IBM CAPSTONE PROJECT – The Battle of Neighborhoods:
Cluster Analysis of Toronto Coffee shops

Introduction: Business Problem

In this project I would like to find an optimal location for a Coffee shop in Toronto.

The idea is to find a place in the city center, near some clothes shop or landmark in order to have more tourists. It's important that this shop is far from other similar shops.

I'll use web scraping in order to get information about possible neighborhoods and Foresquare in order to get information about shops.

The target audience of this report is any one that is interested in opening a Coffee shop but have no idea in which arrondissement.

Data

The most important criteria is the number of existing Coffee shops in the neighborhood. I'll divide toronto according to its arrondissements to define neighborhoods, centered around city center; for each arrondissement I will find the coffee shops in order to calculate the best location for the new shop.

Each arrondissement will be approximated using Google Maps API geocoding, every activity its type and location using Foursquare API. The center of toronto will be approximated using Google Maps API too.

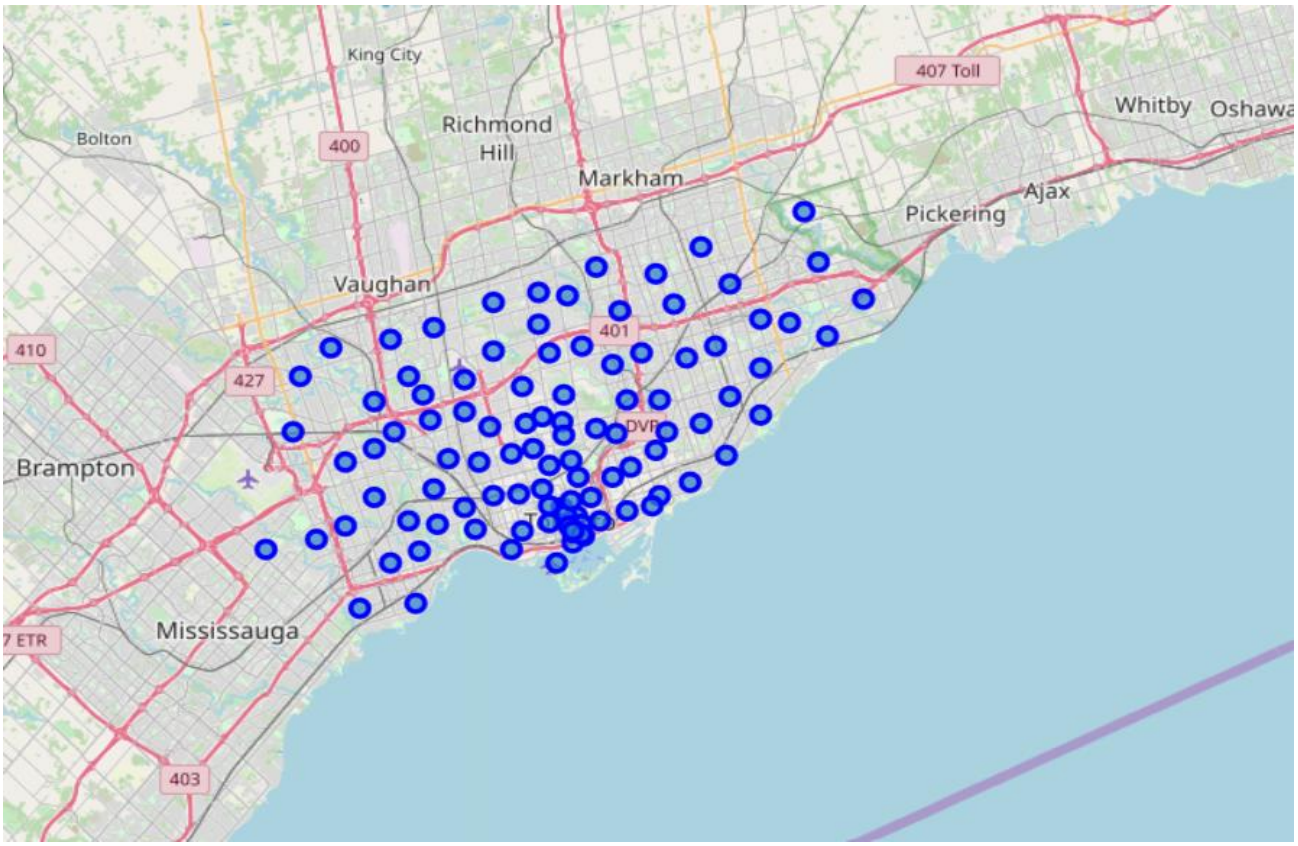
Data Exploration

After extracting and reading the data, we will translate the above data into a Pandas data frame for processing which would look like this. These are the data elements that are needed when we call Foursquare web service call in order to get the venues available in that arrondissement.

