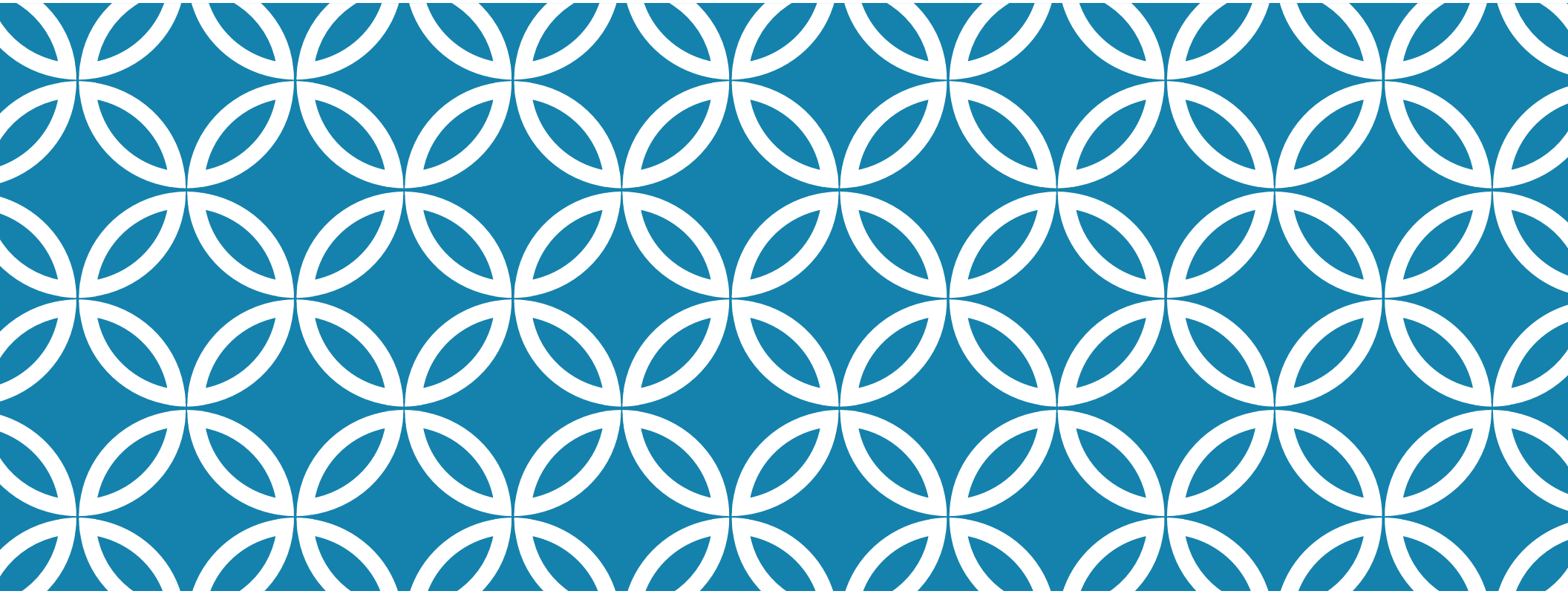


TANZANIA WATER WELLS CLASSIFICATION PROJECT PHASE 3

Presented by: Lydia Mangoa
8th March 2025

INTRODUCTION

The Problem:

- 38% of Tanzania's water wells are non-functional.
- Maintenance inefficiencies lead to wasted resources.
- Lack of predictive insights makes sustainable water access difficult.

Goal:

- Develop a machine learning model to classify wells as functional, non-functional, or in need of repair.
- Provide actionable insights to optimize well maintenance and resource allocation.

STAKEHOLDERS & BENEFICIARIES

- **Government & Policymakers:** Prioritize funding and infrastructure improvements.
- **NGOs & Aid Organizations:** Identify high-risk areas for water investment.
- **Community Water Management Committees:** Plan preventive maintenance strategies.
- **Local Engineers & Planners:** Optimize well construction and repair decisions.

DATA OVERVIEW

Data Source: DrivenData (Pump It Up: Data Mining the Water Table)

- 59,400 water points analyzed.
- 40+ features including location, management type, payment system, and water source.
- Target variable: Well status (Functional, Non-functional, Needs Repair).

METHODOLOGY

1. Data Preprocessing:

- Handled missing values, categorical encoding, and feature scaling.

2. Feature Engineering:

- Created new variables like well age.
- Analyzed geographic distribution of non-functional wells.

