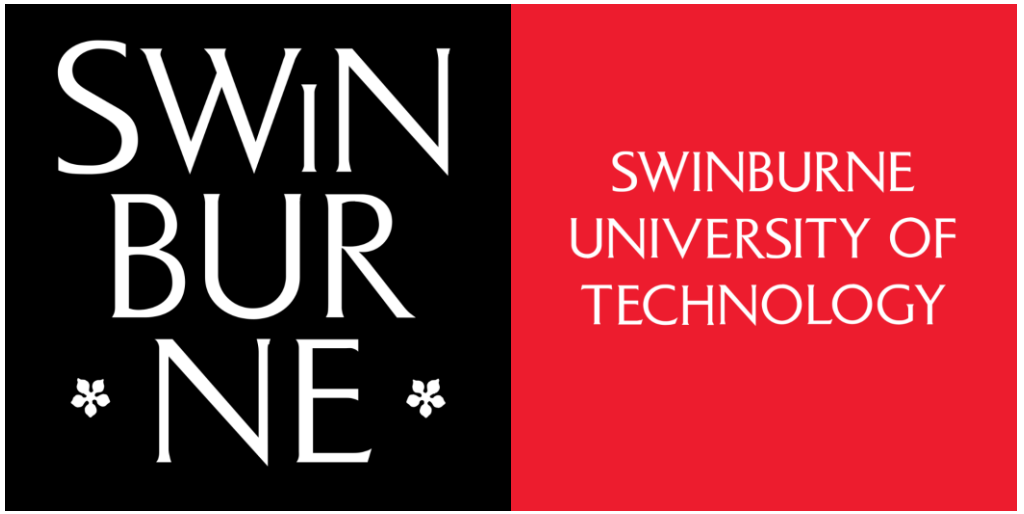
COS30045 DATA VISUALISATION

Team Name – CL01_G21
Name – Yulan yaneth karunanayake
ID - 105107919



DATA VISUALISATION PROJECT

Submission guidelines

- Draft of Project
- Design Book
- Link to team GitHub classroom
- KNIME file

DATA VISUALIZATION PROJECT -

THE ASSIGNMENT'S GOAL

- The background, motivation, and ultimate goal of the visualisation project are all spelled out in detail in the introduction section that has been created.
- The necessary dataset has been found and gathered by the team, and its source and governance considerations have been carefully documented.
- To guarantee a thorough comprehension of every area and its significance, a thorough examination of the data dictionary was conducted.
- KNIME has been used to carry out preliminary data processing tasks, such as determining missing values, performing necessary transformations, and calculating important summary statistics.
- Across jurisdictions and demographics, exploratory data analysis (EDA) has been used to find trends, patterns, and possible problems with data quality.
  For COS30045, a special GitHub Classroom repository has been made to facilitate collaborative code management and tracking.

The goal of this project is to create an interactive dashboard that displays information from drug tests conducted at roadside locations throughout Australia. Using official government enforcement information, the dashboard will enable users to examine variations in testing results by state, year, and detection stage. In order to improve road safety across the country, the main goal is to assist law enforcement, legislators, and the general public in comprehending testing trends, locating high-risk regions, and assessing the efficacy of present enforcement tactics.

WORK COMPLETED SO FAR

- So far, I have written a clear and well-structured project introduction that outlines the reasoning and purpose behind visualising roadside drug testing data. The dataset, sourced from the **Road Safety Enforcement Data Dictionary** provided by the Australian Government, has been collected and properly documented. I thoroughly reviewed the **data dictionary** to understand the key attributes and their relevance to the project.
- Using **KNIME**, I initiated the data cleaning process by addressing missing values, correcting inconsistencies, and filtering for the most accurate records based on the BEST_DETECTION_METHOD flag. I also conducted **preliminary exploratory data analysis (EDA)** to identify early trends and enforcement patterns across different states and demographics.
- As I am working individually, I have created a personal **GitHub Classroom repository for COS30045** to manage version control, organise project files, and ensure a smooth development workflow. This setup allows me to keep track of all changes efficiently and maintain a structured approach throughout the project.

IN PROGRESS

Right now, my main goal is to improve the dashboard's interaction design and visualisation arrangement. This entails choosing the best chart styles to accurately depict important trends in the data as well as setting up filters for state, year, and detection stage. In order to guarantee uniformity in characteristics like testing phases and jurisdiction names, I am also carrying on with data transformation and cleansing. I have started creating interface mockups and using user needs to inform layout choices in an effort to make the website more usable. The goal of these initiatives is to produce a dashboard that is both informative and easy to use for all target groups.

BLOCKERS AND CHALLENGES

Challenge 1: Some jurisdictions—particularly **Tasmania** and the **Northern Territory**—provided incomplete or inconsistent data. Key figures such as total test counts, charges, or arrests were missing in certain records, making it difficult to achieve accurate cross-state comparisons.

Challenge 2: While **KNIME** is effective for data processing and basic visualisation, it has limitations when it comes to generating the **interactive and dynamic layouts** required for the final dashboard. This has created an additional layer of complexity as I plan for external tools or manual solutions.

