Najdi Aboubakr

Final Report

Applied Data Science Capston

## Introduction: Business Problem

The aim of this project is to find a safe and secure location for opening of commercial establishments in Vancouver, Canada. Specifically, this report will be targeted to stakeholders interested in opening any business place like **Grocery Store** in **Vancouver City**, Canada.

The first task would be to **choose the safest borough** by analyzing crime data for opening a grocery store and **short listing a neighborhood**, where grocery stores are not amongst the most common venues, and yet **as close to the city as possible**.

We will make use of our data science tools to analyze data and focus on the safest borough and explore its neighborhoods and the 10 most common venues in each neighborhood so that the best neighborhood where grocery store is not amongst the most common venue can be selected.

## Data

Based on definition of our problem, factors that will influence our decision are:

* finding the safest borough based on crime statistics
* finding the most common venues
* choosing the right neighborhood within the borough

We will be using the geographical coordinates of Vancouver to plot neighborhoods in a borough that is safe and in the city's vicinity, and finally cluster our neighborhoods and present our findings.

Following data sources will be needed to extract/generate the required information:

* [**Part 1**: Using a real world data set from Kaggle containing the Vancouver Crimes from 2003 to 2019](https://colab.research.google.com/github/RamanujaSVL/Coursera_Capstone/blob/master/Capstone_Safest_Neighborhood_in_Vancouver.ipynb#part1): A dataset consisting of the crime statistics of each Neighborhood in Vancouver along with type of crime, recorded year, month and hour.
* [**Part 2**: Gathering additional information of the list of officially categorized boroughs in Vancouver from Wikipedia.](https://colab.research.google.com/github/RamanujaSVL/Coursera_Capstone/blob/master/Capstone_Safest_Neighborhood_in_Vancouver.ipynb#part2): Borough information will be used to map the existing data where each neighbourhood can be assigned with the right borough.
* [**Part 3**: Creating a new consolidated dataset of the Neighborhoods, along with their boroughs, crime data and the respective Neighbourhood's co-ordinates.](https://colab.research.google.com/github/RamanujaSVL/Coursera_Capstone/blob/master/Capstone_Safest_Neighborhood_in_Vancouver.ipynb#part3): This data will be fetched using OpenCage Geocoder to find the safest borough and explore the neighborhood by plotting it on maps using Folium and perform exploratory data analysis.
* [**Part 4**: Creating a new consolidated dataset of the Neighborhoods, boroughs, and the most common venues and the respective Neighbourhood along with co-ordinates.](https://colab.research.google.com/github/RamanujaSVL/Coursera_Capstone/blob/master/Capstone_Safest_Neighborhood_in_Vancouver.ipynb#part4): This data will be fetched using Four Square API to explore the neighborhood venues and to apply machine learning algorithm to cluster the neighborhoods and present the findings by plotting it on maps using Folium.