# BUSINESS PROBLEM REPORT

# CAPSTONE PROJECT



Jabin Choi

Coursera – IBM Data Science


12/30/2019

# CONTENTS

## PROJECT DESCRIPTION

Canada is one of many countries with immigrants for sure. Scarborough, in particular, is one of Canada's largest cities with immigrants and this project is useful for citizens planning to move to Scarborough by showing the results of comparing and analyzing dataset related to their average housing price and school ratings among the venues registered in the app called Foursquare. For more details on Foursquare, please refer to the following url:

** Foursquare URL: https:// foursquare.com

## PROJECT PURPOSE

The aim of this project is to clean and analyze features between neighbourhoods by sending out the names of them. The features mentioned here contain the mean of housing price and school rates etc. The result of data analysis helps people who are considering moving to Scarborough in Canada.

## WHAT IS FOURSQUARE API?

Foursquare is a social location service that provides users to explore the location around the world. The app enables users to download it through iPhone, Blackberry or Android devices and then connects their account to other social media accounts.

API itself is a RESTful set of addresses to which you can send requests. Users has nothing to download onto their server and it just allows developers to interact with the Foursquare platform.

The major purpose of using Foursquare API is to gather data source as it has a database of millions of places and the API provides the ability to perform location search and sharing as well as details on a business.

## CLUSTERING APPROACH

In this project, K-Means clustering which is a form of unsupervised machine learning is used for data analysis. By grouping similar data points (here, 'Neighborhoods' column) together, we will discover underlying patterns and analyze dataset related to the average housing price and school ratings among the venues.

# IMPORTED LIBRARIES

In [11]:
```
import requests
from bs4 import BeautifulSoup as bs
import re

import numpy as np
import pandas as pd
pd.set_option('display.max_columns', None)
pd.set_option('display.max_rows', None)

import json
from pandas.io.json import json_normalize  # transform Json file into a pandas dataframe

!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim  # convert an address into latitude and longitude values

import matplotlib.cm as cm
import matplotlib.colors as colors

from sklearn.cluster import KMeans

!conda install -c conda-forge folium=0.5.0 --yes
import folium
```

- Pandas: It offers data structures and operations for manipulating numerical tables
- Numpy: It supports for large, multi-dimensional arrays and matrices to operate on them
- JSON: It is used to transfer data as text that can be sent over a network
- Geopy: It makes users easy to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources.
- Matplotlib: It is a plotting library and it provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits
- KMeans: It includes algorithm clusters data by trying to separate samples in N groups of equal variances
- Foliums: It helps create several types of Leaflet maps. By default, it creates a map in a separate HTML file