Challenge 3: Comparing results across different **drug testing stages** (e.g., roadside vs. laboratory confirmation) is difficult because this level of detail is **only available in the 2023 data**. The lack of stage-level data in earlier years restricts longitudinal comparisons.

In addition to these technical challenges, working **independently without a team member** has made it more difficult to manage the **overall workload**. Balancing data preparation, design planning, technical development, and documentation alone has been demanding, but I am continuing to make steady progress.

Potential Solutions for Blockers and Challenges

Challenge 1: Incomplete Data (e.g., TAS, NT)
- Use **proxy values** (e.g., fines + charges) where test count data is missing
- Focus analysis on **states with complete and reliable records**

Challenge 2: KNIME Visual Limitations
- Use **KNIME primarily for data cleaning and processing**
- Export cleaned data to build visualisations in **D3.js** or **Chart.js** for better interactivity

Challenge 3: Limited Testing Stage Data (Only 2023)
- Add **clear labels and explanations** to charts using 2023 stage-level data
- Focus year-over-year comparisons on **overall test volumes and positive rates**

Challenge 4: Working Individually Without a Team Member
- Use a **structured project timeline** to manage tasks and reduce overload
- Focus on one task at a time (e.g., clean → analyse → design) to maintain progress

FUTURE STEPS

Completing the report's Visualization Design, Interaction Design, and Website Design parts is the next step. Creating sketches and prototypes will be my first step in defining the dashboard's general structure. I intend to begin developing the visualisation on the technical side by loading the cleaned dataset, configuring shared constants, and generating the essential interactive elements and chart components.

The goal will be to provide an interface that is easy to use and understand. Important interactions will be added to improve usability, like filtering by year, jurisdiction, and testing stage. Every element will be gradually combined to create a functional prototype, which will be improved in response to ongoing testing and feedback.

YEAR	JURISDICTI...	LOCATION	AGE_GROUP	Sum(COUNT)	Sum(FINES)	Sum(CHAR...
2023	ACT	Major Cities of Au	17-25	46	3	46
2023	ACT	Major Cities of Au	26-39	83	3	83
2023	ACT	Major Cities of Au	40-64	73	3	73
2023	ACT	Major Cities of Au	65 and over	1	3	1
2023	NSW	Inner Regional Aus	0-16	21	1	13
2023	NSW	Inner Regional Aus	17-25	3,080	1,504	1,806

hough the Exploratory Data Analysis (EDA) produced insi
dings, a number of difficulties were faced. These include
orcement outcome data from places like the Northern Te
onsistent reporting of testing stages, and an absence of
ail in previous datasets. A thorough examination of mult
ting methods was also limited because not all states pro
ults that were verified by laboratories. Notwithstanding t
wbacks, the analysis was successful in identifying signif
orcement hotspots, noteworthy demographic trends, an
erational approach variations. These findings will all be ta
nsideration when designing the dashboard and will help l
orcement agencies and policymakers make better decis

Hawthorn Campus

BACHELOR OF COMPUTER SCIENCE

Semester 1, 2025

COS30045 – Data Visualisation

SWINBURNE UNIVERSITY OF TECHNOLOGY

Project Stand Up 1

Visualising Drug Testing Enforcement in Australia: State-by-State Analysis

Team Name: CL01_G21

Name – Yulan Yaneth Karunanayake

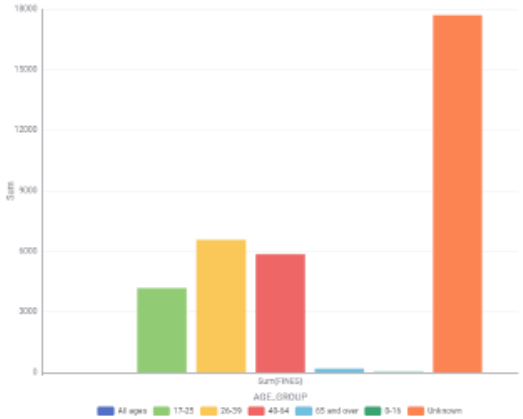
ID - 105107919

Attribute (2023)	Description
Complete drug test	471,217
Drug test result that is positive	11.4%
Fines issued	13,784
Charges	28,716
Arrests	210
Testing stages	Indicator, Laboratory, Not applicable
Filters	Jurisdiction, Year, Testing stage
Drug test conducted and result overtime	From 2008 to 2023 with steady increase
Tests per 10,000 licences	Highest in SA (392), lowes ACT (31)
Update frequency	Annually

rivers. Consistent naming conventions for testing stages and jurisdictions
ere also ensured by applying data standardization techniques.

.3 Data Exploration

r Chart



Data Collection Process

dashboard's dataset is created using police enforcement records tha
ferent states and territories. Jurisdictions have t
ific fields in their submission
the completene

SCREENSHOTS

GEN AI DECLARATION

I admit that generative AI technologies helped with this project in a number of ways. These tools aided in the formulation of ideas, the organization of summaries and important insights in the data analysis documentation, and the enhancement of written material. Nonetheless, every essential activity was finished on its own, including the analysis, coding, processing of KNIME data, and design choices. AI was only utilized as a supporting tool to improve efficiency, clarity, and communication, and the final results were carefully examined and validated.