BATTLE OF THE NEIGHBORHOODS

INTRODUCTION

Background

- Following this trend of increasing number of coffee house establishments, a person who is also interested in establishing needs to conduct a thorough research.
- Research implies:
  - Understanding of the number of coffee houses that are established in various neighbourhoods
  - How many of these coffee houses are successful
  - What factors contribute to the success of that particular establishment

Problem

- The dataset of Toronto has a list of all the boroughs and neighbourhood.
- This project allows us to explore the neighbourhoods of all the boroughs and divide these into clusters using k-means clustering.

Interest

This project is targeted to those people who are interested in setting up a coffee house.

DATA

Data Acquisition

- The data containing the postal codes of Canada was obtained from https://en.wikipedia.org/wiki/List_of_postal_codes_Canada. This dataframe consisted of three columns - Postal Code, Borough and Neighbourhood.
- The next set of data containing the latitudes and longitudes of each neighbourhood was collected from http://cocl.us/Geospatial.
- The dataset that was finally created, consists of four columns – Neighborhood, Borough, Latitude, Longitude.

Data Cleaning

- From the first data set, only the cells that have an assigned borough are selected.

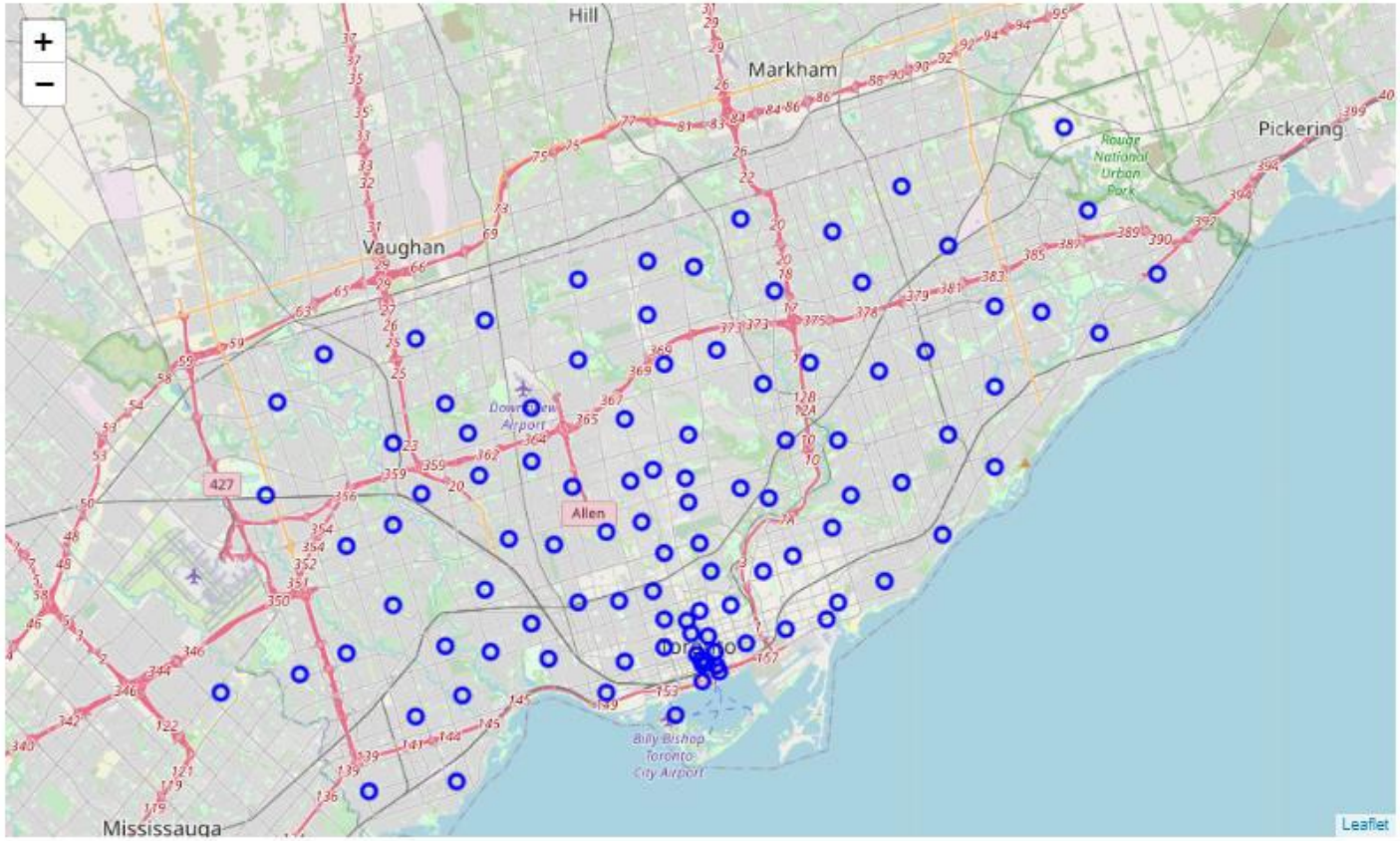
	PostalCode	Borough	Neighborhood
0	M1B	Scarborough	Malvern, Rouge
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek
2	M1E	Scarborough	Guildwood, Morningside, West Hill
3	M1G	Scarborough	Woburn
4	M1H	Scarborough	Cedarbrae

