

# This is my Notebook for IBM Data Science Capstone Project (Classification of Toronto Neighbourhood)

By:- Komal Thakkar

## Introduction & Business Purpose:

As a part of the IBM Data Science Professional Certification Program, I worked on these datasets to get an real experience of data science. Main objectives of this project were to define a business problem, look for data in the web and, use Foursquare location data to compare different neighborhoods of Toronto to figure out the number of coffee shops in neighbourhood of Toronto & thereby decide which neighborhood is suits best for competitive market.

The purpose for this notebook is to explore the neighborhood of Toronto and generate statistics and analytics on top of it. The introductory part has data cleansing & preprocessing, followed by Analysis of most widely spread coffee shop in neighbourhood of Toronto.

## Problem Description :

Assuming that a person wants to setup a new Coffee Shop in neighbourhood of Toronto. And the person wants to explore all competitors around the city of Toronto and who has how many outlets in the neighbourhood, to grasp the hold of particular vendor. So to overcome doubts in his mind whether it is a good idea to open a shop or not, Data Analysis and report is required. In order to make the business profitable, significant analysis before investing is must.

## Area of Benefit:

There can be significant beneficiaries through analysis of this project:

- Business Person who wants to open an outlet of coffee in the neighbourhood.
- The end-users i.e. ardent coffee fans.
- Enterprise of Data Analyst / Data Scientist who analyses the neighbourhood using statistical Data.

## Data Acquisition

Below are the list of sources from which I have collected the data for different purpose:

1. Postal Codes for Canada :- · I have fetched the postal code of the neighborhoods in Canada by scraping Wikipedia. · Link — [https://en.wikipedia.org/w/index.php?title=List\\_of\\_postal\\_codes\\_of\\_Canada:\\_M&oldid=1011037969](https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=1011037969)
1. Geographical Co-ordinates :- · I used a CVS file which consist latitude and longitude of the neighborhoods in Canada. · I chose the CVS file instead using geocoder. · Link for CVS — [http://cocl.us/Geospatial\\_data](http://cocl.us/Geospatial_data)
1. Foursqaure for Fetching Details of the venue : · I used Foursquare API for fetching the details and location of the venues. · And finally visualize using Folium.

From Foursquare API (<https://developer.foursquare.com/docs>),

I retrieved the following for each venue: a) Name: The name of the venue. b) Category: The category type as defined by the API. c) Latitude: The latitude value of the venue. d) Longitude: The longitude value of the venue. e) Likes: Likes of the venue, that the user liked the restaurant. f) Rating: Rating of the venue. g) Tips: Tips given by the users.