

# TANZANIA WATER WELLS

Present by Calvine Dasilver

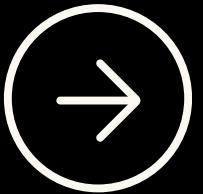
 [github.com](https://github.com/calvinedasilver96)

 [calvinedasilver96@gmail.com](mailto:calvinedasilver96@gmail.com)

 123 Anywhere Street., Any City



# TODAY'S AGENDA



1 Project Goal

2 Business Understanding

3 Data understanding

4 Data Cleaning

5 Data Visualization

6 Model Selection

7 Future Improvement

8 Conclusion



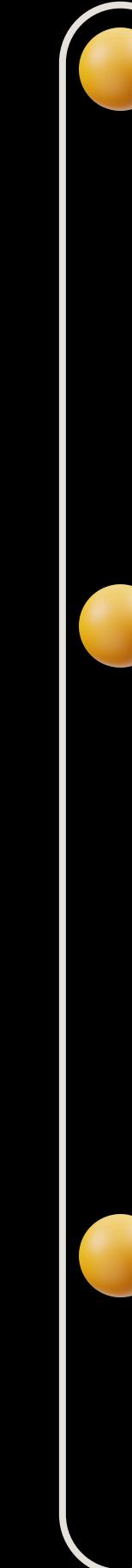
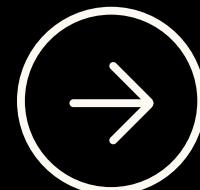
# PROJECT GOAL

We urge the Tanzanian government to adopt our improved water point maintenance model to meet citizens' water needs.





# BUSINESS UNDERSTANDING



## Tanzania's Water Crisis

Nearly 24 million people in Tanzania struggle to find clean water.

## Unlocking Water Security

Water wells have emerged as a crucial solution to water scarcity, providing a reliable source for many communities.

## Predictive Maintenance

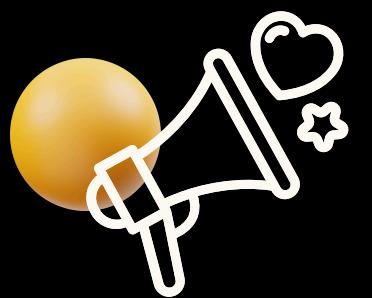
We need to develop a model to identify areas for improvement in well-maintenance operations.

# DATA UNDERSTANDING



## Data Source

Taarifa and Tanzanian Ministry of Water provided original data



## Data-Rich Analysis

Data has 59,400 data points and 41 column



## Building a Well Functionality Model

- will follow the OSEMN framework
- Obtain
- Scrub
- Explore
- Model
- Interpret

# DATA CLEANING

## STREAMLINING THE DATA

We did Dimensionality Reduction. we dropped features that we won't use to model.

## MISSING VALUES

we checked for missing values and we did imputation to solve them.

## OUTLIERS

There were outliers in the data. We capped outliers to the upper bound of the interquartile range (IQR).

## CHECKING CORRELATION

lastly, we correlated using a heat mat to check if features were still correlating to each other, 75% and above.

# DATA VISUALIZATION

From our observation:

