# Coursera Applied Data Science Capstone

## Opening a new Business in city of Toronto

**Author: Arjun P**

## Data section:

The objective of this project is to propose the most suitable location for starting a new business in the city of Toronto. Since Toronto is a large metropolis with diverse business venues and ventures, we require acute and exhaustive data corresponding to Toronto neighbourhoods and venues.

To solve this problem, the following Data are needed:

1. List of Boroughs and Neighbourhoods in Toronto, Canada
    - Data source: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
    - Description: This is a list of postal codes in Canada where the first letter is M. Postal codes beginning with M are located within the city of Toronto in the province of Ontario.
    - Data extraction: A python code is written to scrap this Wikipedia page and extract the required data using pandas read_html method.

2. Latitude and Longitudinal co-ordinates of the Neighbourhoods in Toronto,Canada
    - Data source: http://cocl.us/Geospatial_data , which is a .csv file
    - Description: This is a .csv file that contains the list of neighbourhoods in Toronto with their Latitude and Longitudinal co-ordinates which is used to make Foursquare API calls.

- Data extraction: A python code is written to read the .csv file into a pandas dataframe for further processing

3. The data corresponding to various venues/businesses and their categories in Toronto neighbourhoods. This data will be used for clustering the neighbourhoods.
    - Data source: https://api.foursquare.com. We use Foursquare API to get the venue and venue category data related to each neighbourhood. The exact url for data extraction is: https://api.foursquare.com/ v2/venues/explore?&client_id=CLIENT_ID& client_secret=CLIENT_SECRET&v=VERSION&ll=lat,lng&radius=1000&limit=100
    - Description: This is a .json file that contains the details about a particular neighbourhood location and 100 different venues around each neighbourhood within a radius of 1km.
    - Data extraction: A python code is written to read the required fields in the .json file into a pandas dataframe for further processing.