

Internal Competition Guidelines: Individual Challenge

Instructions

Welcome to the first round of our internal competition! From here, we will select the top 12 students to proceed to the second round—a group challenge.

Participants will be provided with a dataset and are expected to build a complete data science pipeline, including:

- Data cleaning and preprocessing
- Exploratory Data Analysis (EDA)
- Feature engineering
- Model selection and training
- Evaluation and interpretation
- Communication of findings through a voice-over slide presentation

Deliverables:

1. A codebase (Jupyter Notebook)
2. A slide deck with voice-over narration (maximum 5 minutes)

Submission Deadline: 7 June 2025, 17:00

Note: No extensions will be granted for this challenge.

Data Abstract

The *Marriages and Divorces (MD)* dataset is one of three primary sources of marriage and divorce statistics in South Africa. Unlike population censuses and household sample surveys, the MD dataset is compiled from administrative data and based on continuous recording via civil registration systems and administrative records.

Statistics South Africa (Stats SA) regularly publishes this data series, beginning in 2006. The most recent dataset available is MD 2023.

Divorce Data: The data on divorces are obtained from regional courts that handle divorce cases. These records are based on successfully granted divorce decrees issued by the Department of Justice and Constitutional Development (DoJCD). Divorce records may originate from marriages and filings in different years.

Data Description: [Click here](#) for full variable definitions and descriptions.

Use the provided dataset to predict the **main grounds for divorce** (`GroundsForDivorce`).

Tips for Participants

- **Understand the problem:** Carefully read the dataset documentation and understand your data.
- **Tell a story:** Use your slides to clearly and compellingly communicate your process and findings.
- **Visuals matter:** Keep your charts clean, focused, and uncluttered.
- **Keep code clean:** Write modular, readable, and well-commented code.
- **Validate your model:** Use appropriate evaluation metrics and watch for overfitting.
- **Practice the voice-over:** Speak clearly, and explain the reasoning behind your decisions.

Bonus points for creativity, originality of insights, and actionable recommendations!

Rubric for Presentation (60% of Total Score)

Criteria	Weight	Score (0–10)
Clarity of problem definition and objectives	10%	
Narrative and flow of the story	15%	
Visual quality and slide design	10%	
Insightfulness of analysis	15%	
Explanation of model and evaluation metrics	10%	
Voice-over delivery (clarity, pace, engagement)	10%	
Total Presentation Score (out of 60)		

Rubric for Code (40% of Total Score)

Criteria	Weight	Score (0–10)
Code structure and readability	10%	
Data preprocessing and cleaning techniques	10%	
EDA and feature engineering	10%	
Modeling approach and justification	5%	
Evaluation and interpretation of results	5%	
Total Code Score (out of 40)		