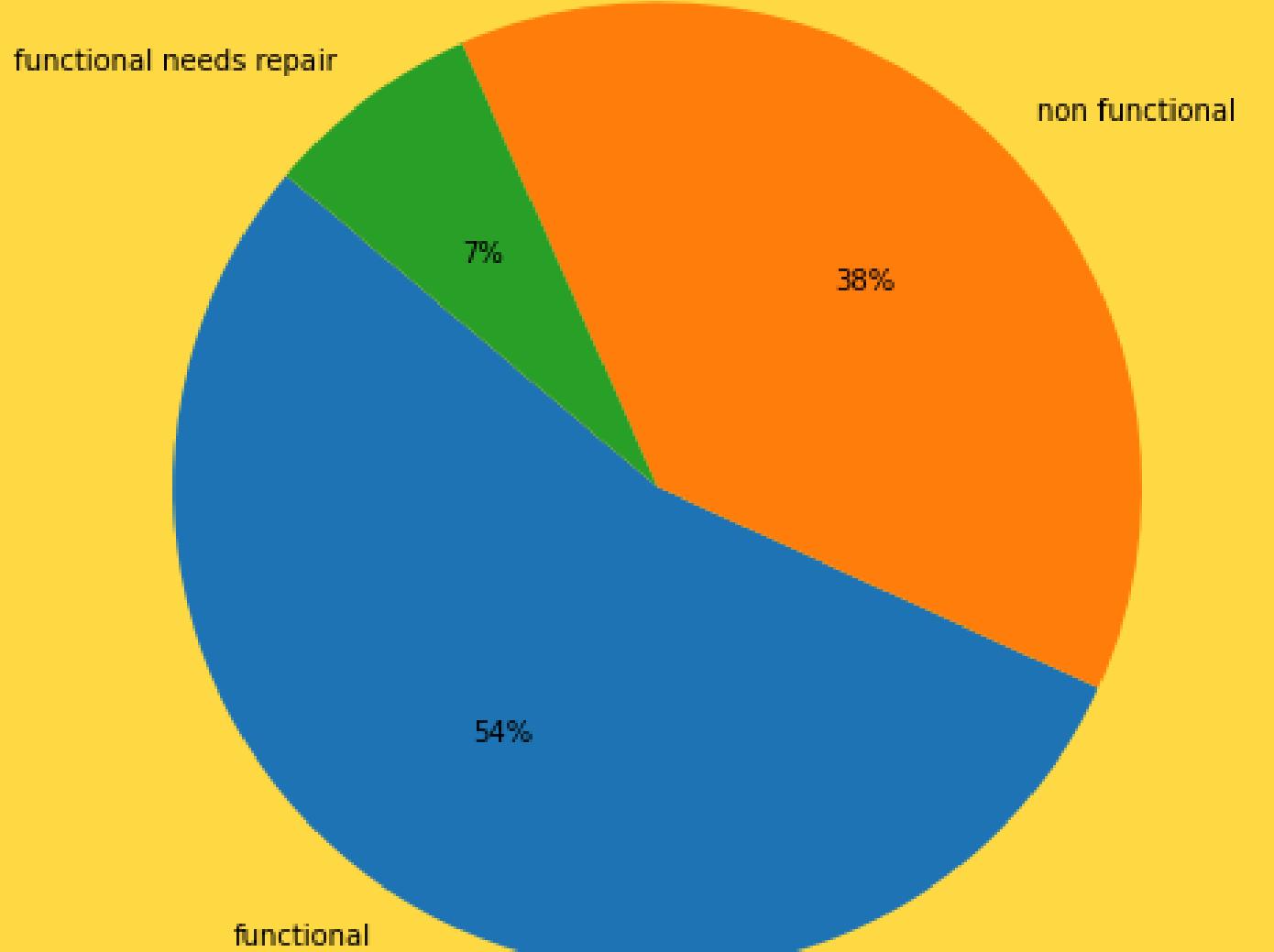
A. 54% of water points are functional,

B. 38% are non-functional

