PREDICTING FAULTY PUMPS

DATA MINING FOR SUSTAINABLE WATER MANAGEMENT

MIGUEL BARRIOLA ARRANZ

05/21/2024

INTRODUCTION

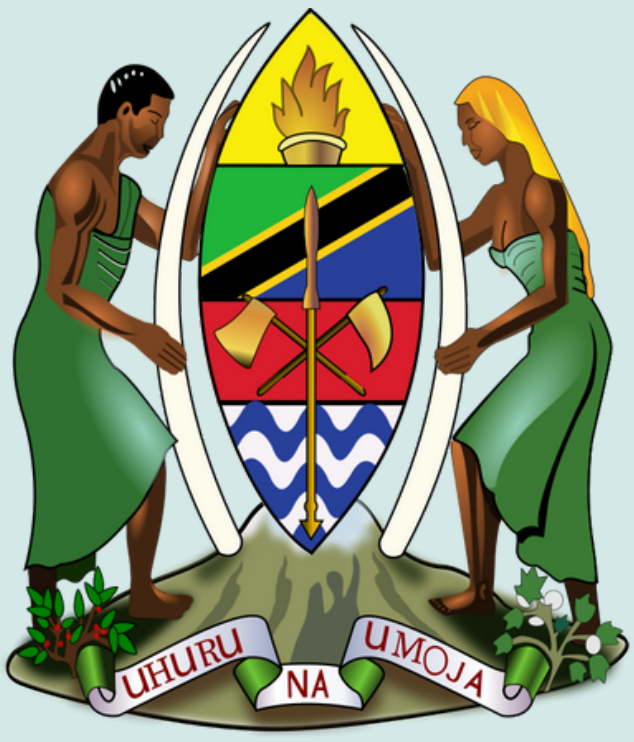
- Project to predict the status of water pumps in Tanzania

- Objectives of the analysis:
  - Classify pumps in Tanzania as functional, or non-functional
  - Improve water access in Tanzania

STAKEHOLDERS

- Stakeholders, including government agencies and NGOs, will use these findings to prioritize and streamline efforts towards ensuring reliable water access.

- Primary stakeholders for this project are the Tanzanian government and international development organizations focused on improving water access in the region.



BUSINESS CASE

- Core objective:

1. Enable the identification of functional and non-functional pumps in Tanzania

- Results implications:

1. Guide decisions on maintenance, investments, and resource allocation.
2. Support sustainable water management in Tanzania.

DATA

DATASET OVERVIEW

Driven Data provided the following datasets:

