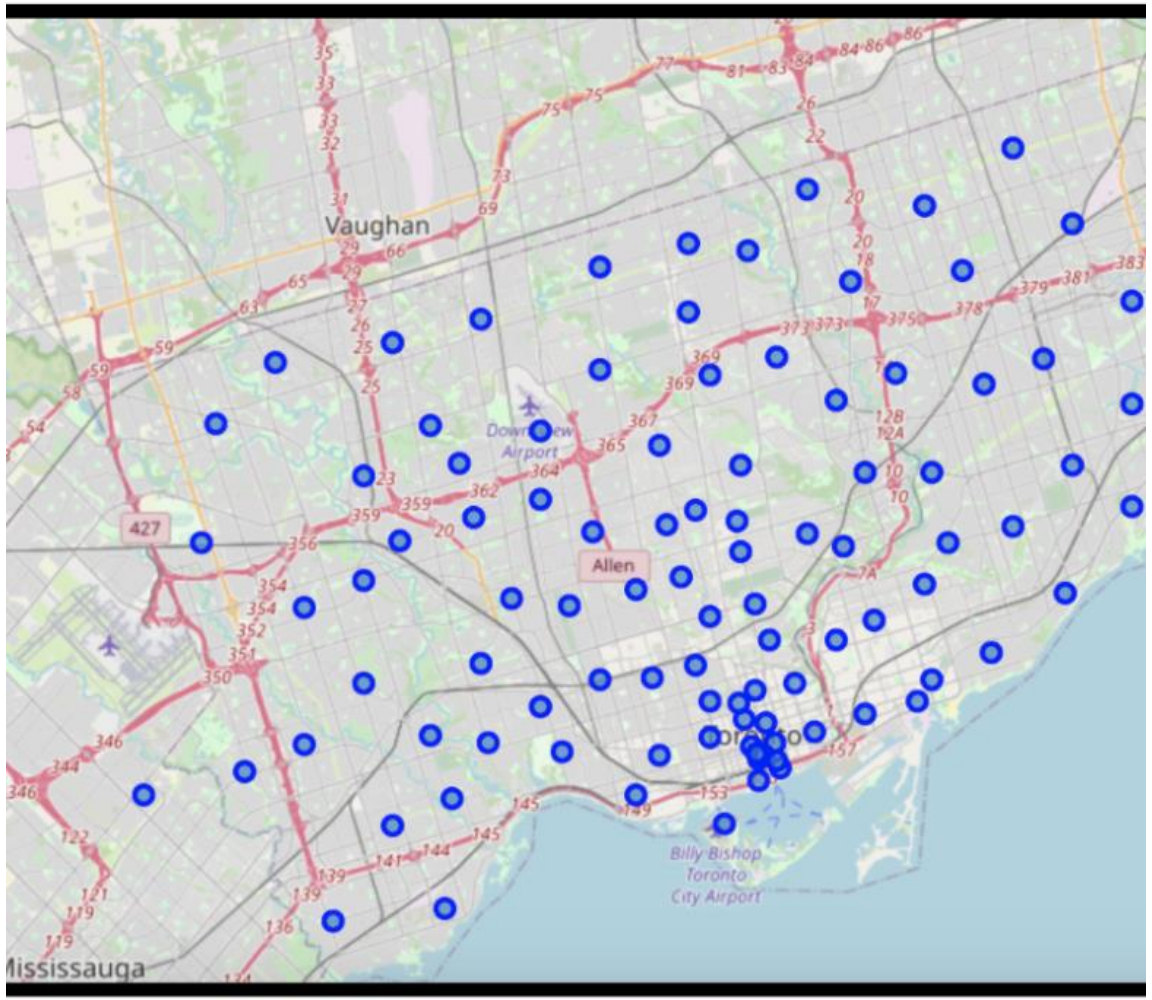
Clustered Neighborhoods in Toronto

RESULT

- The neighborhoods in the Red cluster are mainly on the outskirts of Toronto.
- There are 2 neighborhoods that form their own cluster and one cluster (indicated in purple) consists of only 2 neighborhoods.
- Since the neighborhoods were clustered based on the similarity in the categories of popular venues, it can be observed that most neighborhoods have the same category of venues.

Conclusions

- The clusters allow interested people to understand how similar neighborhoods are in Toronto.
- Using the data about the 10 most popular venues in each neighborhood group allows people to choose the right neighborhood for starting a new business or opening a new branch for their already existing business.
- The similarity in neighborhoods allows tourists to decide which places they should add on the to-visit list without making redundant choices. It also tells them what are the most popular places n each neighborhood that they must visit.



Neighbourhoods in Canada


THANK YOU!