**Machine Learning Project 1**

* 10 years of Crime volume by crime type and area.

2 types of questions

1. Fundamental operations - query the data for summary statistics, subsets, lists.
2. Problem Solving/ Investigation - Translate a business/social question into a technical statement and solve.

**Question 1**

* Join Population data with crime statistics table.
* Group by: Province and sum crimes over each year.
* What is the most dangerous province overall?
* Rank average crimes per year for all crime types.

Question 2

* Contrast volume vs density of crimes.
* Compare population density and crime rate - do places with higher **population** density tend to have higher crime **rates** (correlation)?

Question 3

* Group by: Crime Type.
* In which provinces does drug-related crime occur more than 1000 times a year (on average).
* Which Province has the highest number of stations? What is the average crime rate per station in this province?

Question 4

* Correlations: What types of crimes would you expect to go hand in hand?  does the data confirm?

Question 5:

* Plots/Visualizations: Boxplots and edf/Kernel density estimates of some crime type distributions across time.
* Histogram of the distribution of some crime type across provinces.
* Create overlayed time series plots showing trend of all crime types for Kwazulu/Natal.
* Create overlayed time series plots showing trend of total crime types for all provinces.

Question 6:

* Do any other exploration or analysis with the data apart from the questions mentioned above.