Introduction:-

Toronto is an international centre of business, finance, arts, and culture, and is recognized as one of the most multicultural and cosmopolitan cities in the world.Starting a restaurant in toronto can be a great business opportunity, but you need to open it in a very good location.

Business problem:-

Which neighbourhood in Downtown Toronto borough is best to open a restaurant?

Analytic approach:-

As the problem is a measure of which neighbourhood is best to open a restaurant we will cluster the venues for travellers on each neighbourhood using K-means cluster

Data

Data requirements :- Based on definition of our problem, A factors that will influence our decission is number of venues for travellers

Data source :- The data used in this analysis is gotten from https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M it has the postal codes of each neighbourhood and the borough it falls under, I will also use the csv file provided in week 3 of this capstone it has the latitude and longitude of the neighbourhoods

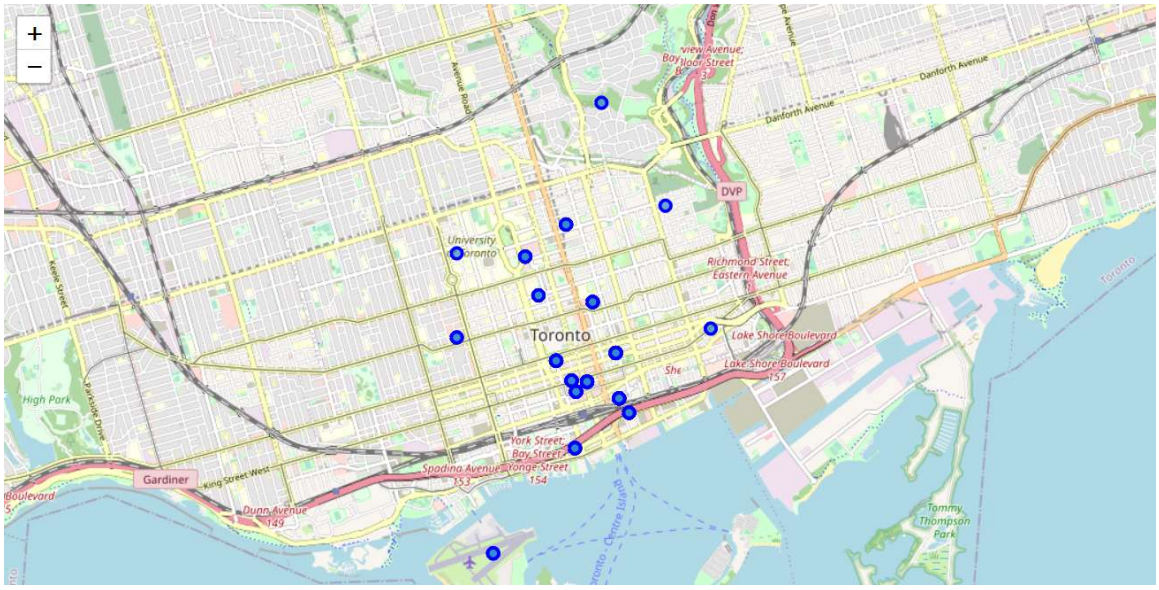
Data Understanding or preparation:- Read the datas into Pandas DataFrame and merge them into a single dataframe get the venues for travellers in each neighbourhood using foursquare API.

Methodology

In this project we will direct our efforts on Downtown Toronto Borough. We will limit our analysis to area ~5km around Borough center.

In first step we have collected the required data: location and neighorhoods within 5km around Downtown Toronto Borough . We have also identified venues for travellers (according to Foursquare categorization).

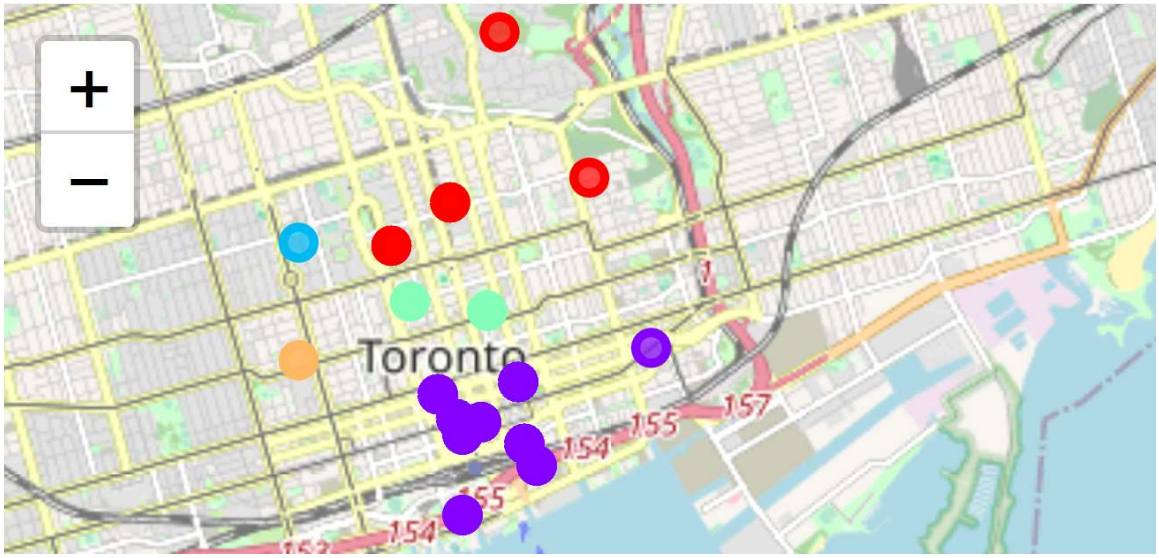
Second step in our analysis will be calculation and exploration of 'number of venues for travellers in each neighorhoods



In third and final step we will focus on most promising areas and within those create clusters of locations that meet some basic requirements established in discussion with stakeholders: we will take into consideration neighorhoods with more number of venues for travellers in each neighorhoods, . We will present map of all such locations but also create clusters (using k-means clustering) of those locations to identify general zones / neighborhoods / addresses which should be a starting point for final 'street level' exploration and search for optimal venue location by stakeholders.

Model / Evaluation

The model is built using K-means cluster



Conclusion

From the analysis it can be seen that it is best to open a restaurant in the Commerce Court, Victoria Hotel neighghborhood