	name	categories	lat	lng
0	Roselle Desserts	Bakery	43.653447	-79.362017
1	Tandem Coffee	Coffee Shop	43.653559	-79.361809
2	Cooper Koo Family YMCA	Distribution Center	43.653249	-79.358008
3	Body Blitz Spa East	Spa	43.654735	-79.359874
4	Morning Glory Cafe	Breakfast Spot	43.653947	-79.361149

Foursquare

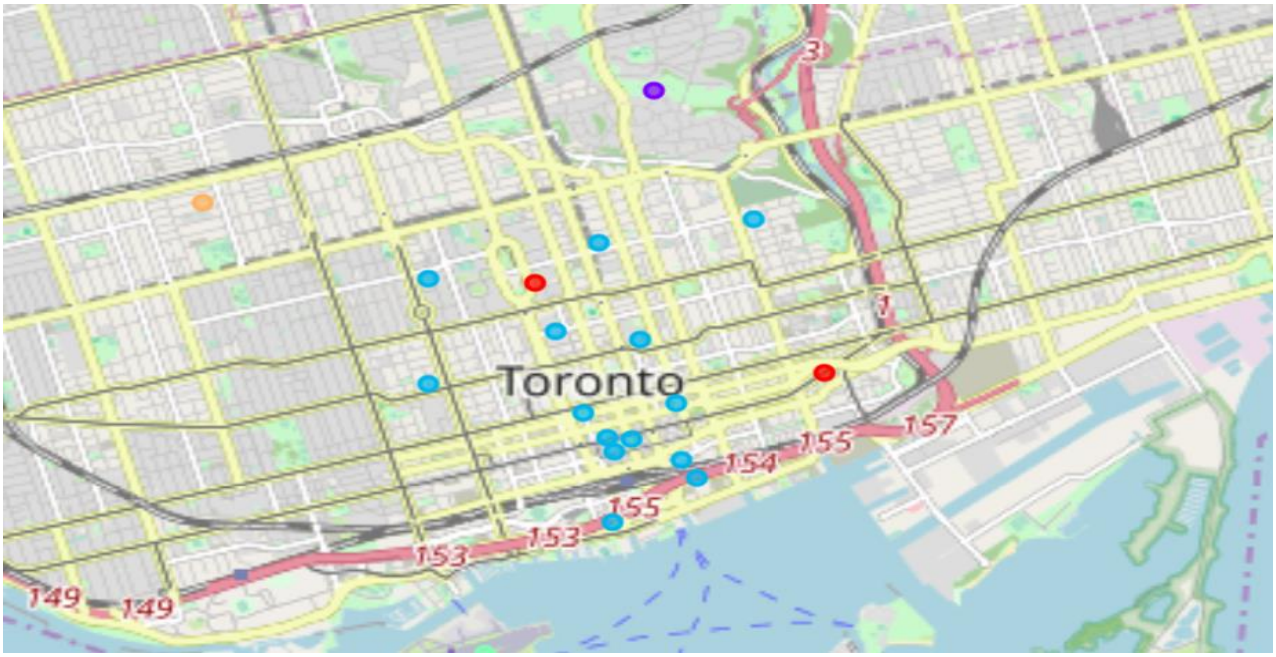
I use Foursquare API to get info on each activity in each neighborhood.
I'm interested in activity that sell coffee...
In the map all this kind of activity are in red



I prefer to concentrate only in coffee shops so I filter the data and I try to identify cluster for each arrondissement.

Arrondissements 8,10,11 have not any coffee shop so are perfect place.

However I visualize cluster related to the other arrondissement



As we can see most of them are concentrated in the city center buti f the client dosen't matter about the distance from the center in the rest of the city there a lot of good place near other activity as we have seen above.