Tanzania Water Wells Project

MUNIU

PAUL

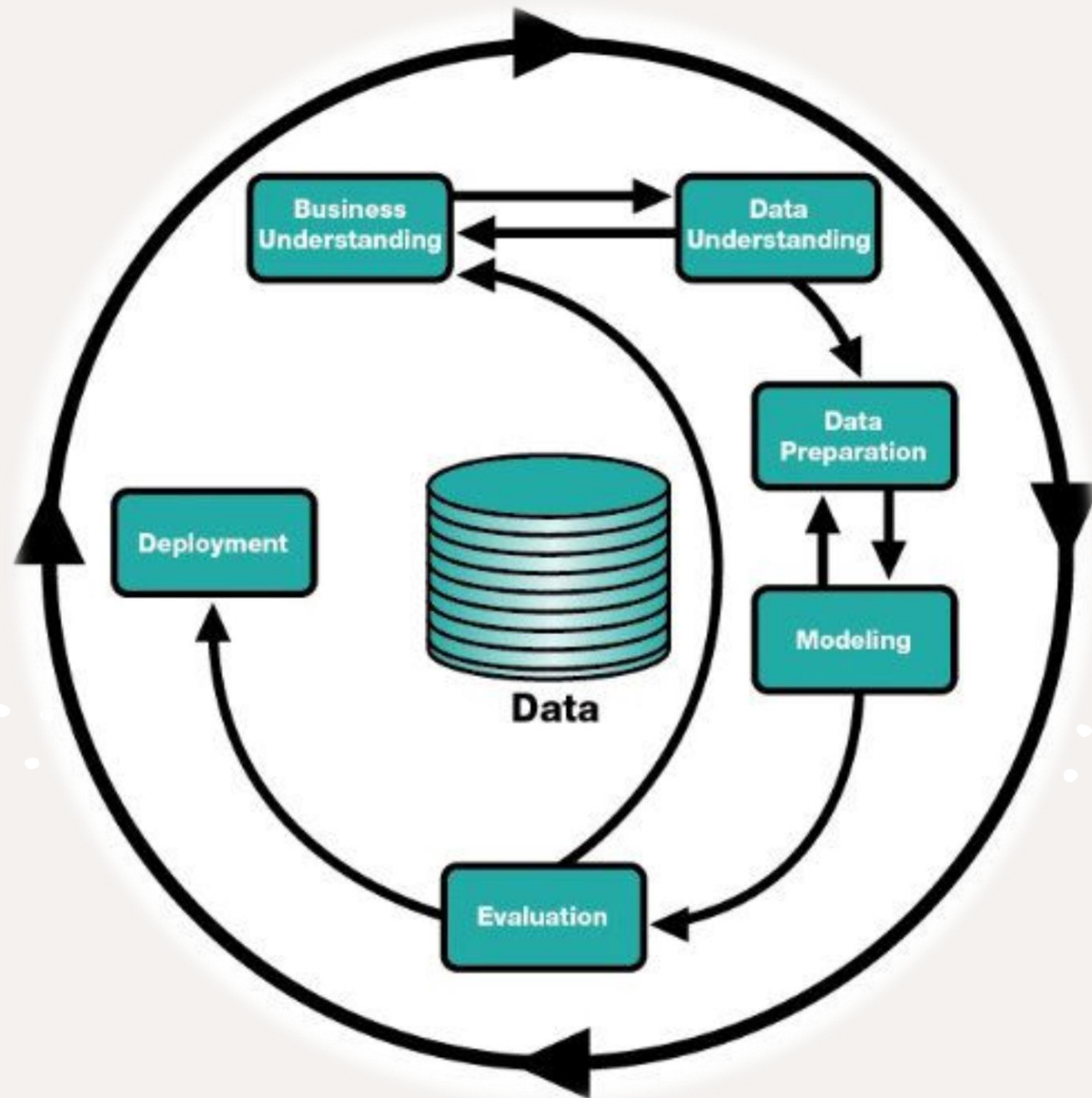# Overview

Stakeholder: `WaterAid International.`

Objective:

➤ To predict the functionality of water points in Tanzania using machine learning.

➤ To assist WaterAid in making data-driven decisions to improve water access.

Business and Data Understanding

Key Takeaways:

- ❖ Understanding water point status.

- ❖ Identifying the best predictive model.

- ❖ Providing actionable recommendations.

Dataset:

• Includes attributes such as region, construction year, water quality, and status group (functional, non-functional, functional but needs repair).