- SUBMISSIONFORMAT
- TEST_SET_VALUES
- TRAINING_SET_LABELS
- TRAINING_SET_VALUES

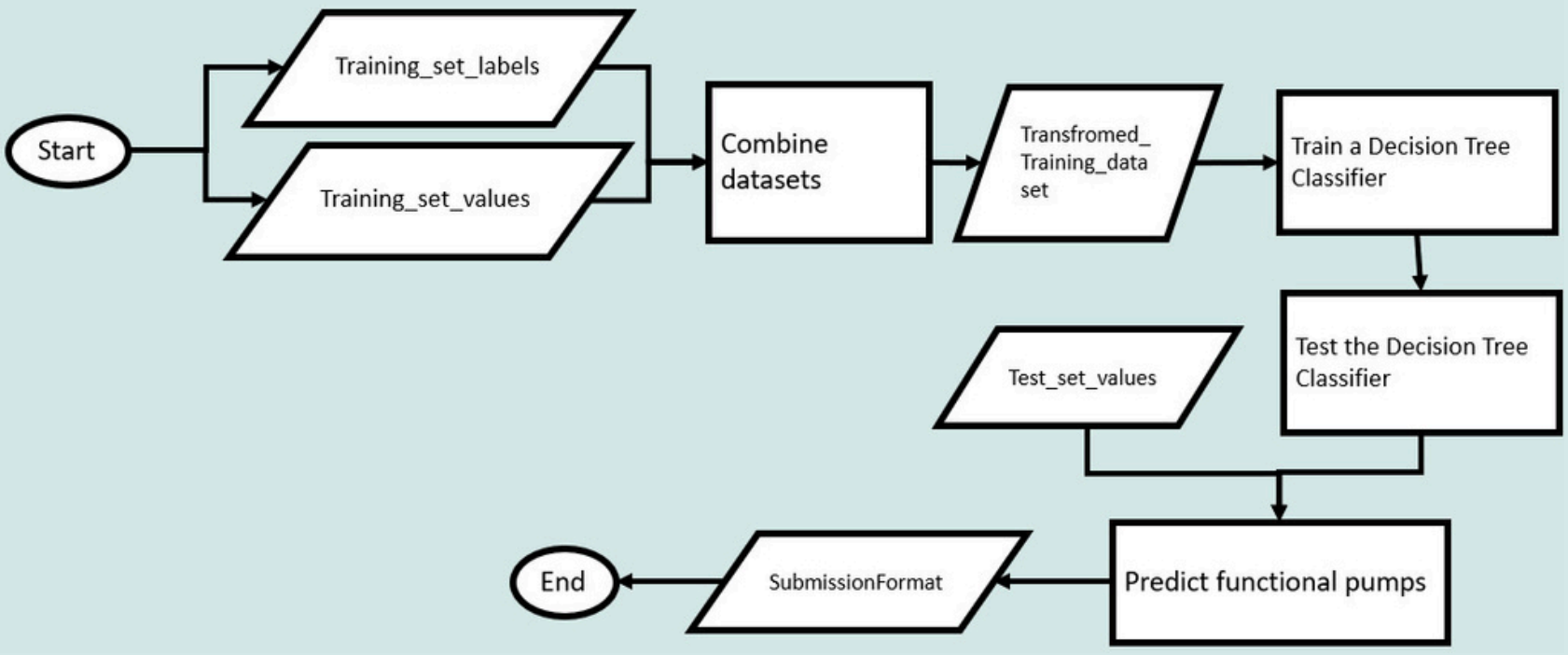
DATASET DESCRIPTION

