TANZANIAN WATER PUMP PROJECT

Presented By Carol Kilonzo

Write date 11/2/2023

AGENDA

Add Company Name

INTRODUCTION

Water is the backbone of agriculture and human life sustenance.But most African countries face the challeng of making this important need accessible to everyone.Tanzania is no different.With a population of more than 64 milion people 4 million lack access to water.

To make water accesible to we need to creat a model that will most accurately predict water pumps functionality.

Back to Agenda

BUSINESS CONTEXT

Problem Statement

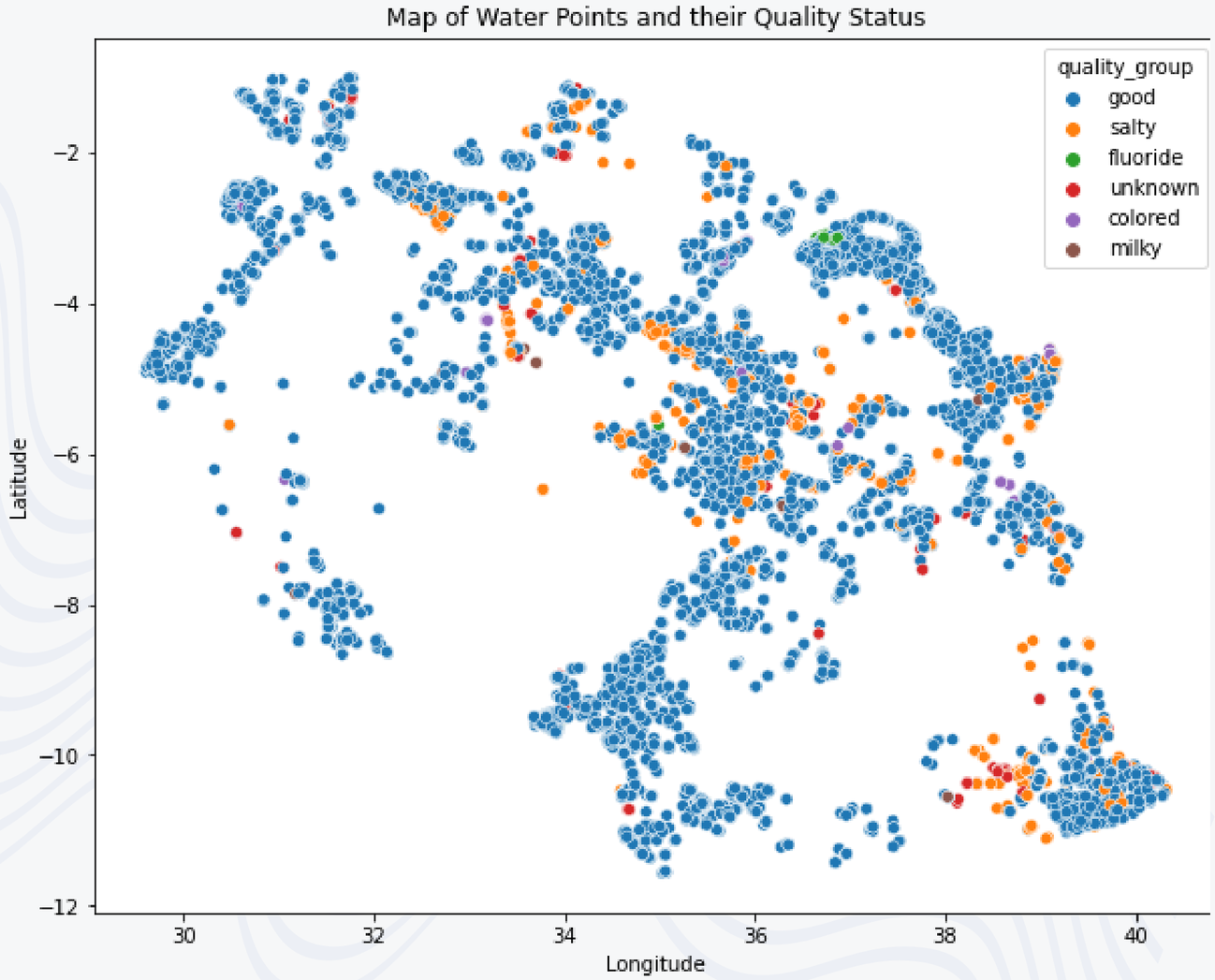
To curb the water and sanitation crisis Tanzanian Government had installed water pumps through out the country. The problem however, is that some these water pumps need repairs or replacing.