- The second dataset containing the co-ordinates of all the boroughs is merged with the above dataset at the postal code.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Malvern, Rouge	43.806686	-79.194353
1	M1C	Scarborough	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
2	M1E	Scarborough	Guildwood, Morningside, West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

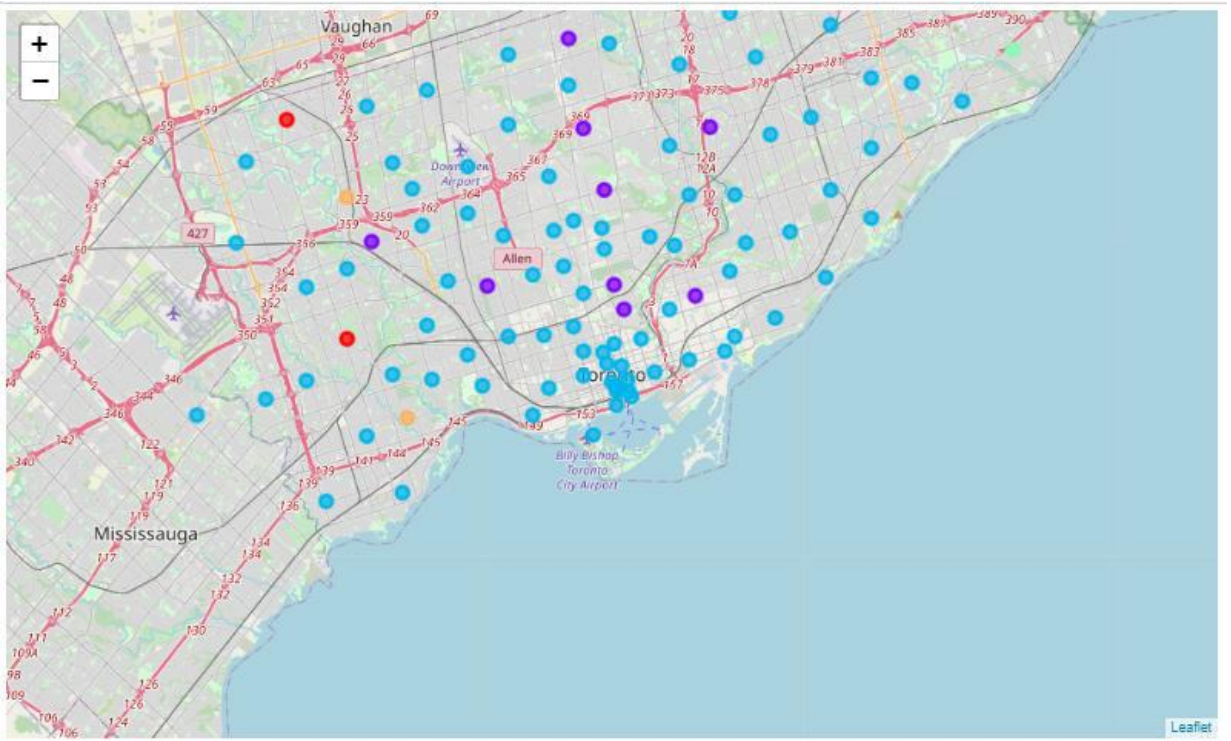
METHODOLGY

There are about 102 boroughs in Toronto, which are clustered using k-means clustering.



RESULTS

With the help of k-means clustering, these boroughs are segregated into five clusters. These clusters were further examined to determine the 10 most visited venues with these clusters.



DISCUSSION

- The aim of this project was to help someone who is interested in setting up their own café or coffee house.
- The third cluster is an extremely active clusters with a lot of establishments that are popular among the masses.
- The remaining four clusters have lesser establishments, especially few cafés or coffee houses. Establishing one here would be profitable as the people living here would not have to travel far to access a popular one.

CONCLUSION

This project helps anyone who is interested in establishing a coffee house in Toronto. It is always a good option to use technology to stay ahead in the game and not act on the basis of trial and error, as setting up any establishment would require large number of investments.

In this project, I have taken coffee house as an example, but it can be used for other establishments also. One simply has to analyse which are the commonly visited venues in each cluster and make a decision accordingly.