Features in our Dataset.
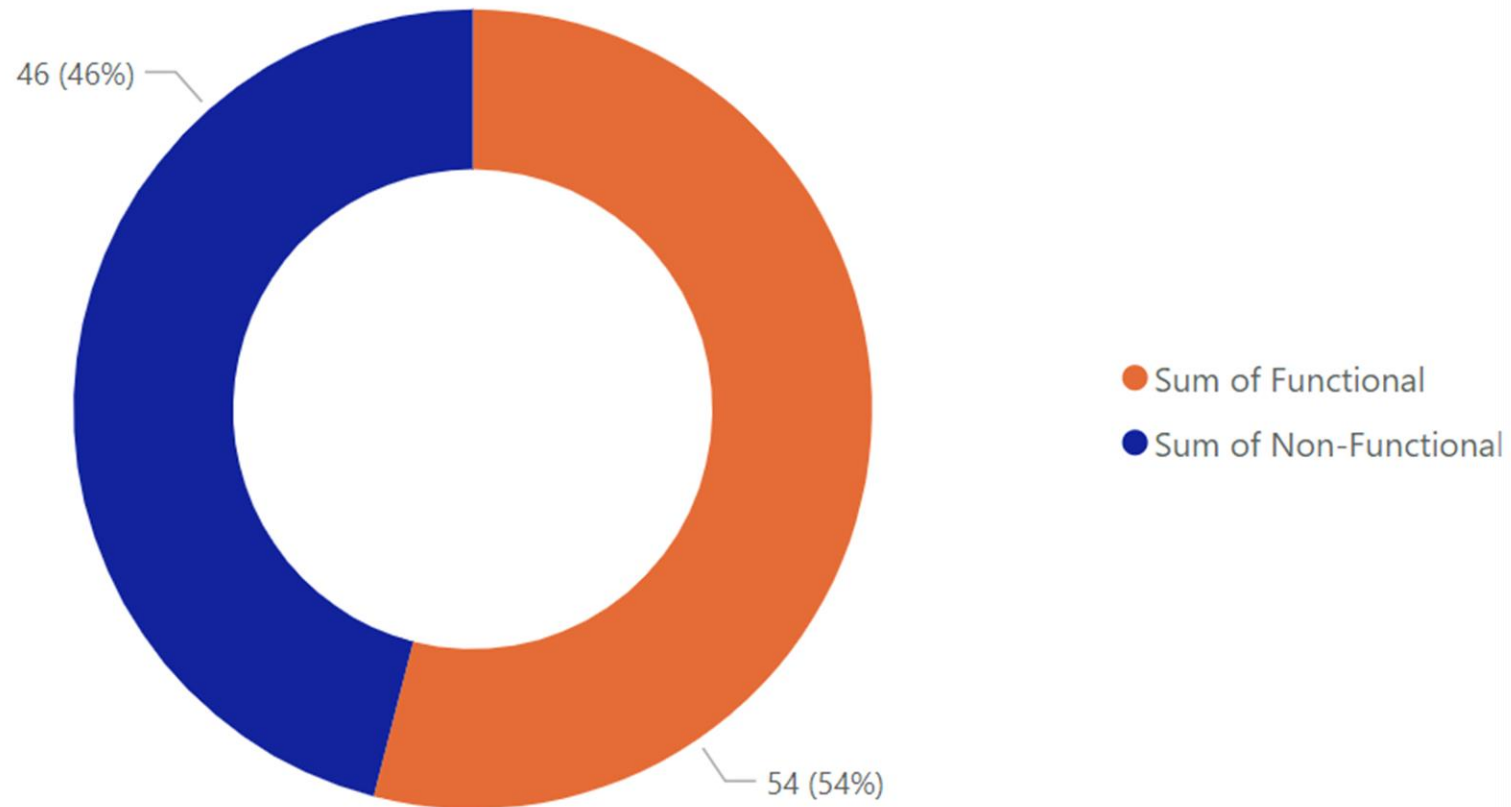

FEATURES ~ VISUALIZATION.
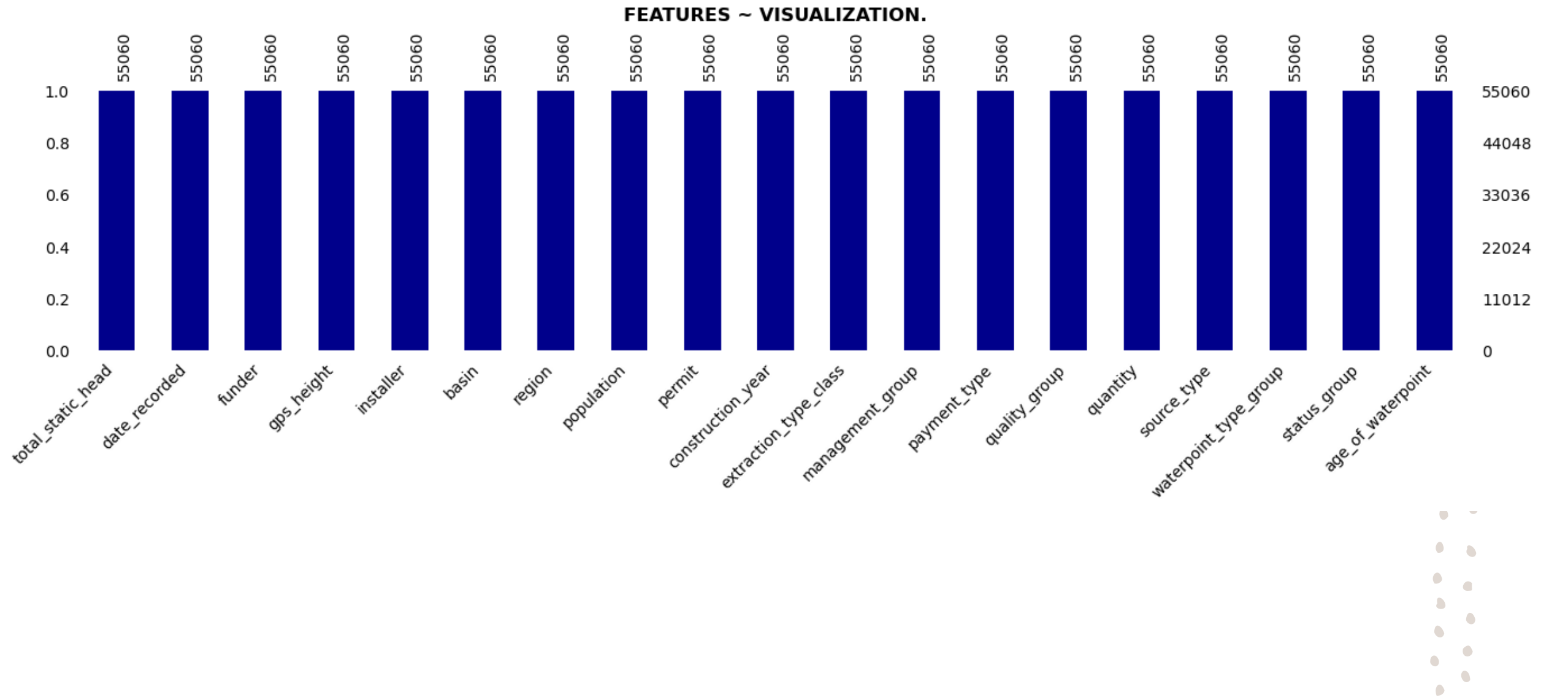
# Exploratory Data Analysis

EDA Insights:

- Distribution of water point statuses across different regions and types.

- Key features impacting functionality.

- Visualizations depicting data distributions and relationships.

DISTRIBUTION OF FUNCTIONAL AND NON-FUNCTIONAL WELLS.

46 (46%)

54 (54%)

- Sum of Functional
- Sum of Non-Functional

Data Preprocessing.