Main Objective

The goal is to create a model that will be used to implement an actionable plan for fixing and replacing water pumps throughout Tanzania.

EXPLORATORY DATA ANALYSIS

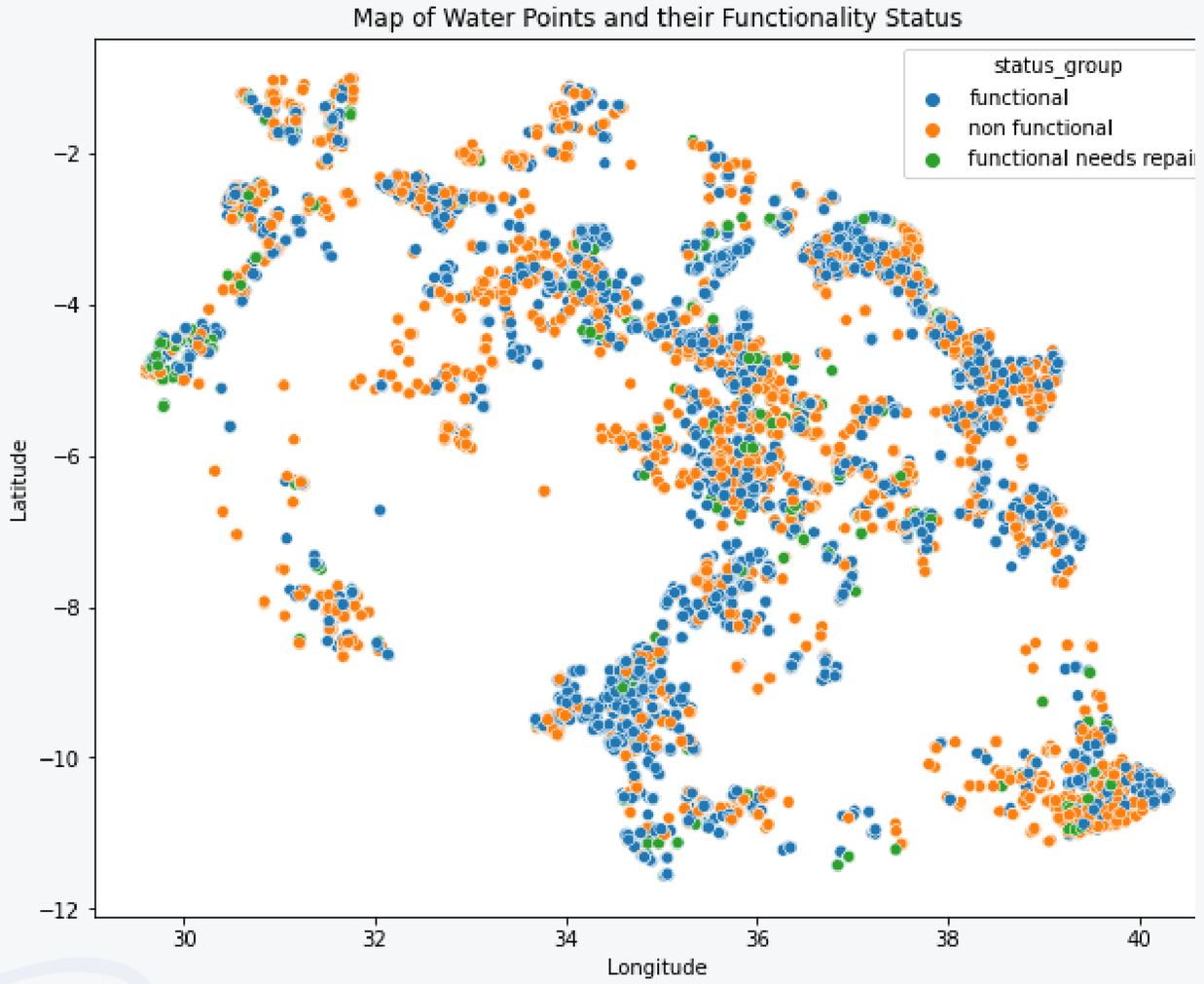
