PREDICTING FAULTY PUMPS

DATA MINING FOR SUSTAINABLE WATER MANAGEMENT

MIGUEL BARRIOLA ARRANZ

07/12/2024

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

- Government agencies and NGOs
- Tanzanian government and international development organizations

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

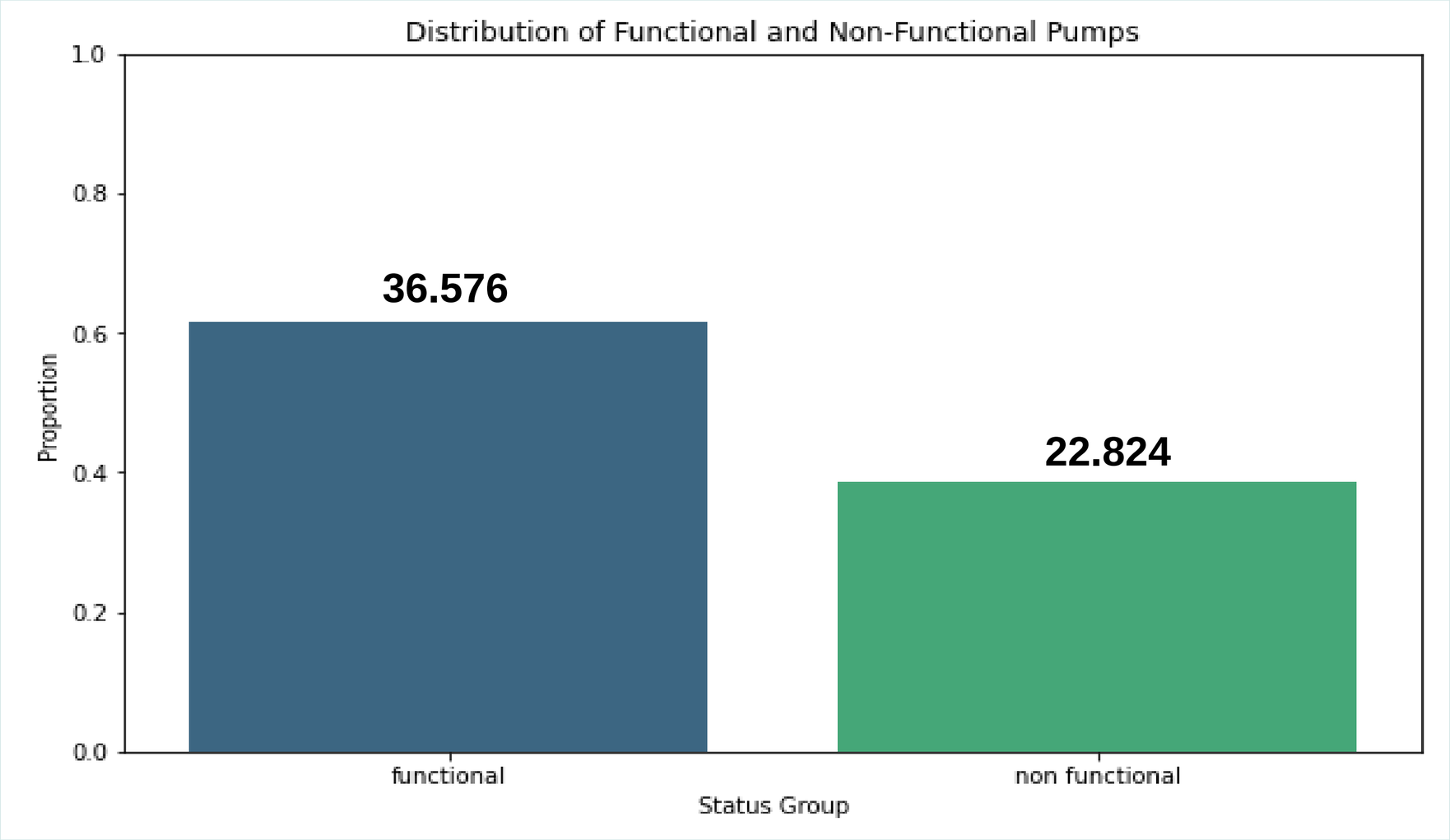
- Source: DrivenData
- Link to the source: https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/page/23/