✓ Handling missing values

✓ Encoding categorical variables

✓ Feature scaling

✓ Addressing class imbalance using SMOTE

✓ Feature Selection: Identified and retained the most important features for model training.
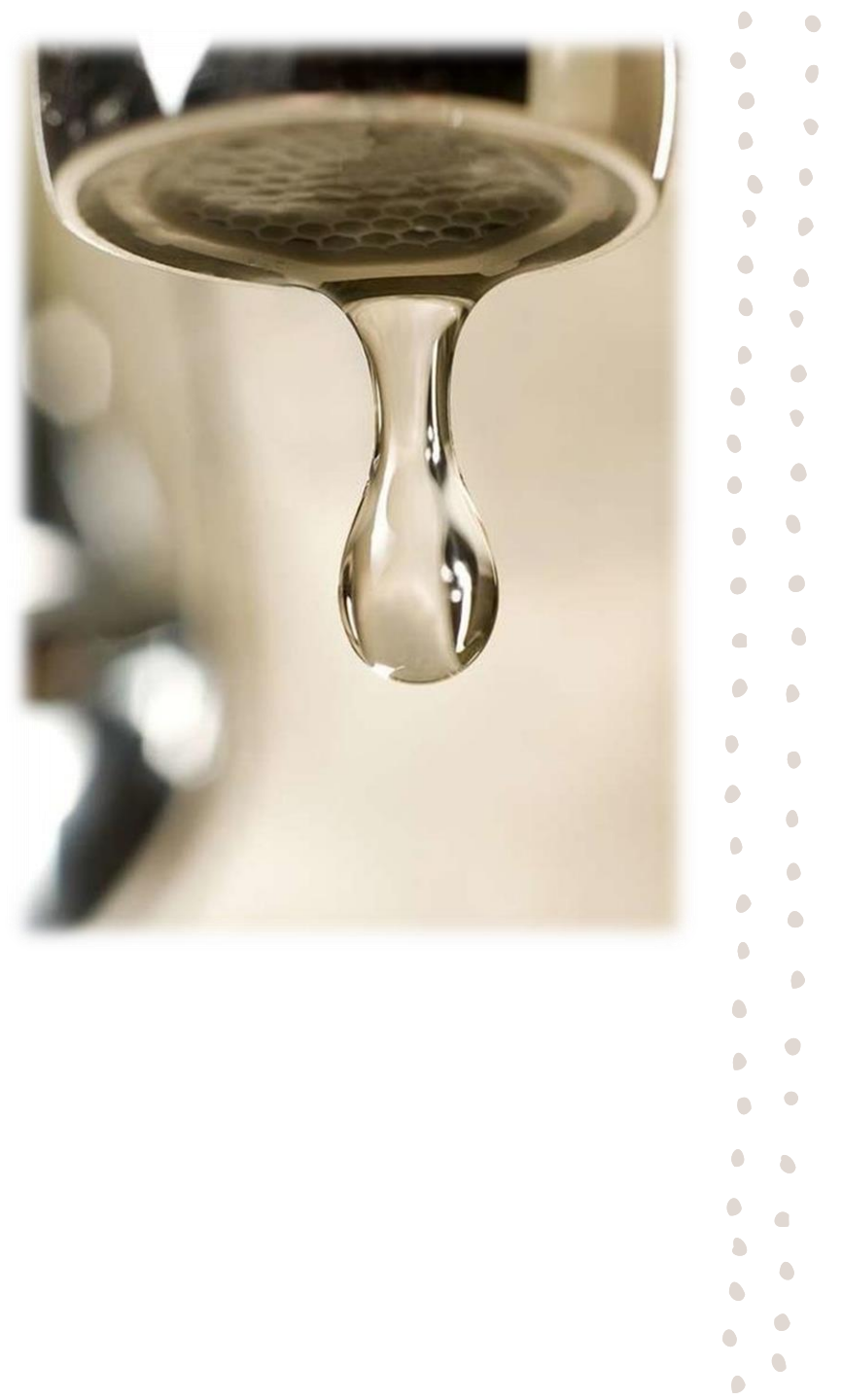
Feature Importance

- Highlighted the most important features influencing water point functionality.

- Interpretation:

These insights can guide WaterAid in prioritizing factors needing attention to maintain water point functionality.

Modeling Models.

✓ Dummy Classifier

✓ Decision Tree

✓ Random Forest

✓ K-Nearest Neighbors (KNN)

✓ XGBoost

Approach: Trained multiple models and evaluated their performance

Model Evaluation.

EVALUATION METRICS :
✓ Accuracy
✓ Precision
✓ Recall
✓ F1 Score
✓ ROC AUC Score

Top Performer:

•Random Forest classifier with 80% accuracy, 80% precision, 80% recall, 80% F1 score, and 88% ROC AUC score.

Next Steps.

- **Implementation:** Deploy the Random Forest model in a production environment to predict water point functionality in real time.

- **Monitoring:** Regularly update the model with new data to maintain accuracy.

- **Further Analysis:** Explore additional factors and data sources that may impact water point functionality.

"Nothing in life is to be feared; it is only to be understood. Now is the time to understand more, so that we may fear less."

- Marie Curie.

Thank you.

- ❖ Muniu Paul.

- ❖ Data Scientist.

- ❖ muniupauljr@gmail.com

- ❖ https://github.com/PaulMuniu

- ❖ https://www.linkedin.com/in/paul-muniu-a641ba187/www.firstupconsultants.com