# Calgary Crime and Disorder Project

## Description and Preprocessing of Data

The Community_Crime_and_Disorder_Statistics.csv is made up of 114 848 thousand rows and 11 columns. The columns are as follows:

- Sector – This is the column that has the sector in which the offense was committed. The possible values for this column are CENTRE, NORTH, EAST, SOUTH, WEST, NORTHWEST, NORTHEAST, SOUTHEAST, and SOUTHWEST. However, the public usually divides the city into 4 sectors, NORTHWEST, NORTHEAST, SOUTHEAST, and SOUTHWEST. The addition of the other five sectors would probably cause confusion as to which communities make up each respective sector. Based on that I decided not to include this column in the analysis and visualisation.
- Community – States the name of the Community in which the offense was committed
- Group Category – States whether the offense is a crime or disorder
- Category – Goes into more detail of the type of crime that was committed. Some Crimes do not seem to be included in this dataset which is leading me to suspect that this dataset is incomplete. Crimes such as Homicides.
- Crime Count – I could not find the description for this column and the entries did not seem to follow any logical pattern I could think of. This led me to drop this column.
- Resident Count - This column gives the population for the community in which the offense was committed.
- Date – Contains the date in which the offense was committed. However, no matter situation, the date is always the 1$^{st}$ of the month in which the offense was committed, and the time is always 12:00:00. The Date column also displays the month and year, but we already have specific columns for both of those in the dataset. The only info of use we can get from this column is the time of day, AM or PM. In my Python program I preprocessed the column to only keep the AM/PM.
- Year – This is the Year in which the offense was committed
- Month – This is the Month in which the offense was committed
- ID – This is the ID for the offense that was committed. I removed this column as I did not have use for it in my analysis.
- Community Center Point – The is the geographical center point of the community in which the offense happened. I was not interested in a geographical visualisation therefore I dropped this column.

Project Description

For this project I was aiming to perform an analysis based on how Crime and Disorder were distributed over the communities, months, and years. I also wanted to find the most common type of crime for each community, month, and year. I decided to do this by writing a Python program that would output the programs as csv files. The data in each csv file is explained by the title of the csv file. I then used Tableau to create a visualisation based on the top 50 crime and disorder communities in Calgary, and the total crime and disorder count for each month of the year, whereas the csv files show the total number of crimes for each month and each year. The Python file is called **Calgary_Crime_and_Disorder_Analysis.py** and the original excel file is called **Community_Crime_and_Disorder_Statistics.csv.** The Tableau visualisation is called **Calgary_Crime_and_Disorder_Visualisation.twb.** There is a screen capture of the visualisation in case you do not have Tableau to open the visualisation.