- Training dataset: 60000 records
- Testing dataset: 15000 records

DATASET DESCRIPTION

- Certain fields: Geographic location, water access point, quantity of water

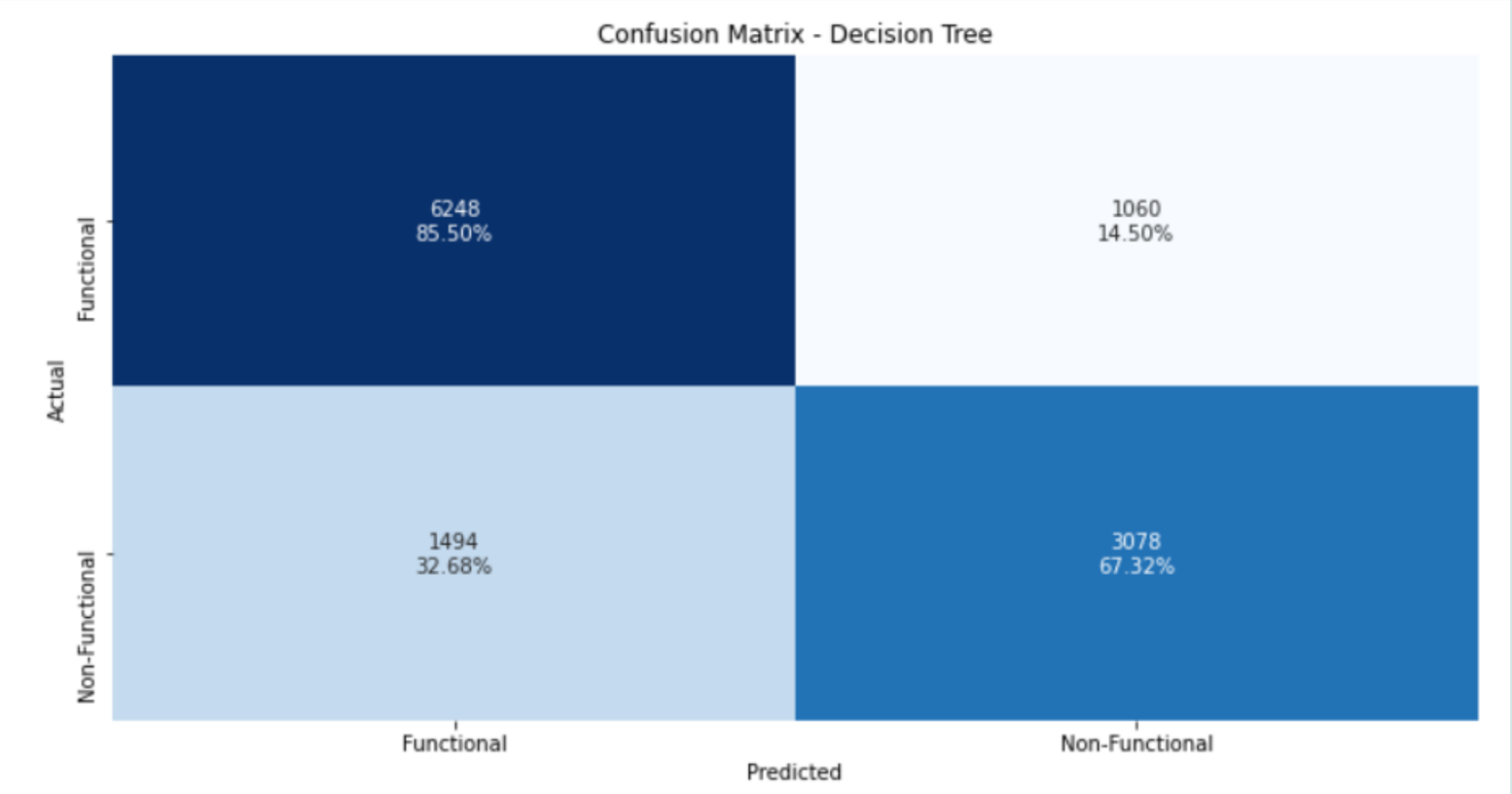
PUMPS STATUS



MODELING APPROACH

Models tested: Logistic

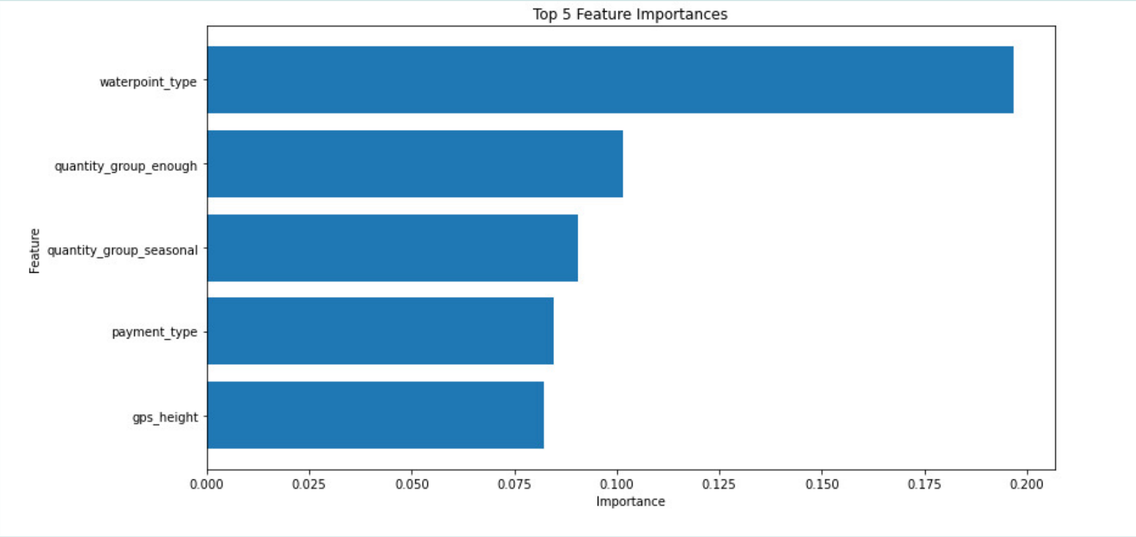
Regression Decision Tree



FEATURE IMPORTANCE

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

1. Waterpoint_type (ie. Water Access point)