WATER QUALITY VS LOCATION



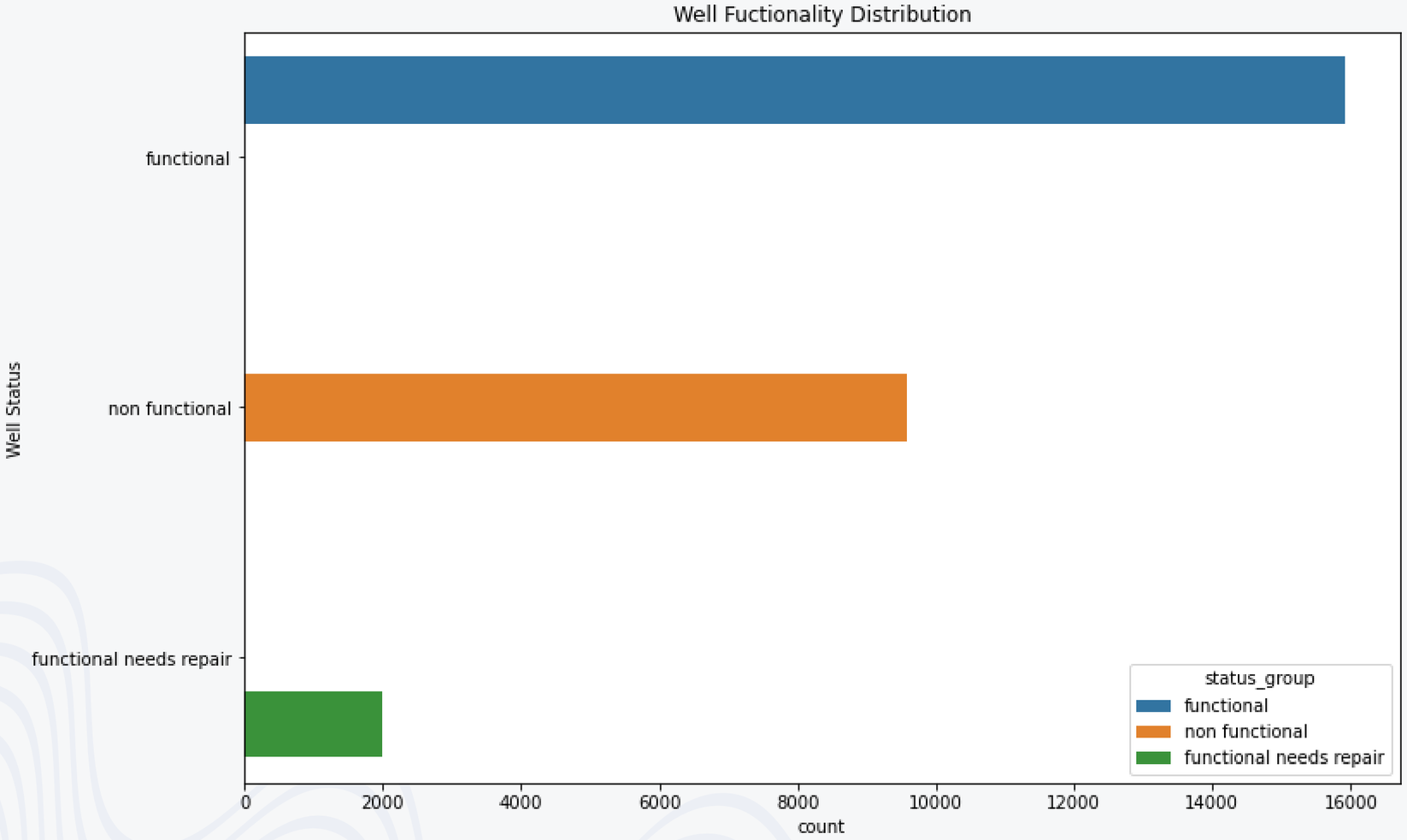
The quality of water is mostly good in most areas.

FUNCTIONALITY VS LOCATION

The number of non functional well pumps seems to be high in most locations.