C. 7% are functional and need repair

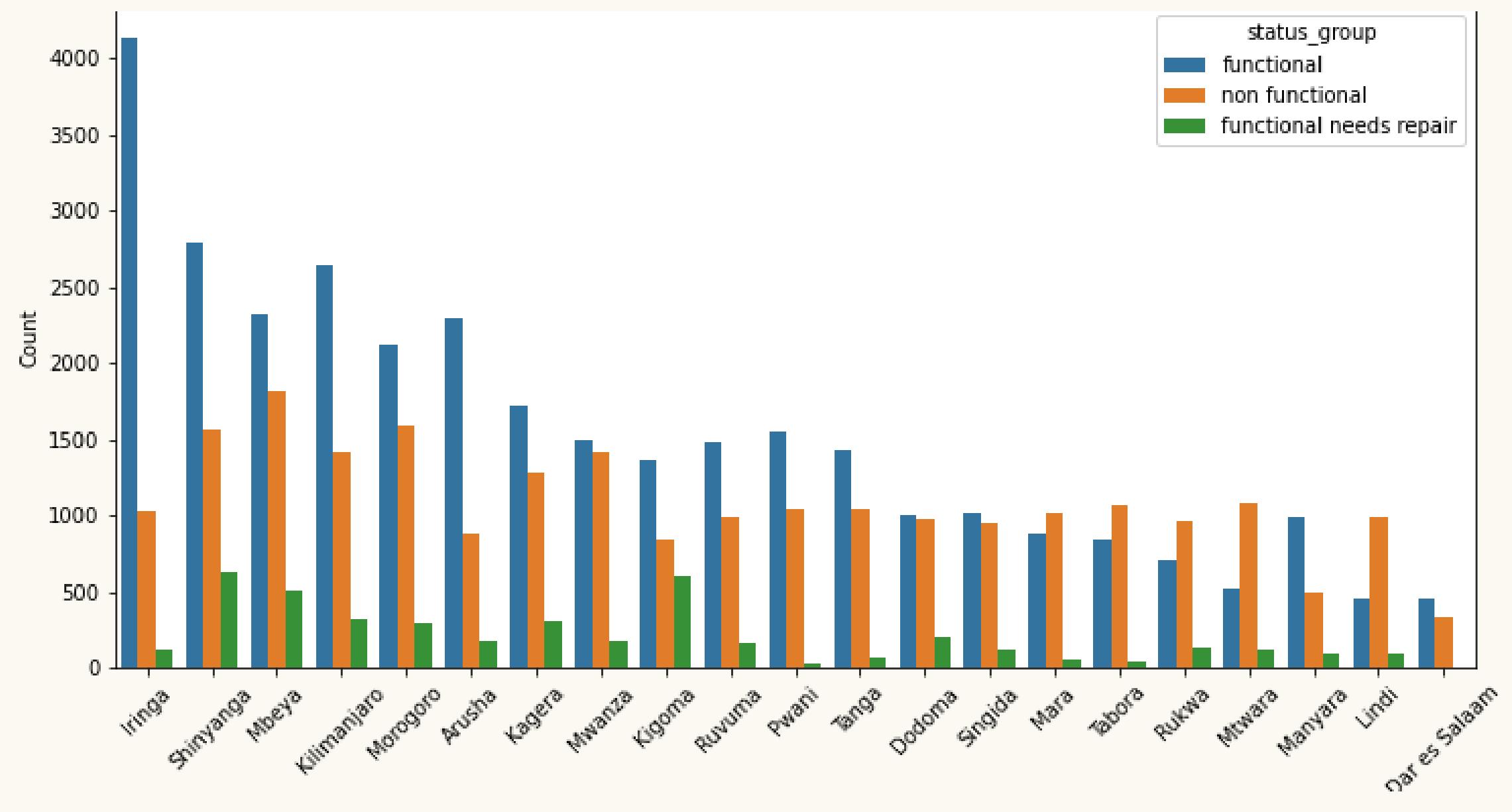
Target variable (status\_group)

