# The Capstone Project Report

## Business Problem

Assuming one wants to move to Vancouver and they are contemplating on which neighborhood to move to, they might want to know which neighborhood performs better in terms of crime rate. What would be more interesting rather is which neighborhoods has low **break and enter** crime rate or **vehicle accidents**. It all depends on what one really is moving there for.

All other factors held constant, an investor who wants to construct a business infrastructure might be interested in neighborhoods that have low **commercial break and enter** neighborhoods while a couple trying to raise a family might more keen on neighborhoods with low **residential break and enter** crime rate. This information is essential for such a large group of individuals given their reasons of moving to Vancouver or given the data, any other location in the world.

We will use the **2019 Vancouver crime dataset** to cluster which neighborhoods are ideal either to live in or do business in depending on the rate of specific crime rate in those neighborhoods. Further, we will transform the Foursquare neighborhoods data into a dataframe that can be merged with the clusters to find if there is high-view relation between the number of crimes and specific number of venues.

## Crime Dataset Justification and Exploration

For this exercise, I have obtained two sets of datasets:

1. Vancouver crime report data, the csv version of which the raw data is from here
2. *Foursquare location data*, to see how the Foursquare fair into our crime clusters

As we will be comparing with the venues dataset, we don't have information about the date when the each venue was constructed enough to related a venue with a crime occurrence but we know that the venue exist now (2019), therefore, we will consider 2019 crime dataset only. The data is stored on github