

# Bachelor of Informatics Honours

## Information & Knowledge in Organisations: Data Analytics

### *Assignment 2: Advising the South African Government on Crime*

Crime is a pervasive issue that affects societies worldwide, and South Africa is no exception. The country has grappled with high levels of crime, impacting individuals, communities, and the nation as a whole. To better understand the nature and dynamics of crime in South Africa, you are tasked with analysing a comprehensive dataset containing records of all crimes reported in the country between 2005 and 2016.

The primary goal of this analysis is to uncover trends, patterns, and insights that can help policymakers, law enforcement agencies, and communities address and mitigate crime effectively. By examining the data over a twelve-year period, I hope that you can identify temporal, spatial, and categorical patterns in different types of crime. This analysis seeks to provide a comprehensive overview of the crime landscape in South Africa, highlighting key areas for intervention and improvement.

The dataset used for this assignment includes detailed information on crimes reported in South Africa, including the province, the police station, the crime category, and the count for each station-category combination per year for the 12 year period. Your task is to conduct a statistical analysis and modelling of the data to identify key areas of focus for the government.

It is up to you to decide what aspects you will focus on for the analysis, but the following questions provide an indication of the types of analyses and insights that I expect you to consider. For some, you may need to identify additional data sources to augment the crime data. You are free to add any other data that you see fit to this dataset.

- How has the overall crime rate in South Africa changed over the twelve-year period?
- Are certain types of crimes becoming more or less prevalent over time?
- Which provinces or regions have the highest and lowest crime rates?
- Are there specific areas within provinces or regions that experience higher levels of certain types of crimes?
- How do crime rates vary between urban, suburban, and rural areas?
- What are the most common types of crimes reported in South Africa?
- Are there any trends or patterns in the occurrence of specific types of crimes?
- Are certain types of crimes more prevalent in specific regions or provinces?
- Can predictive models be developed to forecast crime trends based on historical data?
- How accurate are these models in predicting the occurrence of certain types of crimes?
- How do factors such as poverty or unemployment correlate with crime rates?
- How does the crime rate in South Africa compare to other countries?
- Based on historical data, what are the potential trends in crime rates in South Africa?
- Are there any correlations between specific policy changes and changes in crime?
- Which areas in South Africa are considered crime hotspots, and what factors contribute to this status?

The assignment has two parts:

1. Your quarto notebook that contains your EDA and modeling
2. A PDF report produced using quarto describing the findings from your analysis.

Your report should include the following elements:

1. A data-based descriptive background on crime in South Africa, including current challenges and trends over the period under investigation
2. Motivation for the selection of relevant variables from the comprehensive datasets based on your EDA of key variables in the datasets.
3. You should include key summary statistics, relationships, and visualisations in the report as well as the results of statistical models, predictions, and comparisons, with appropriate interpretations (using both NHST and machine learning approaches). It is expected that you will use regression, classification, and clustering analyses.
4. Presentation of recommendations for the South African government, based on the results of your analysis and modelling.

The report should be written in a clear and concise manner and should be supported by appropriate tables and figures. Alongside your report PDF, you will also submit your analysis code in a zip file. **The assignment should be completed in groups of three (3).**

**The following elements are required for submission:**

- Your analysis notebook (.qmd)
- The final PDF report (.pdf)
- The quarto file that produced the report (.qmd)
- Any other data/style/supporting files

Submit all the above in a single .zip file.

**Assessment criteria**

1. Data-based descriptive background on crime in South Africa (10)
  - Clarity and coherence of the presentation (writing, tables, figures) (5)
  - Quality of the code used to produce descriptive statistics and visuals (5)
3. Exploratory data analysis (25)
  - Summary statistics and visualizations (10)
  - Relationships (correlations, t-tests, etc.) between variables (5)
  - Quality of the code used to produce the analysis (10)
5. Statistical modelling (55)
  - regression analyses (15)
  - classification analyses (15)
  - clustering analyses (15)
  - Clarity of written explanation of methods, results, interpretations (10)
6. Recommendations (5)
  - Relevance and feasibility of evidence-based recommendations (5)
7. Code, document, and Data Management (10)
  - Clarity and organization of code (5)
  - Style and organization of the report (5)

**See the following resources for more information making documents with quarto.**

- <https://quarto.org/docs/gallery/>
- <https://youtu.be/lbE5uOqfT70?si=IQ4pZ4Nxn-2bqq6R>
- <https://www.youtube.com/watch?v=HI9TvhlfV4>
- <https://nrennie.rbind.io/blog/making-pretty-pdf-quarto/>

**Due Date: Friday 10 May, 13h00 (via SUNLearn link)**