TANZANIA WATER WELL STATUS AND DISTRIBUTION: A PREDICTIVE ANALYSIS

Overview

Project Overview

- Location: Tanzania, East Africa
- Population: 62 million
- Geography: Diverse, including Serengeti, Ngorongoro Crater, Zanzibar

Water Access

- Rural Dependence: Groundwater sources (wells, boreholes, springs)
- Challenges: Uneven distribution, seasonal variations

Research Focus

- Objective: Predict water well functionality using machine learning

Business and Data Understanding

Business Understanding

Objective: Ensure reliable water supply in Tanzania

Goal: Predict water well functionality to prioritize maintenance

Impact: Improve water access for rural communities, optimize resource allocation

Data Understanding

Source: Driven Data competition "Pump it Up: Data Mining the Water Table"

Key Datasets:

X_train.csv: Features for training (e.g., pump type, installation date)

y_train.csv: Well statuses (functional, non-functional, needs repair)

Test set: Features for prediction

Data Highlights

Some example of Features:

Type of pump

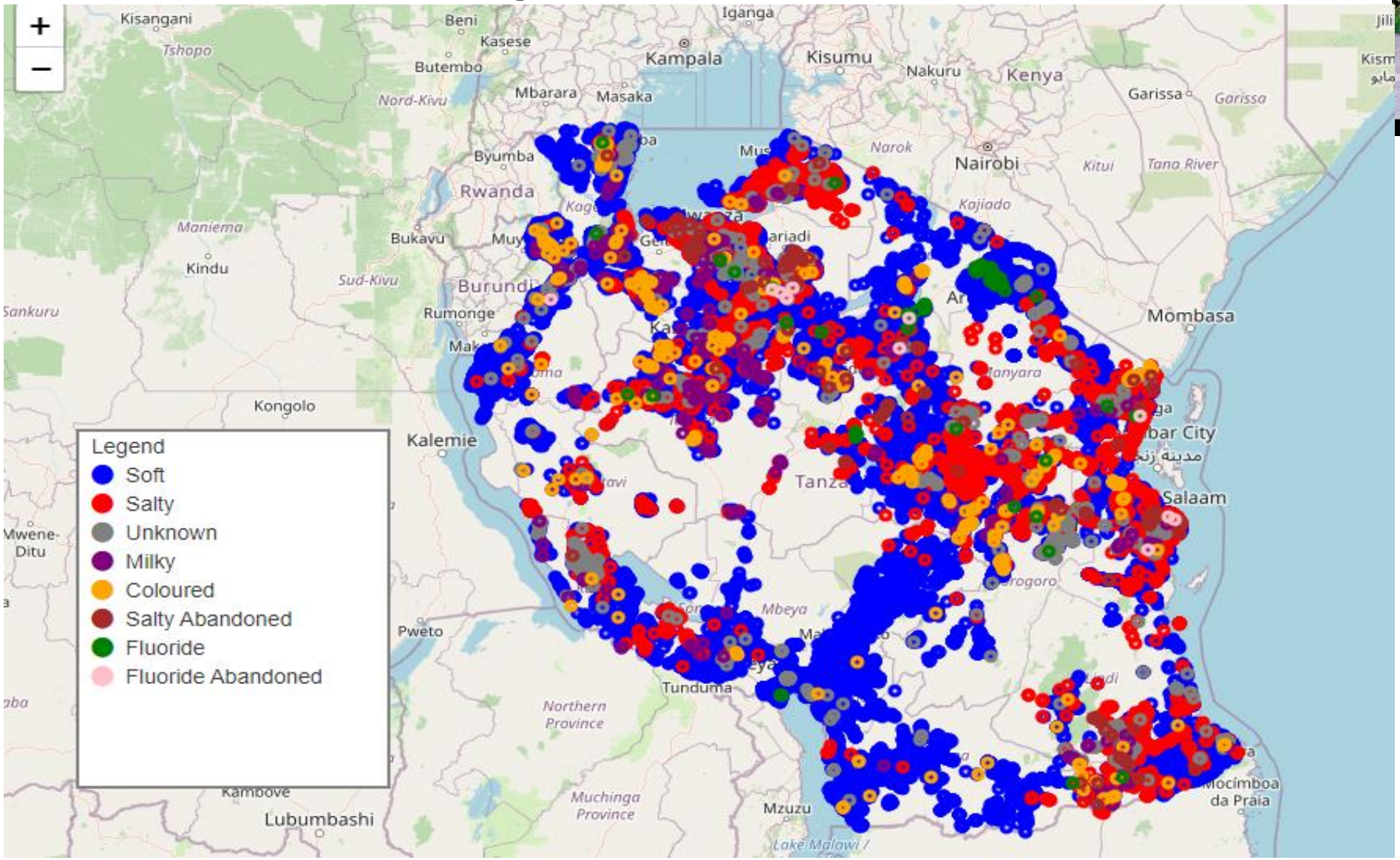
Installation date

Installer

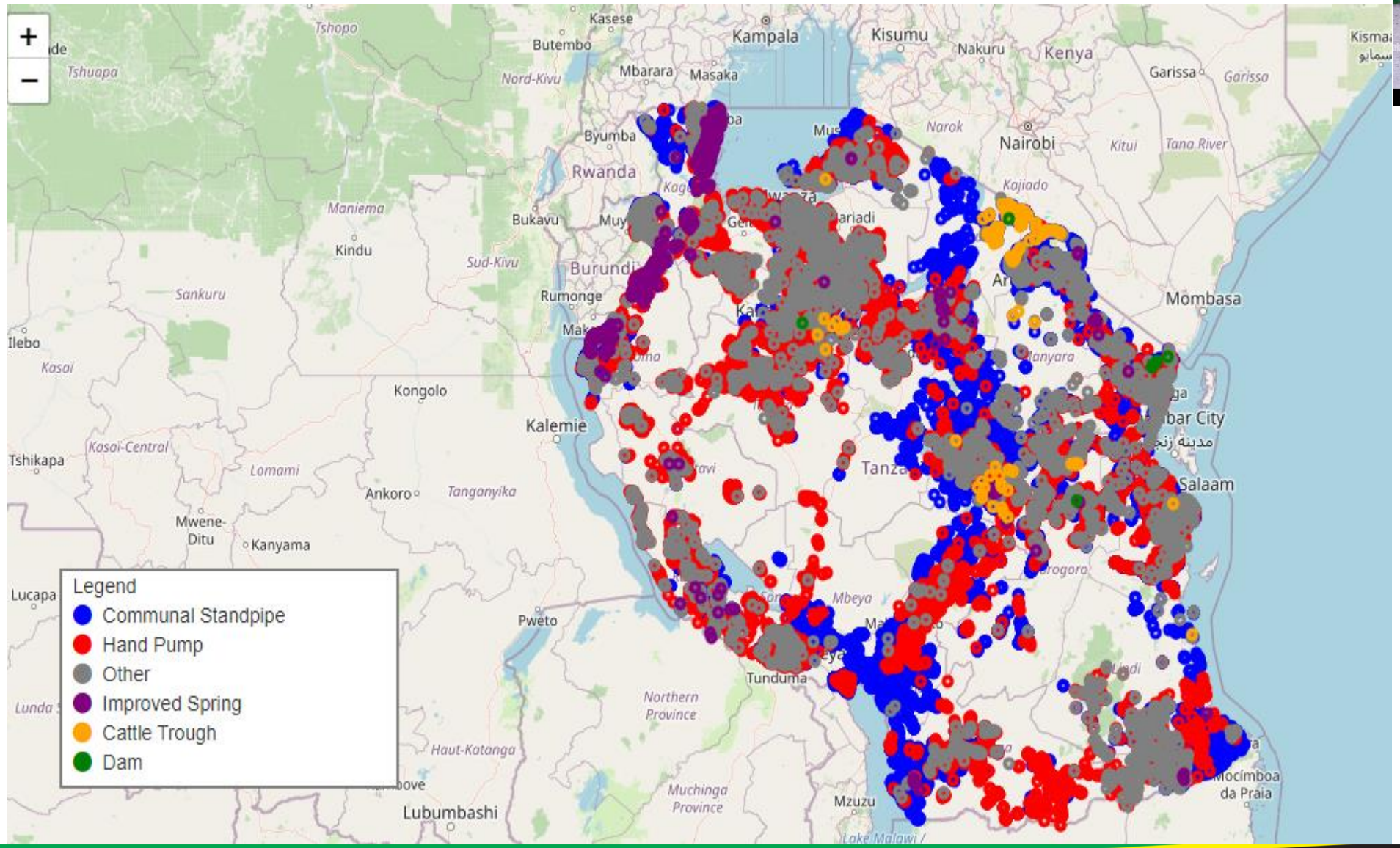
Region

Labels: Well status (functional, non-functional, needs repair)