- TRAINING_SET_LABELS and TRAINING_SET_VALUES were used for model building

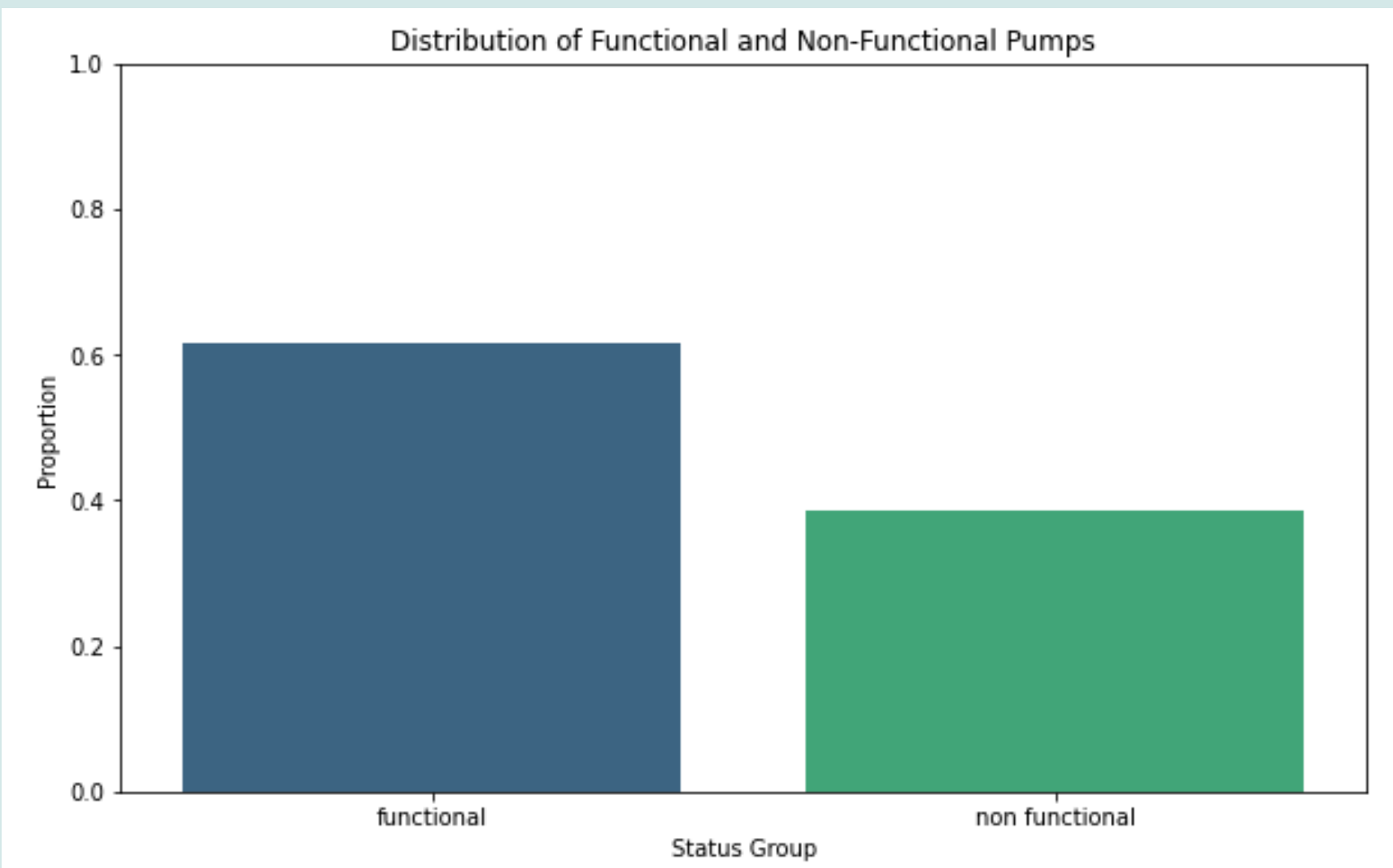
- Same transformations applied to TEST_SET_VALUES for predictions