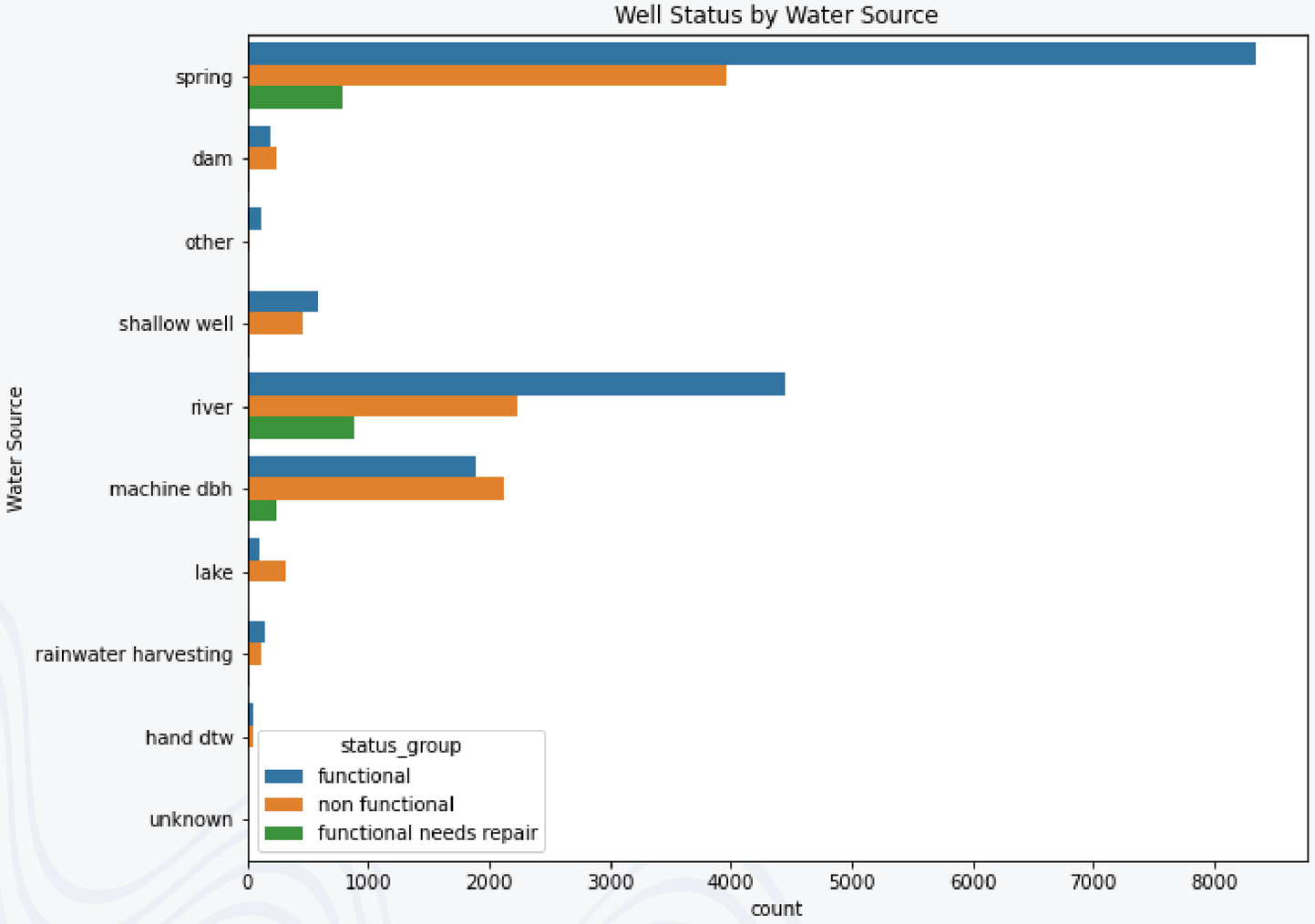


FUNCTIONALITY DISTRIBUTION



A huge number of well pumps are functional while over 8,000 need replacing.

