**Data Science Capstone Project**

**Introduction to the Business Problem:**

We are going to use k-means algorithms to cluster Downtown Toronto , Toronto. Downtown Toronto is pretty expensive place which has great amenities such as restaurants,bars,cafes,gyms,playgrounds,schools,saloons all of which are some of the most quite sophisticated in the world.The people moving to this area spend a lot of money to move into such an expensive neighborhood because of their world class architecture/amenities and breathtaking landscapes.They want to leave no stones unturned before moving into this neighborhood so that they can have an amazing life which they won’t regret.Since Downtown Toronto is a large area it can have its differences which caters to different individuals.As a data scientist i have to provide the best bang for my client’s buck using various machine algorithms (i.e k-means in this case) and most importantly to leverage the power of data to make better choices. This problem focuses to subdivide the Downtown Toronto area into clusters so that clients can make better decisions based on their preferences.



source:Unsplash ,courtesy of Krishna C Koganti

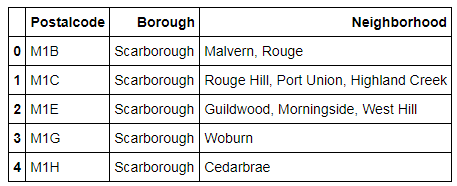
**Data**

As a data scientist the most important and time consuming part of this process is to obtain,clean and analyse the data.For this project i have obtained from 2 sources mainly :

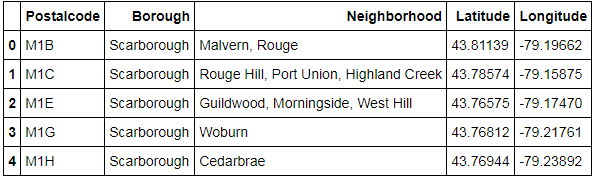
1. Wikipedia:https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_of\_Canada:\_M

2.Foursquare api

Wikipedia was used to obtain details of various Boroughs (since Downtown Toronto happens to be one such Borough) along with their respective postal code and neighborhood.



We then add the longitude and latitude corresponding to the postal code to the above table with the help of python libraries i.e pandas,geocoder



Now the above table only contains a small section,in reality the table is large consisting of data of all the neighborhoods (i.e their postal code,borough,neighborhood,lat,long).

However we are only concerned with the data relating to Downtown Toronto,so we create a new table which is a subset of the above table that contains the data only about Downtown Toronto. We use the pandas library of python to select data related to Downton Toronto.