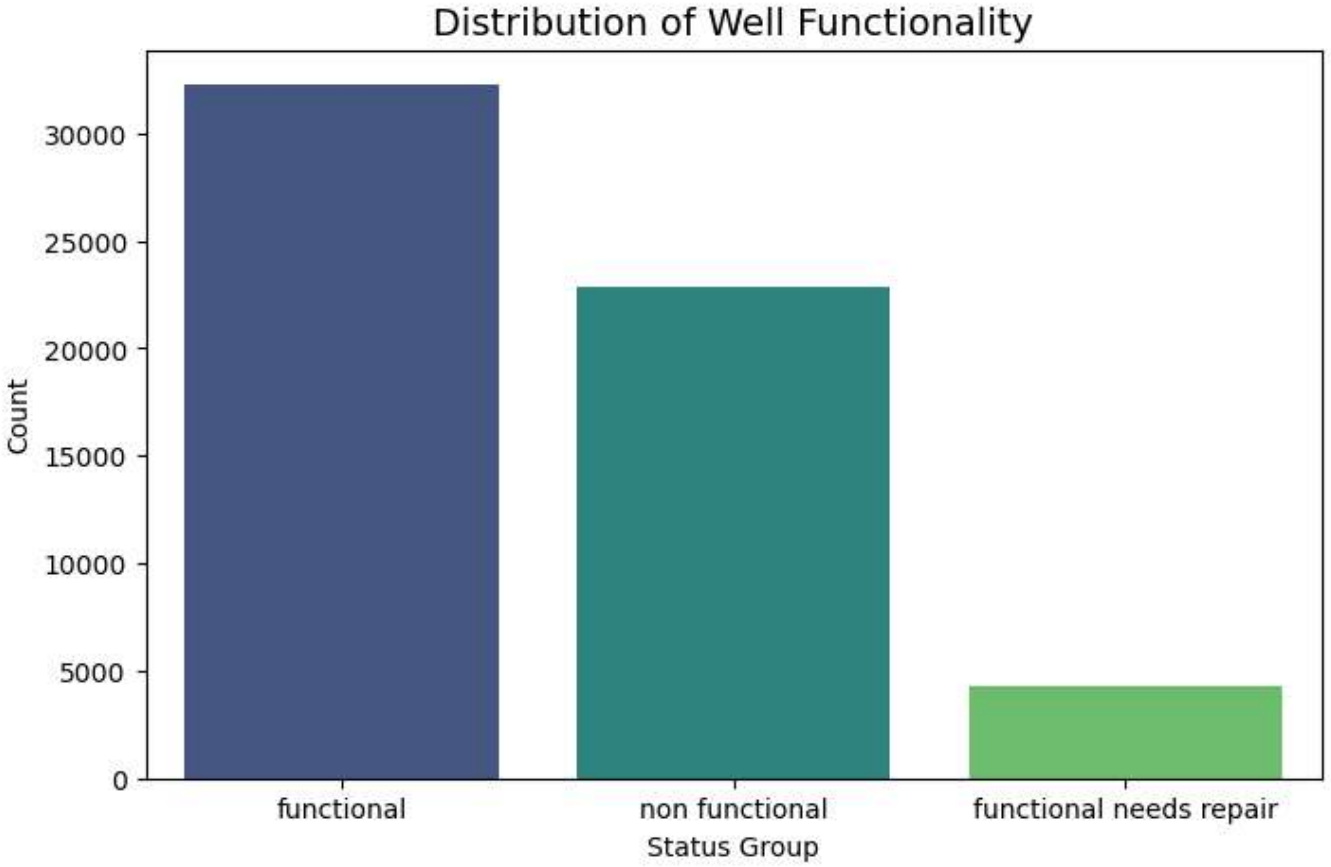
3. Model Selection:

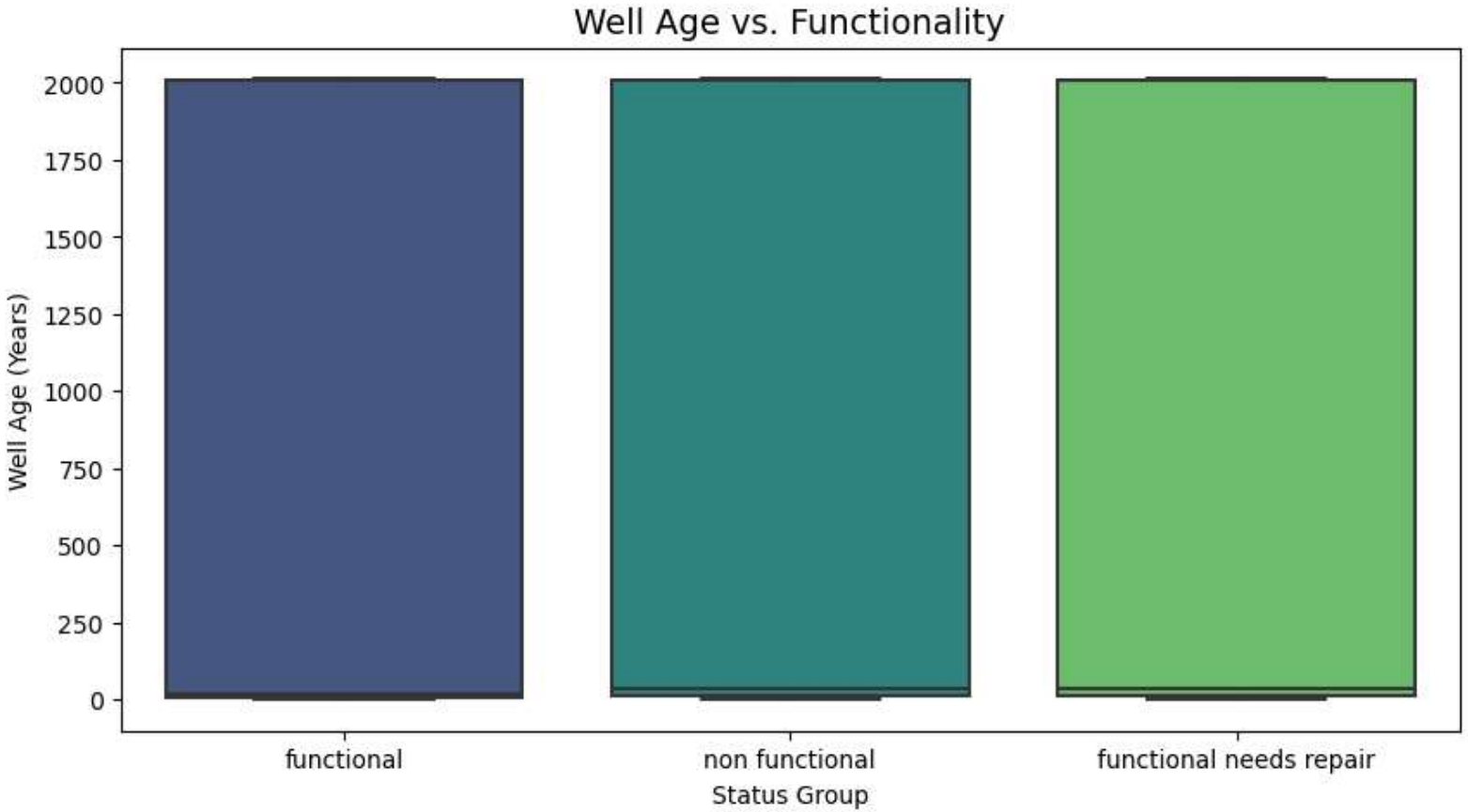
- Tested multiple classification models.
- Evaluated based on accuracy, precision, and recall.