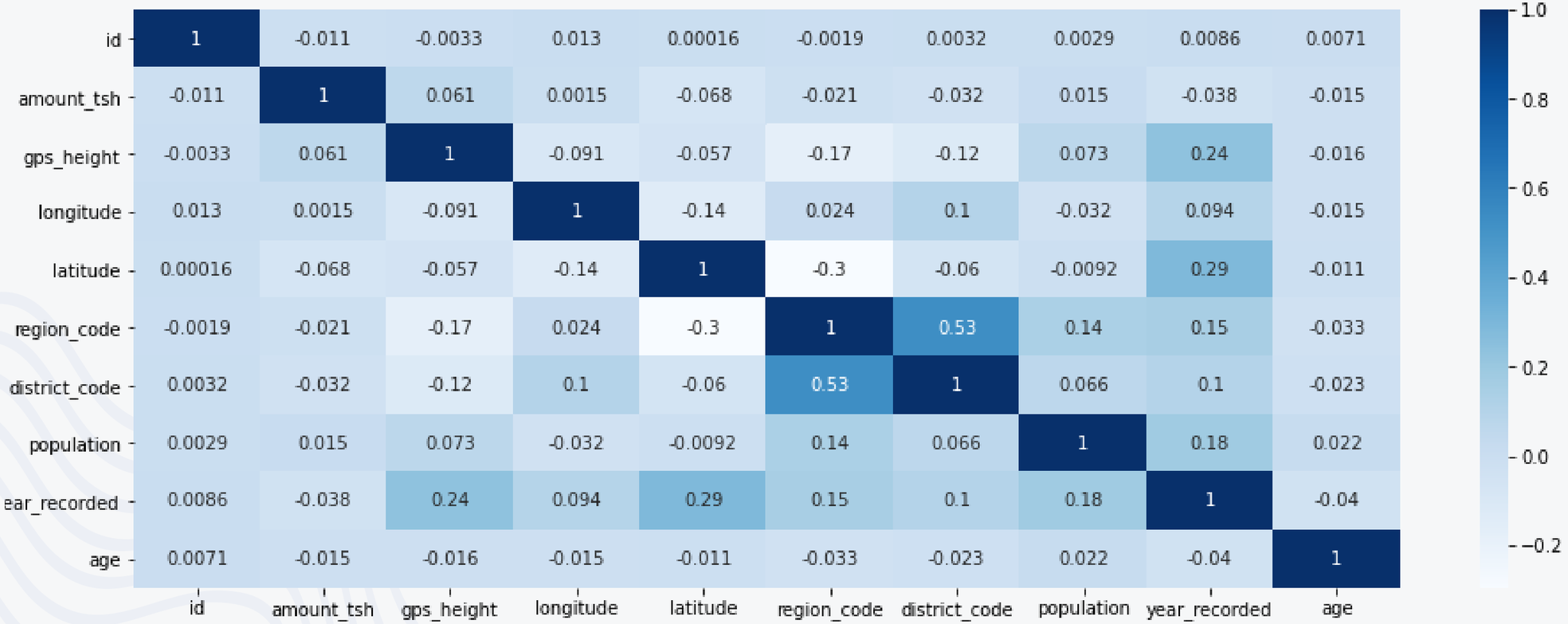
SOURCE VS FUNCTIONALITY



Spring has the highest number of functional pumps and those needing replacement while river has the highest number of pumps needing repairs.

