

Women/Men Focused Training

Ghana R-Users Trainers

Group Project 2 – Analysing and Reporting Insights from a Dataset

Learning Objectives:

- Apply data preprocessing, visualization, and analysis techniques from previous sessions
- Perform clustering, PCA, or regression analysis on a real-world dataset
- Interpret and present findings using Quarto for reporting

Session Outline

1. Introduction to Group Project

Project Objective

Each group will:

1. **Choose a dataset** (From the provided data)
2. **Preprocess the data** (handle missing values, normalize, etc.)
3. **Apply analysis techniques** (PCA, clustering, regression, etc.)
4. **Create visualizations and customise** using `ggplot2`
5. **Compile findings into a Quarto report**

Datasets to Choose From:

- **Student Performance data** – Use clustering to identify groups
- **Top 100 tiktokers data** – Perform PCA and regression analysis

6. Group Presentations

Each group will present:

1. **The cleaned dataset** – How they handled missing values and normalization.
2. **EDA insights** – Trends, correlations, and summary statistics.
3. **Advanced analysis** – PCA, clustering, or regression models used.
4. **Visualizations** – Any plots that support their conclusions.
5. **Key takeaways** – Summary of findings and recommendations.

8. Recap and Q&A

- Recap key skills: EDA, PCA, clustering, and regression.
- Discuss challenges and insights from the project.