**CAPSTONE PROJECT**

The project, I intend to do is based on a hypothetical scenario where a coffee shop chain company want to start a coffee shop in the Central Toronto location. The data can also help other small coffee shop retailers. The main motivation is to aid any Coffee Shop chain or CCD chains, or any other retailer help set up shop in the Central Toronto region. The analysis can aid CCD chains or retailers in picking locations for the setup of their coffee shops. The main aim is to help a Coffee Shop retail chain help find locations for setting up shops in the Central Toronto region.

**OBJECTIVE**

The objective of the project is to analyse the locations in Central Toronto and recommend good locations to open Coffee Shop for a CCD and Coffee Shop chain retailer. The project aims to use location data and data manipulation and other statistical techniques such as clustering to recommend best locations for a coffee shop. The main problem is which are set of best locations to open a CCD in Central Toronto

**TARGET AUDIENCE**

The target audiences are CCD coffee shop chains and other small retailers who can leverage the data for opening coffee shops. The retail chains like Starbucks use location data to open their stores in locations without affecting the business of their other stores and for maximizing the profit in the new store or shop.

**DATA**

The data required are as follows.

Toronto neighbourhood data and postal code data

Toronto neighbourhood location data

Data for top coffee house location. The data related to coffee shops in the neighbourhood

The Toronto neighbourhood data will be scrapped from this site

<https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

The neighbourhood location can be taken from the csv file or can be extracted using Geopy or Google API key.

The FourSquare API will be used to fetch the criterion-based data of the neighbourhoods in the Central Toronto region. The FourSquare API based data will be taken where the top venues of about 100 in number will be grouped together. The data will be accessed through API calls to FourSquare. The criterion-based data will be filtered from that data for the coffee shops and the top venues can be categorized based on their frequency of the location. The data from the venues can be filtered for the coffee shops and then clustering can be done on them to identify the location cluster where the coffee shops can be opened.