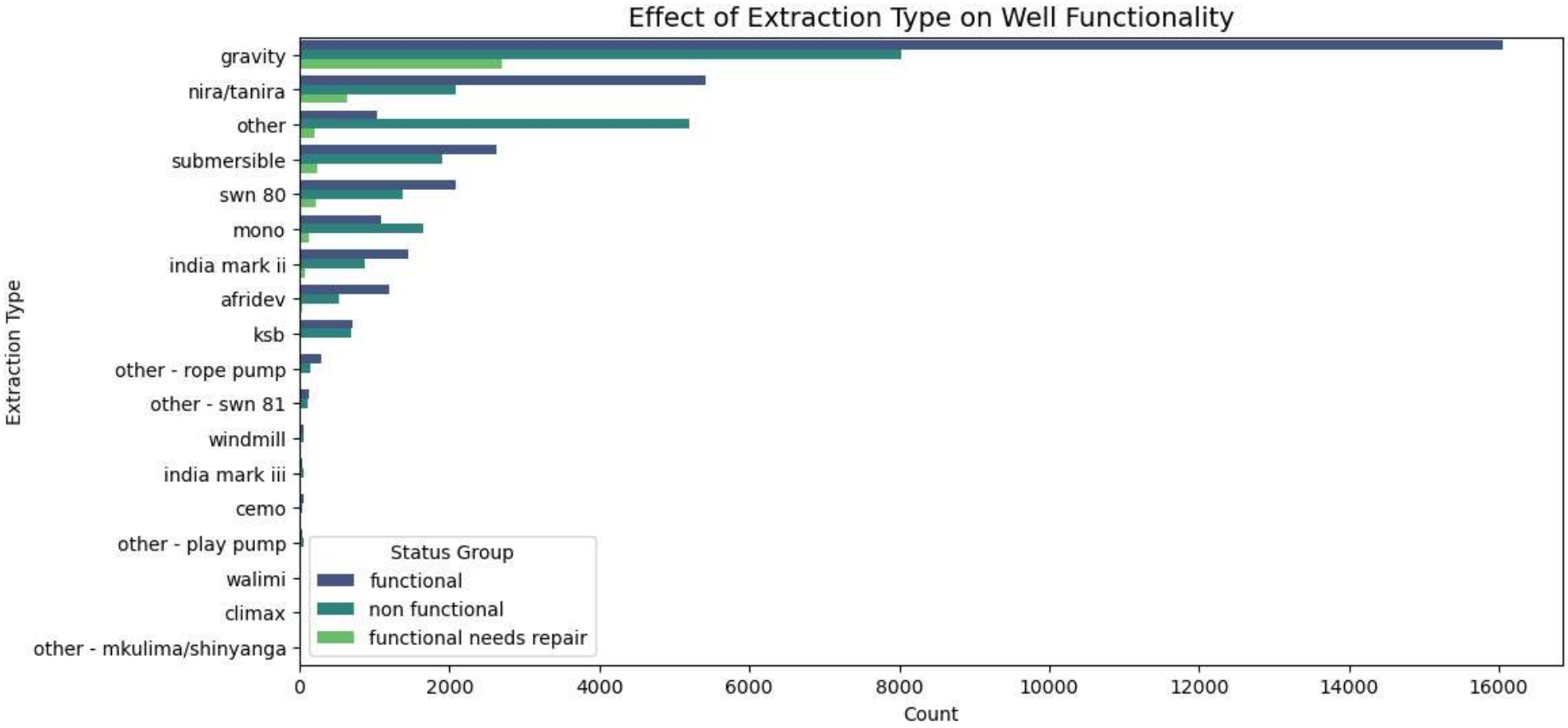
KEY INSIGHTS FROM DATA ANALYSIS

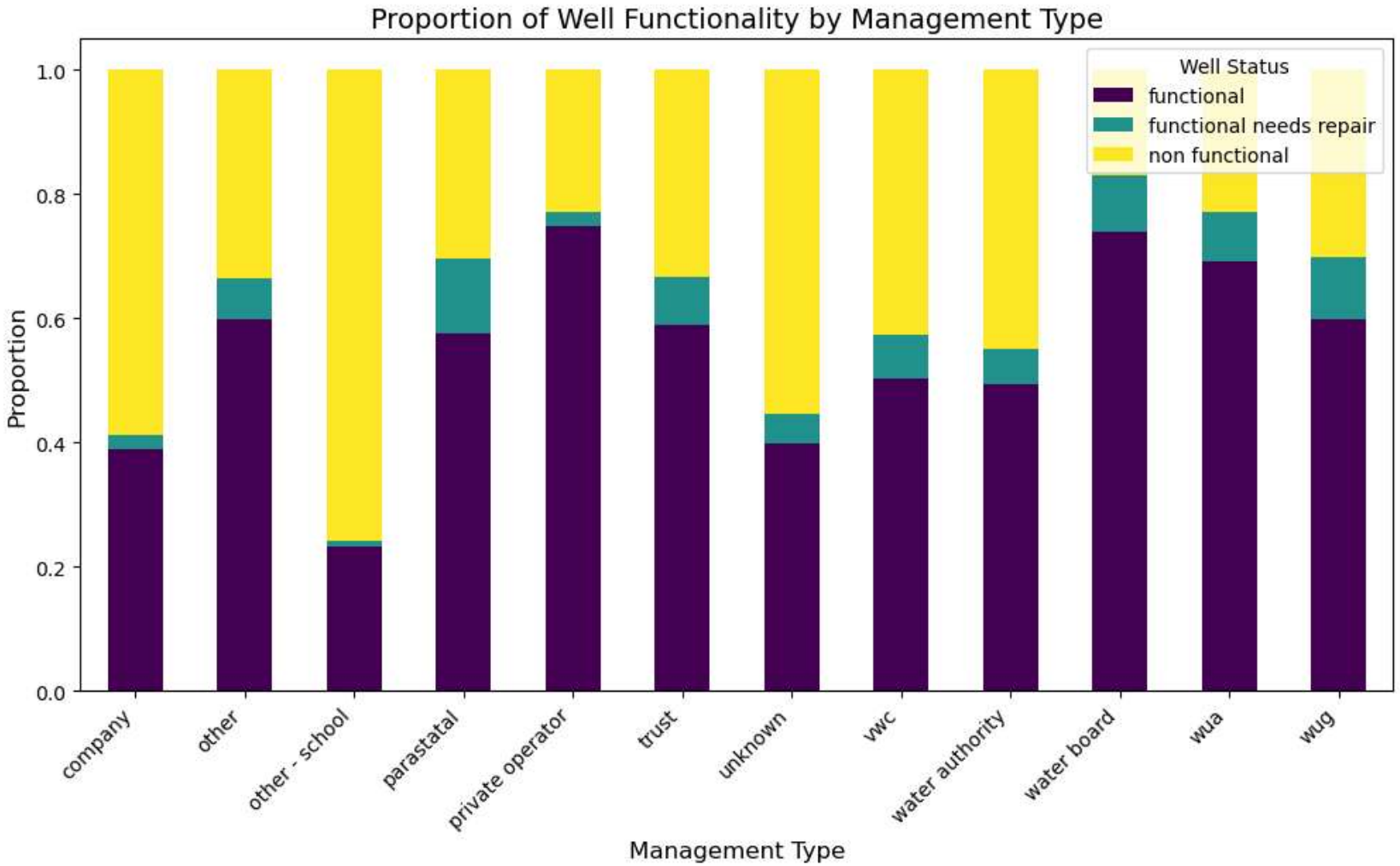
Older wells (>20 years) are more likely to fail.