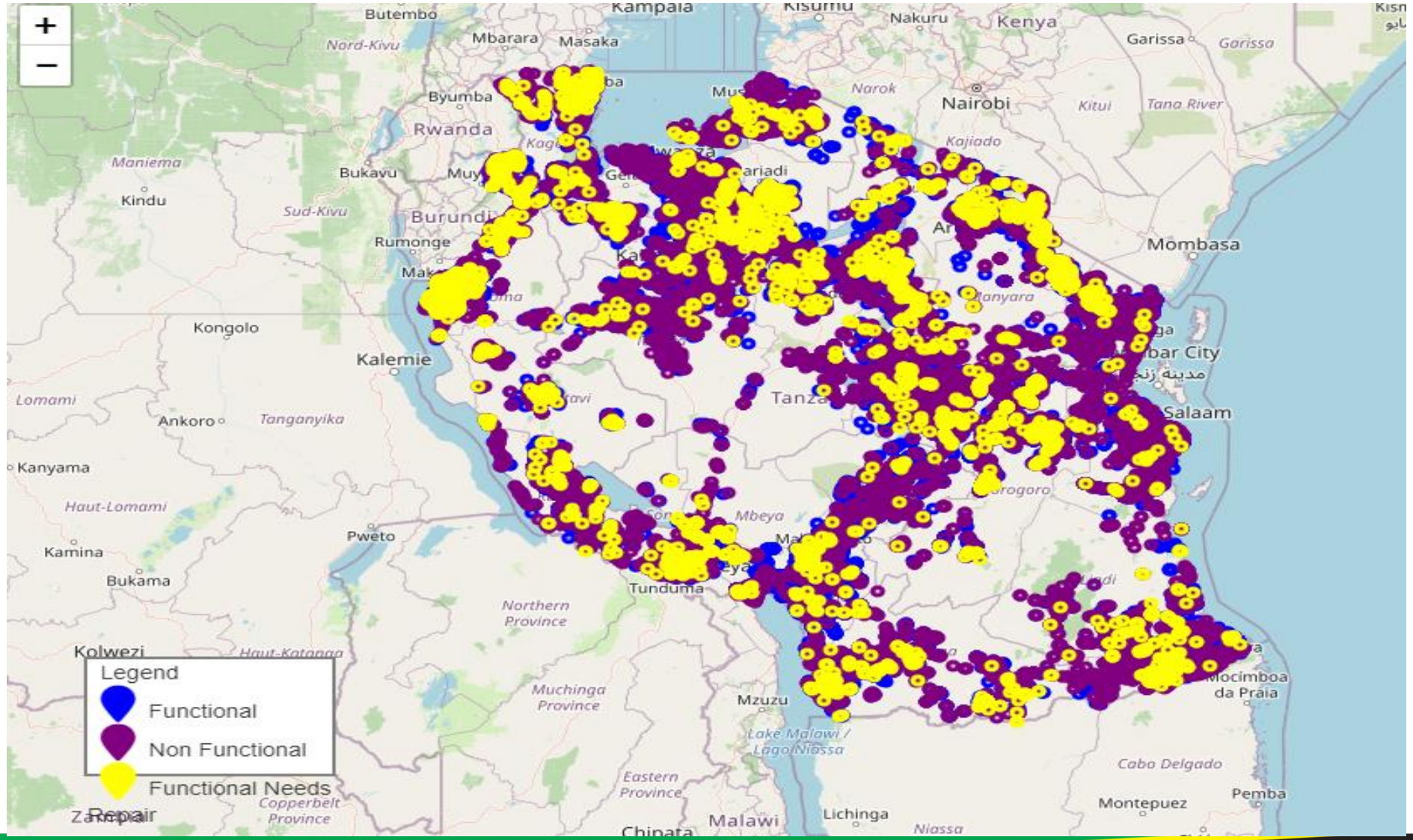
Water Quality Distribution in Tanzania



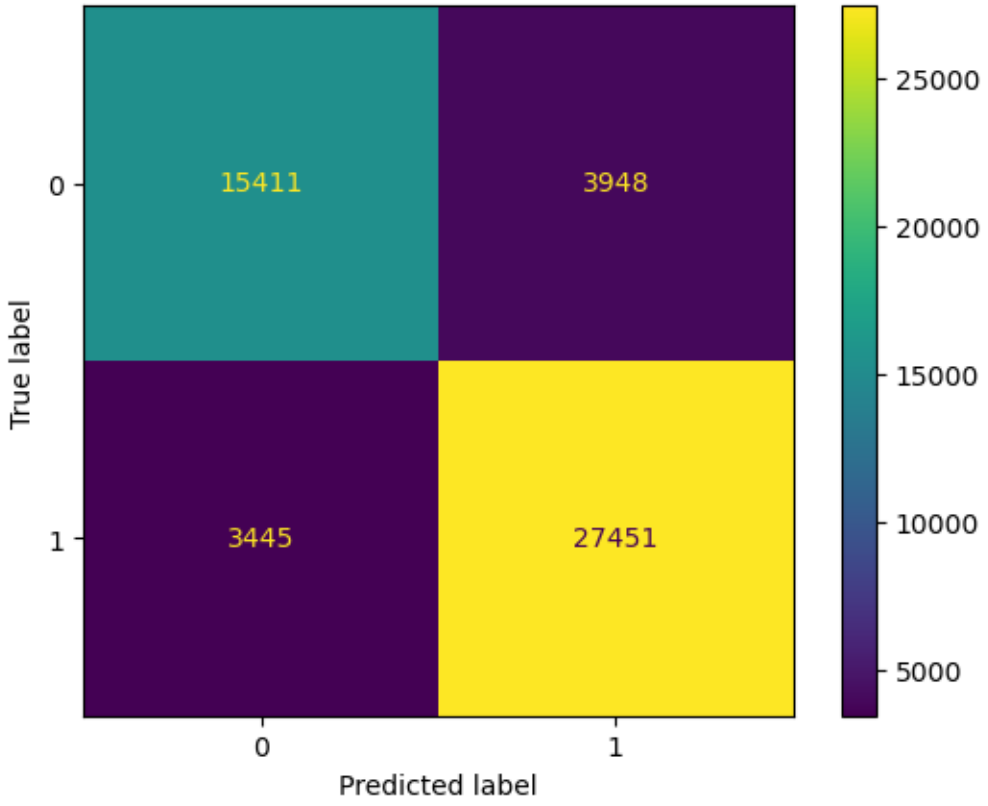
Water Point Distribution in Tanzania



Water Source Status



Best Model- Random Forest



- Mean cross-validation score: 0.85
- F1 Score (Cross-Validation): 0.85
- Recall (Cross-Validation): 0.852
- Precision (Cross-Validation): 0.8523

Recommendations

Use Random Forest Model: Predict well pump conditions

- **Accuracy:** At least 85% success rate

Northern Regions: Focus on Bukoba and Arusha

- **Issue:** High density of pumps needing repairs

Central and Southern Regions: Focus on Dodoma and Mtwara

- **Issue:** High density of non-functional pumps

Next Steps

Monitor Performance

Ensure model maintains at least 85% accuracy

Improve Data Collection:

Enhance data accuracy for better predictions

Ongoing Evaluation

Track Progress: Regularly assess the impact of interventions and adjust based on new data and findings

Thank You