Water Point Functionality Distribution



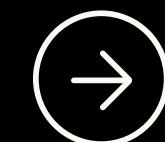
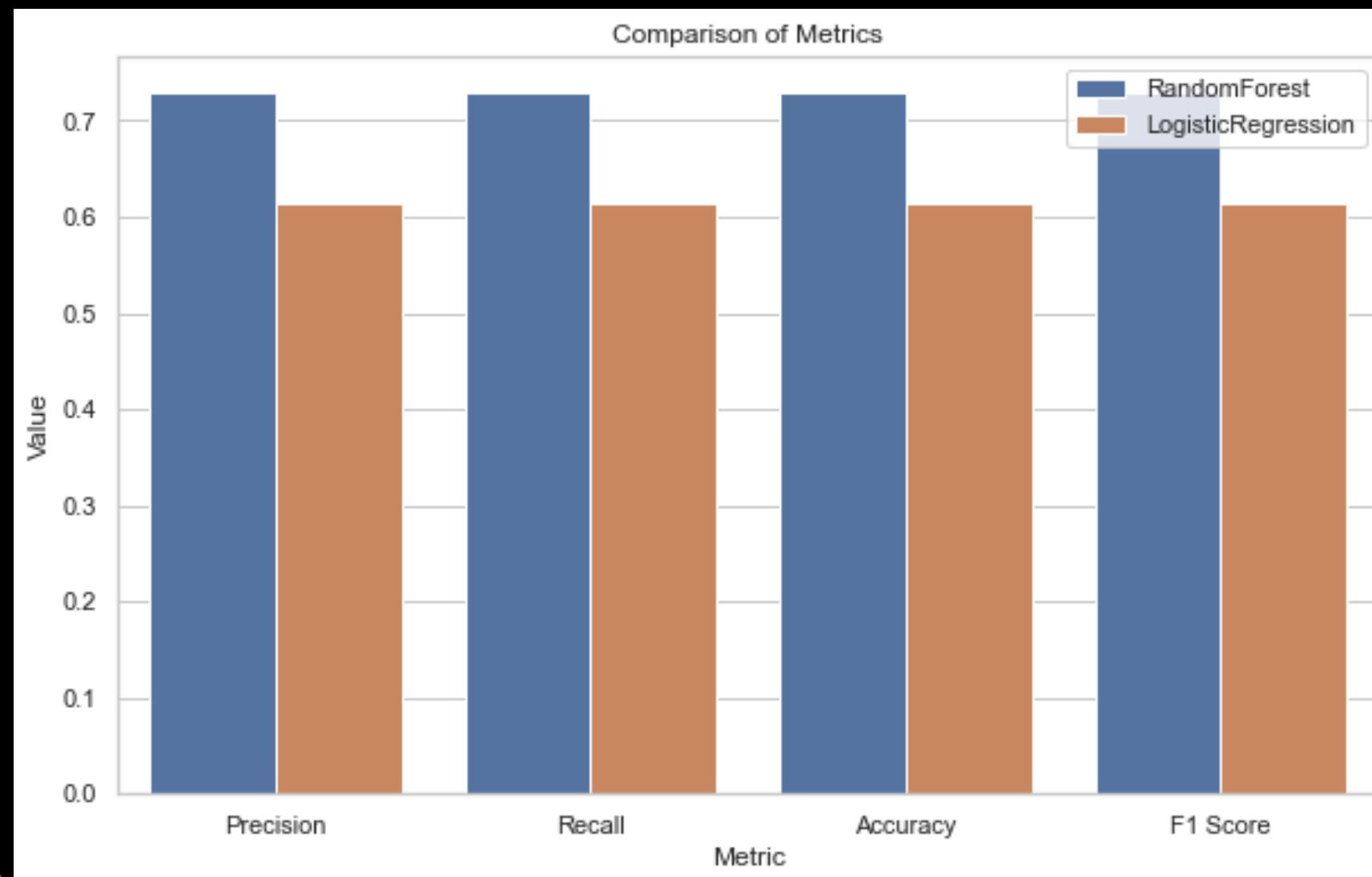
## REGION VS FUNCTIONALITY

## OBSERVATION



Analysis shows regional pump functionality varies. Iringa, Shinyanga, and Kilimanjaro have the most functional pumps, while Kilimanjaro and Morogoro have the most non-functional. Kigoma has the most functional pumps needing repair, suggesting targeted maintenance efforts there.