2. Quantity_group (ie. Quantity of wáter)
3. Payment_type (ie. Payment method)



RECOMMENDATIONS

1. Align Payment Plans:

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

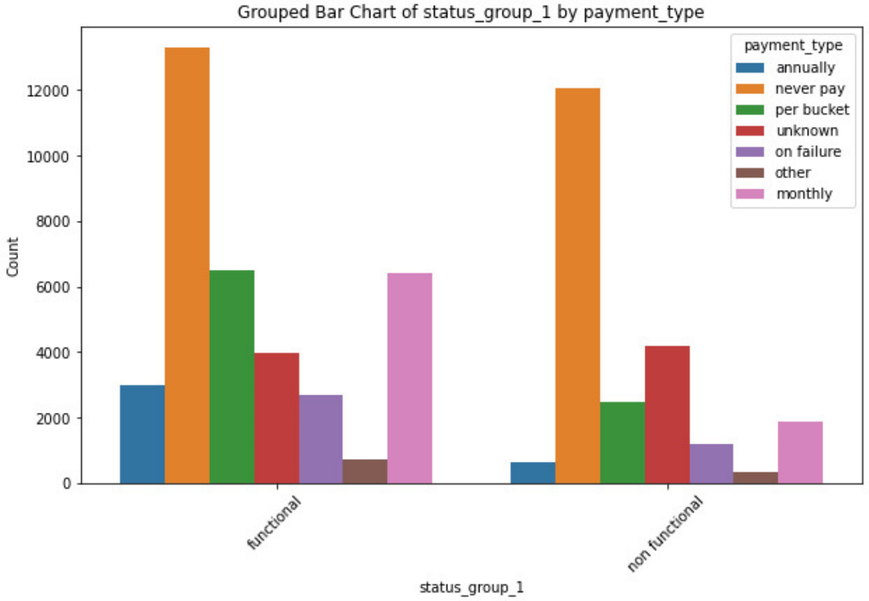


Table for payment_type:

payment_type	annually	monthly	never pay	on failure	other	per bucket	unknown
status_group_1							
functional	8.17%	17.52%	36.32%	7.40%	1.99%	17.77%	10.83%
non functional	2.87%	8.29%	52.85%	5.29%	1.42%	10.89%	18.39%