- Certain management types maintain more functional wells.
- Wells in specific regions show higher failure rates.
- Payment systems influence functionality (prepaid wells perform better).


Distribution of Well Functionality
Count
30000
25000
20000
15000
10000
5000
0
functional
non functional
functional needs repair
Status Group


Well Age vs. Functionality
Well Age (Years)
2000
1750
1500
1250
1000
750
500
250
0
functional
non functional
functional needs repair
Status Group


Effect of Extraction Type on Well Functionality
Extraction Type
gravity
nira/tanira
other
submersible
swn 80
mono
india mark ii
afridev
ksb
other - rope pump
other - swn 81
windmill
india mark iii
cemo
other - play pump
walimi
climax
other - mkulima/shinyanga
Status Group
functional
non functional
functional needs repair
0
2000
4000
6000
8000
10000
12000
14000
16000
Count


Proportion of Well Functionality by Management Type
Well Status
functional
functional needs repair
non functional
Proportion
1.0
0.8
0.6
0.4
0.2
0.0
company
other
other - school
parastatal
private operator
trust
unknown
vwc
water authority
water board
wua
wug
Management Type

MODEL PERFORMANCE COMPARISON

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.72	0.74 / 0.67	0.84 / 0.53	0.79 / 0.60
Decision Tree	0.75	0.80 / 0.67	0.80 / 0.67	0.80 / 0.67
K-Nearest Neighbors (KNN)	0.76	0.79 / 0.71	0.84 / 0.64	0.81 / 0.68
Support Vector Machine (SVM)	0.76	0.76 / 0.76	0.89 / 0.56	0.82 / 0.65
Gradient Boosting	0.72	0.72 / 0.70	0.88 / 0.46	0.79 / 0.56
Naïve Bayes	0.61	0.79 / 0.49	0.50 / 0.79	0.61 / 0.61
Tuned Random Forest	**0.80**	**0.81 / *0.77***	**0.88 / *0.67***	**0.84 / *0.72***

Tuned Random Forest - The Best Performing Model

- Highest accuracy: 80%
- Best balance of precision & recall
- Strong generalization on test data
- Can be fine-tuned for further improvements

Why Random Forest?

- Handles class imbalance well.
- Robust against overfitting.
- Provides feature importance insights for better decision-making.

BUSINESS IMPACT & RECOMMENDATIONS

Actionable Insights for Stakeholders:

- **Preventive Maintenance:** Focus on older wells & high-risk areas.
- **Policy Adjustments:** Encourage prepaid water systems.
- **Targeted Funding:** Prioritize NGOs' & government investments in high-failure zones.

Future Enhancements:

- Deploy predictive dashboards for real-time monitoring.
- Incorporate additional data sources (weather, soil conditions).
- Explore deep learning for improved accuracy.

CONCLUSION & NEXT STEPS

- Machine learning enhances decision-making for sustainable water access.
- Predictive analytics enables smarter well maintenance.
- Implementation of insights can reduce well failures and improve water access across Tanzania.

Next Steps:

- Deploy the best model in a pilot program.
- Monitor real-world impact & refine recommendations.
- Scale solutions across broader regions.

THANK YOU.

ANY QUESTIONS?

Name: Lydia Mangoa

LinkedIn: http://www.linkedin.com/in/lydia-mangoa-2b5b68a8