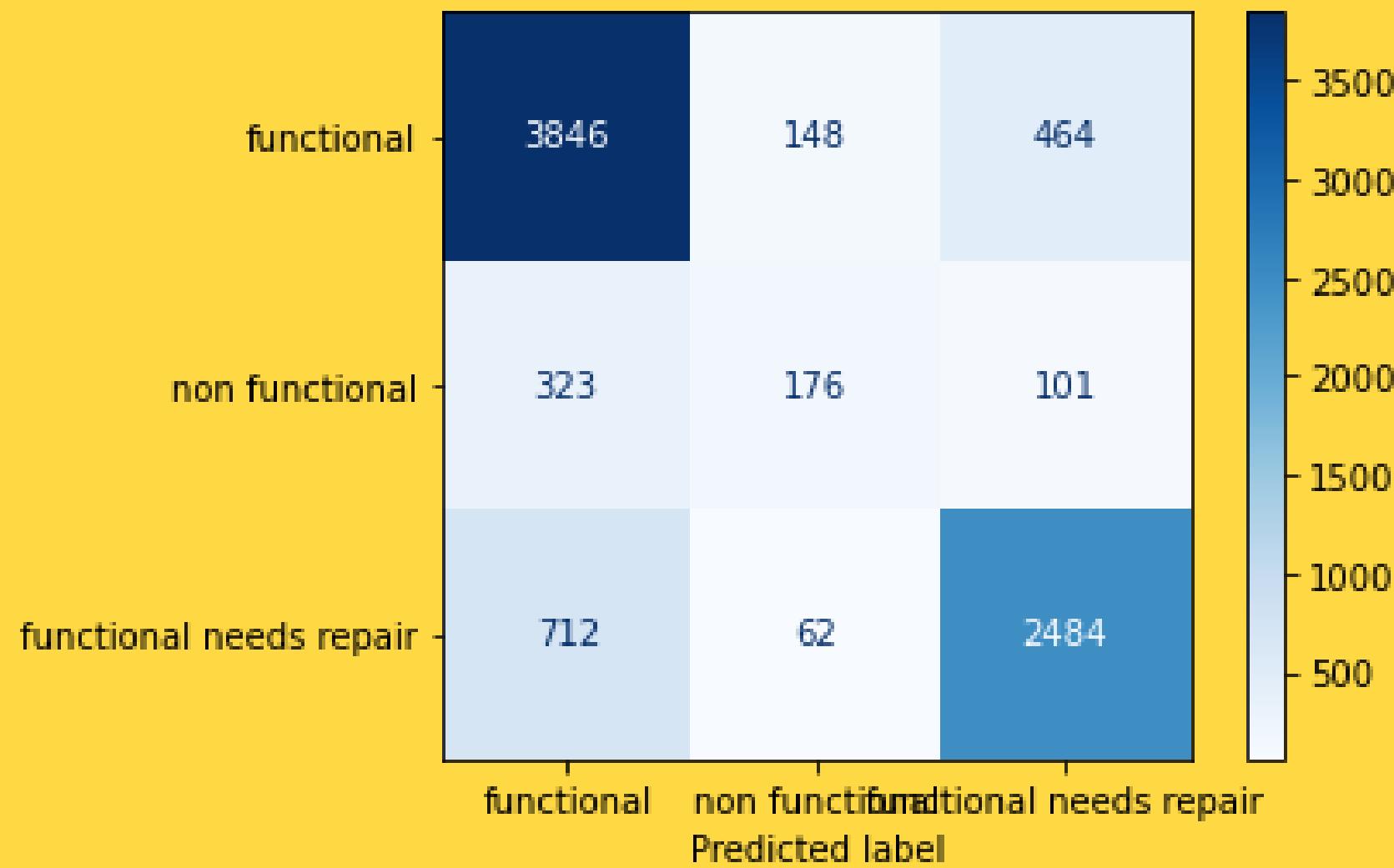
# MODEL SELECTION



Developed 4 different classifiers based on the metrics.  
Out of the 4 RF was our best model while LR had the lowest Accuracy.  
The Classifiers are: RandomForest LogisticRegression KNN XGboost

## RANDOMFORESTCLASSIFIER MODEL

True label



model report:

Validation accuracy: 78%

Streamline Maintenance and Repairs

# FUTURE IMPROVEMENT

## IMPROVE DATA

Enrich the model with coded qualitative data

## MONITOR WELLS

Continuously improve the model to predict maintenance needs for well pumps

## GEOGRAPHIC REGION

The model has to consider regional factors like rainfall, climate, etc.





Ingoude Company

# THANK YOU

for your time and attention

Present by Calvine Dasilver