RECOMMENDATIONS

2. Use Dry Pumps as Indicators:

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

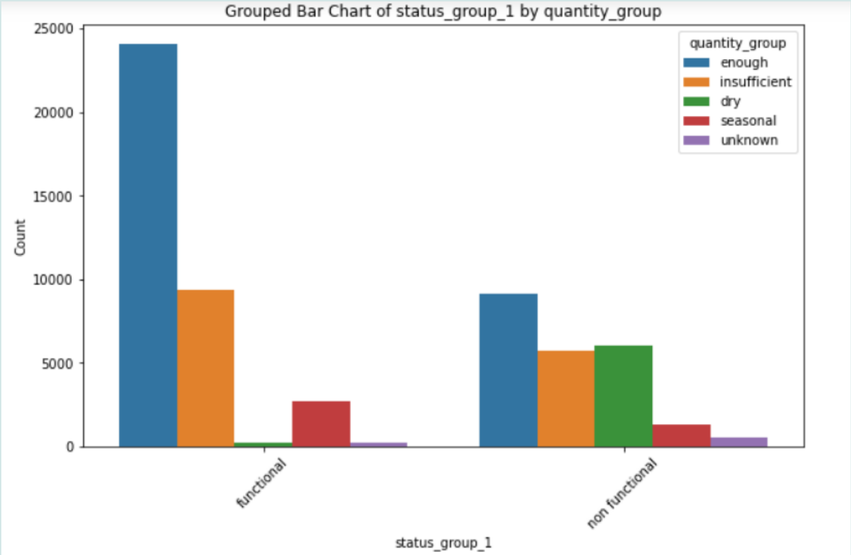


Table for quantity_group:

quantity_group	dry	enough	insufficient	seasonal	unknown
status_group_1					
functional	0.53%	65.75%	25.61%	7.49%	0.62%
non functional	26.52%	40.04%	25.25%	5.74%	2.46%

RECOMMENDATIONS

3. Identify Non-Common Types:

Invest in communal standpipe multiple as it best detects pump functionality

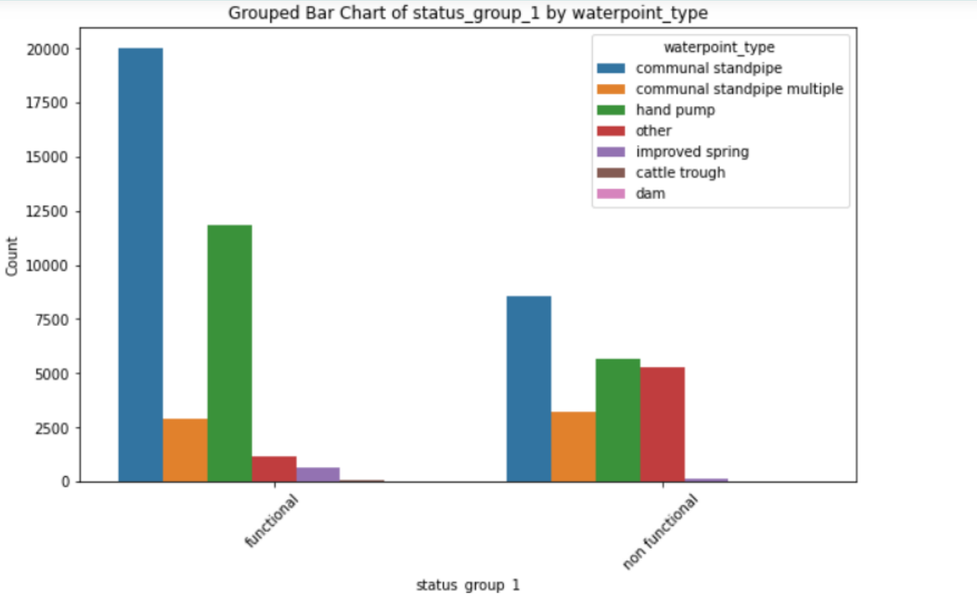


Table for waterpoint_type:

waterpoint_type	cattle trough	communal standpipe	communal standpipe multiple	dam	hand pump	improved spring	other
status_group_1							
functional	0.24%	54.64%	7.88%	0.02%	32.35%	1.77%	3.10%
non functional	0.13%	37.40%	14.11%	0.00%	24.77%	0.60%	22.99%

THANK YOU VERY MUCH

Miguel Barriola Arranz

- LinkedIn: https://www.linkedin.com/in/miguel- barriola-arranz/
- Medium: https://medium.com/@mbarriolaarranz