- SUBMISSIONFORMAT contains the predicted pump status

PROJECT OVERVIEW



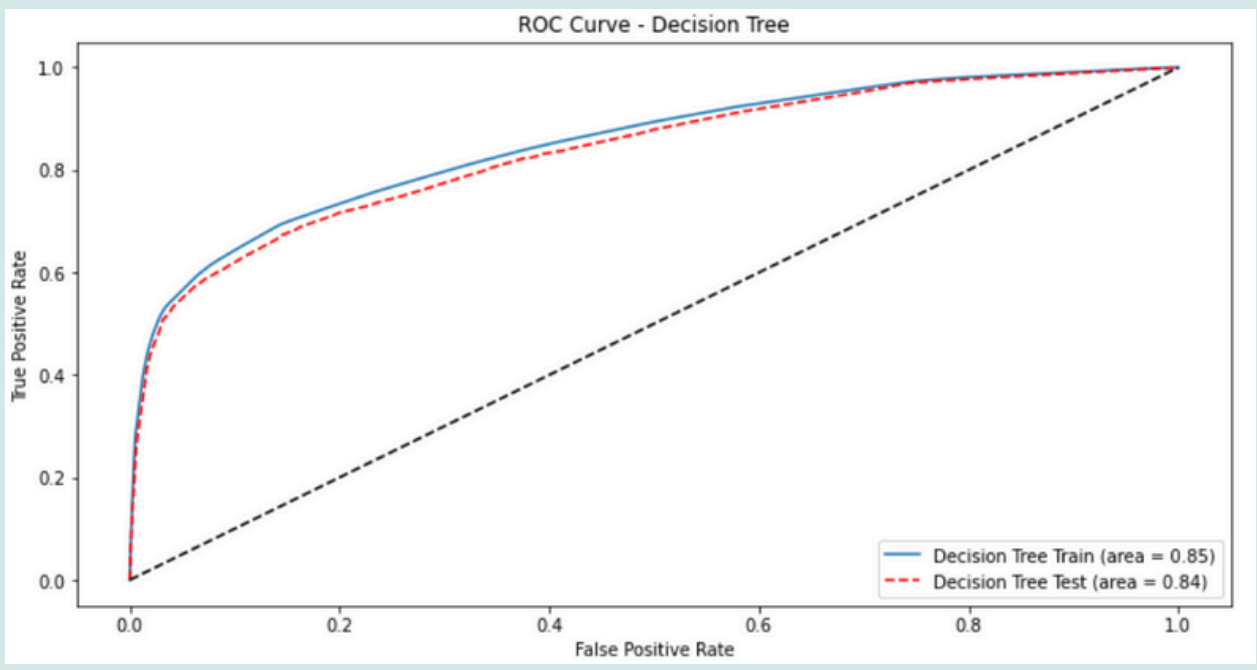
PUMPS STATUS



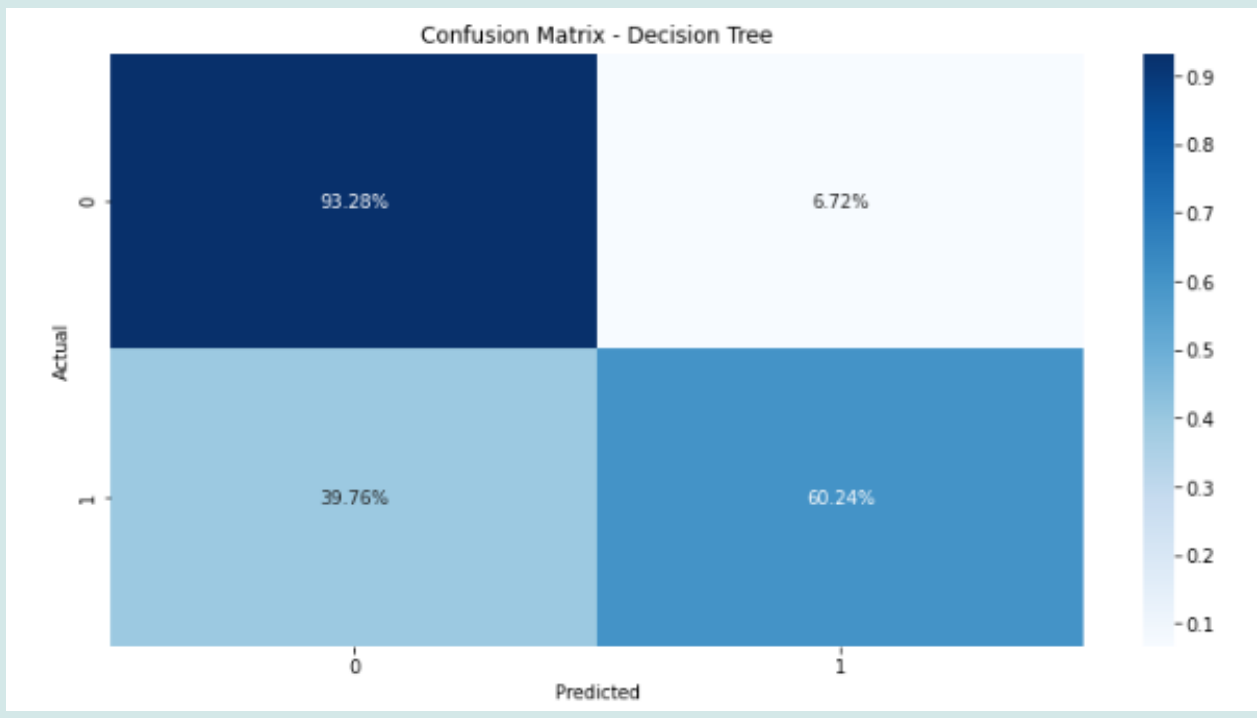
MODELING APPROACH

- Models tested:
  - Logistic Regression
  - Decision Tree

- Evaluation metrics considered were ROC, AUC, and the confusion matrix