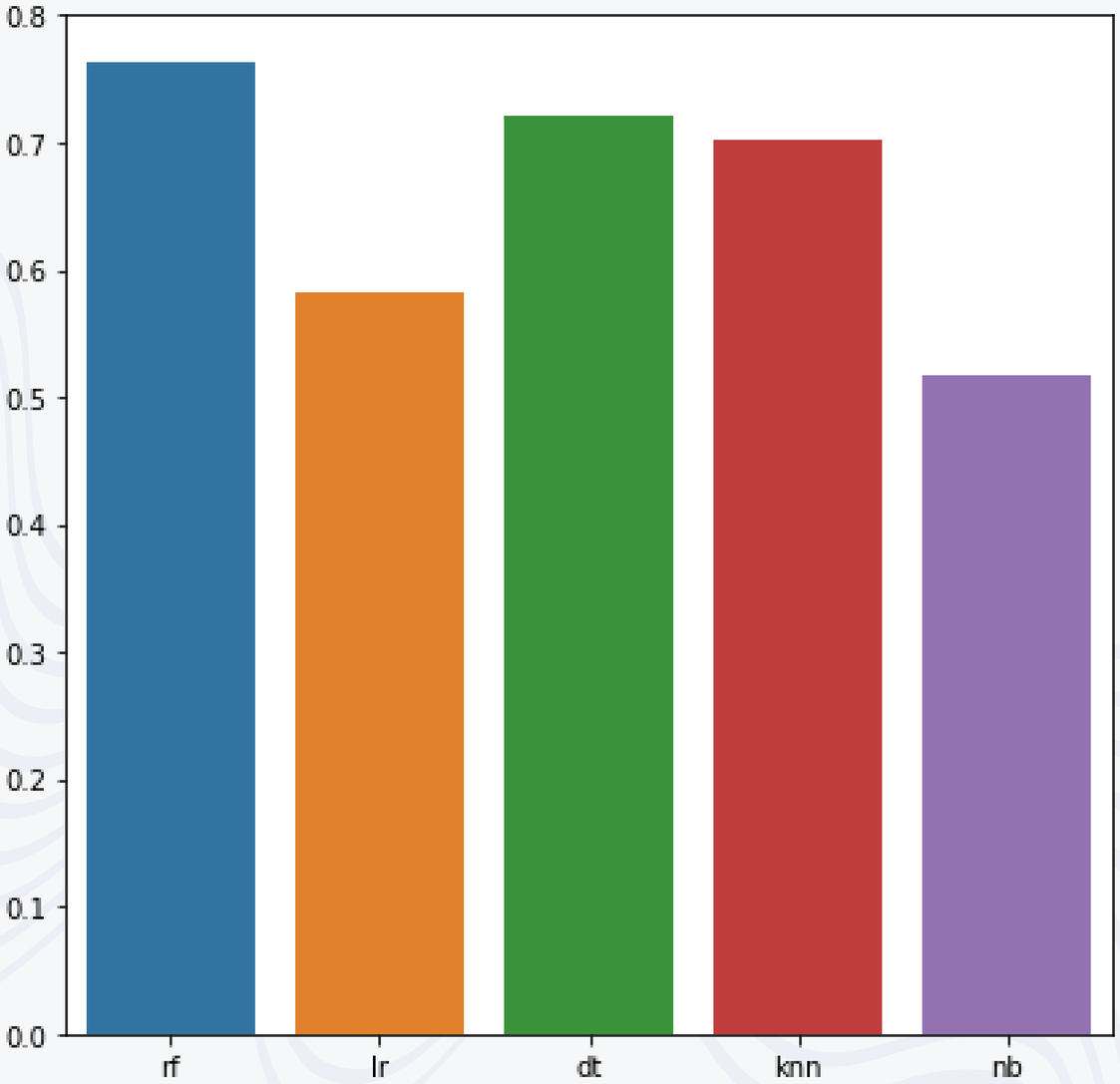
DATA MODELING

CORRELATION



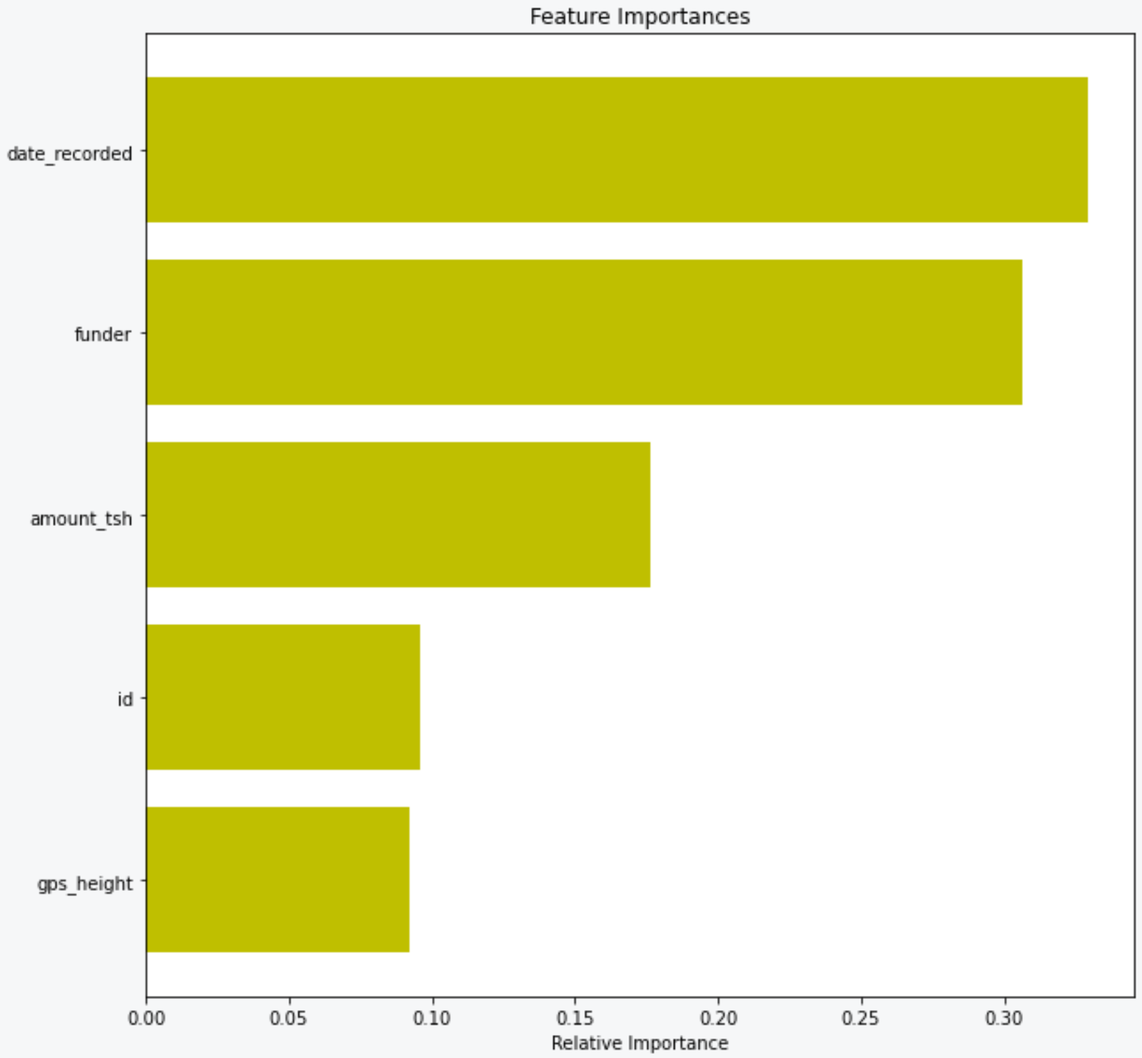
The above heatmap gives us an idea of how each feature correlates to the target value before doing any classification modeling

BASELINE MODELS



- Random Forest Classifier (rf)- Train Accuracy = 100% - Test Accuracy = 76.22%
- Logistic Regression Classifier(lr) - Train Accuracy = 58.80% - Test Accuracy = 58.19%
- Decision Tree accuracy(dt) - Train Accuracy = 100% - Test Accuracy = 72.19%
- KNN accuracy(knn) - Train Accuracy = 80.16% - Test Accuracy = 70.21%
- Gaussian accuracy(nb) - Train Accuracy = 51.93% - Test Accuracy = 51.79%

EVALUATION



After preprocessing and hperparameter tuning the Random Forest Model had the highest accuracy of 76%.This graph show the relative importance of variables as per the model.

CONCLUSION

The final Random Forest Model shows that we can predict the condition of each water pump with 76% accuracy.

We chose this model because of its priority with classifying False Nonfunctional over False Functional. This model is most likely not cost effective because it will prioritize classifying a pump as needing to be replaced over being functional. Because of that prioritization though, this model does provide us with the most humanitarian solution and given the data and our project needs, provides the most useful results.

RECOMMENDATIONS

- Given the above conclusions,the priority should be replacing the pumps that needs replacing as this will go along way in ensuring Tanzania reaches its Vision 2025.
- Water points in densly populated areas should be monitored as this are prone to a lot of wear and tear and serve many people.
- More reseach need to be done to areas with pumps that need replacing so as to establish the real cause before replacing.
- Research should also be carried out on the pumps that require repairs.This is for better understanding on which repairs take priority.

Clean water is essential for life.