Hyperparameter tuning process to get the optimal parameters

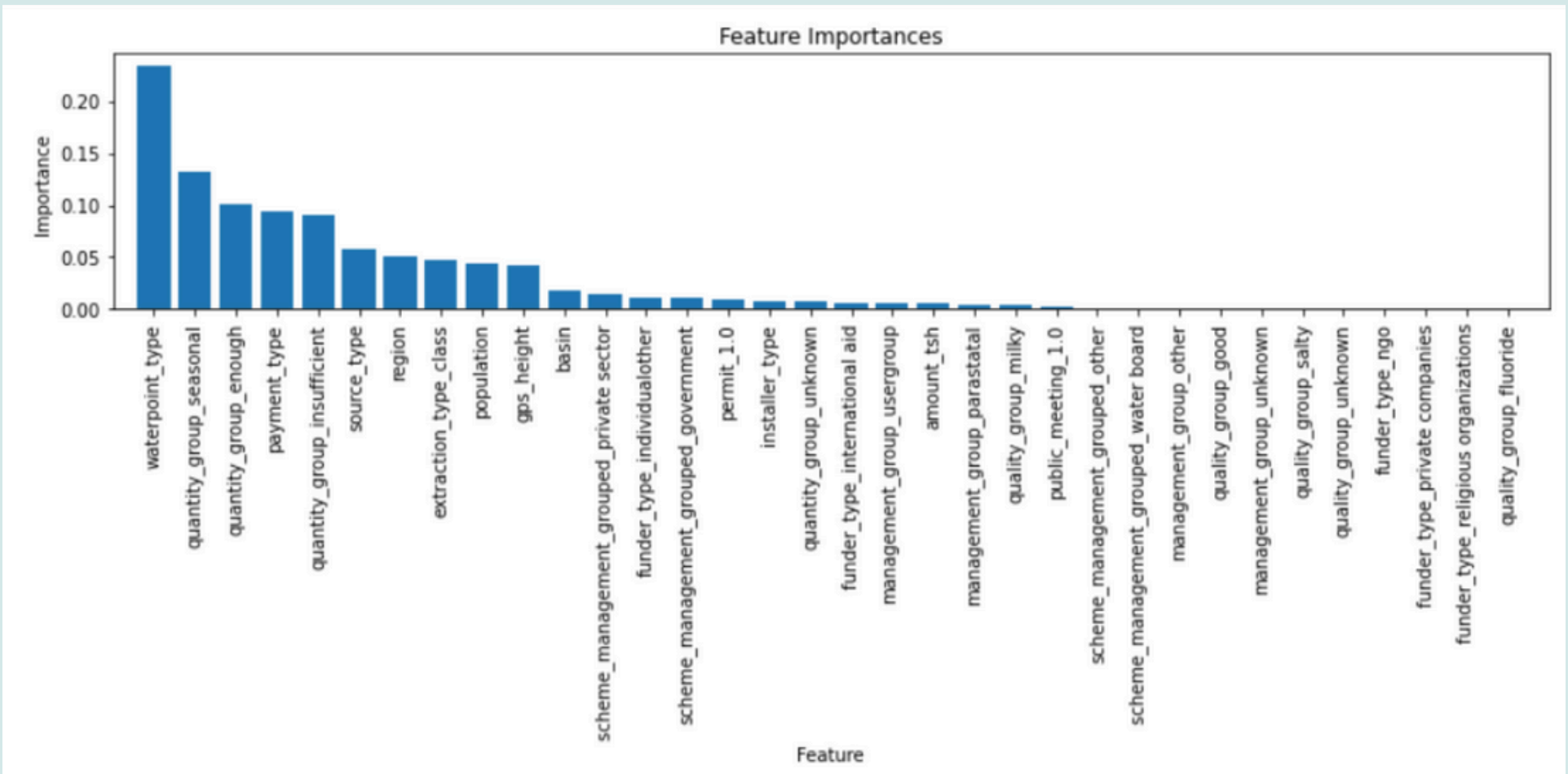


Focus on reducing false negatives for prompt repairs

FEATURE IMPORTANCE

- Here are the most important variables that better discriminate between functional and non-functional:

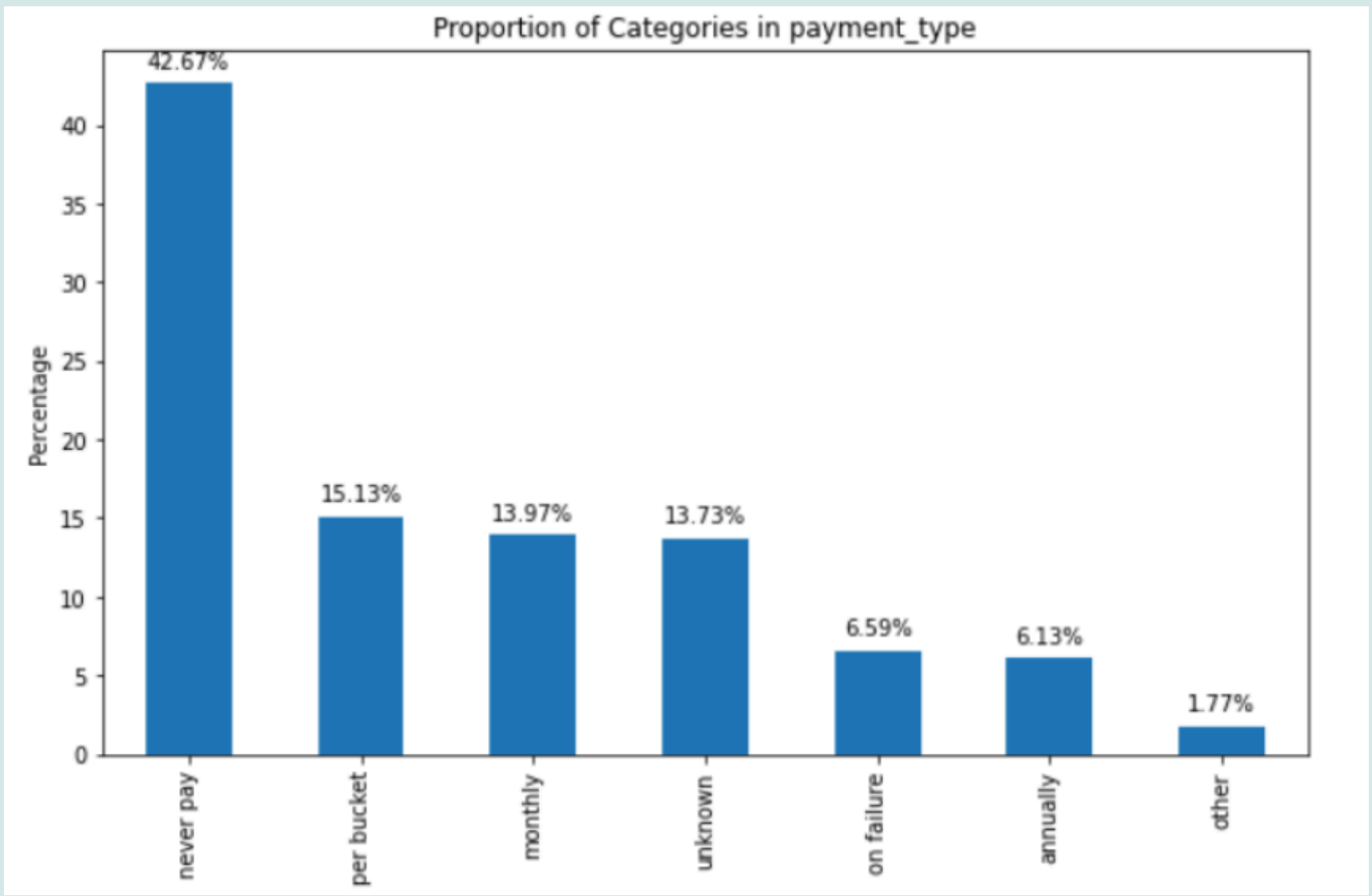
    a. waterpoint_type (ie. water access point)
    b. quantity_group (ie. quantity of water)
    c. payment_type (ie. Payment method)



RECOMMENDATIONS

1. Align Payment Plans:

Providing financial assistance more accessible (with monthly or per-bucket payments)



RECOMMENDATIONS

2. Use Dry Pumps as Indicators:

Using dry as an indicator helps identify non-functional pumps for targeted repairs

quantity_group	functional	non-functional
dry	0.49	26.52
enough	67.11	40.04
insufficient	24.54	25.25
seasonal	7.21	5.74
unknown	0.66	2.46

3. Identify Non-Common Types:

Invest in communal standpipe multiple points to best detect pump functionality

waterpoint_type	functional	non functional
cattle trough	0.26	0.13
communal standpipe	54.95	37.40
communal standpipe multiple	6.93	14.11
dam	0.02	0.00
hand pump	33.49	24.77
improved spring	1.75	0.60
other	2.60	22.99

THANK YOU VERY MUCH

Miguel Barriola Arranz

- LinkedIn: https://www.linkedin.com/in/miguel-barriola-arranz/

- Medium: https://medium.com/@mbarriolaarranz