

# TANZANIA WATER WELLS ANALYSIS

PRESENTATION TEMPLATE



# Overview

03 Introduction

02 Objectives

05 EDA

06 Modelling

07 Conclusion

08 Recommendations





# Introduction

Tanzania, as a developing country, struggles with providing clean water to its population of over 65.5M (as of 2022).

More often than not, the clean water a developing community desperately needs is right underneath them and a relatively small investment to get to it can make such a dramatic difference in so many lives.



# Objectives



- GreenPace as an NGO aims to alleviate the water crisis in Tanzania by establishing more water wells in the country.
- They also aim to partner with the best organizations and choose the most suitable geographical locations to ensure the best outcome.
- The model built is supposed to help make predictions that will inform the NGO's final decisions.

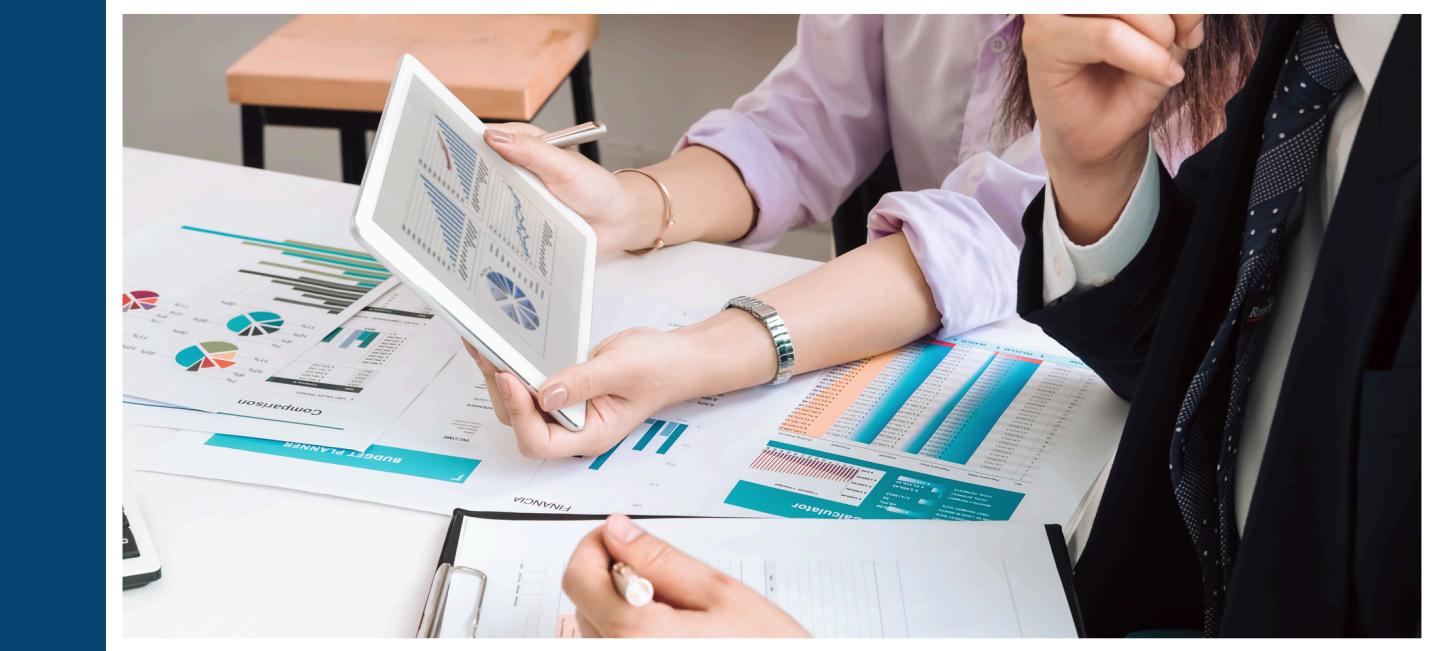


# Specific Objectives

- To build a classifier that predicts the condition of water well as either functional or non-functional.
- To determine the regions with more chances of having functional water points for future planning
- To determine the funders funding a high number of functional water points.

# Exploratory Data Analysis

The dataset used in this analysis is the Tanzanian Water Well dataset, which is subdivided further into three datasets; Training\_set\_values, Training\_set\_labels and Test\_set\_Values.



The data contained in the dataset contained many null values, similar values(duplicates) and missing values. Data cleaning was thereby paramount to address these issues, before the modeling process.

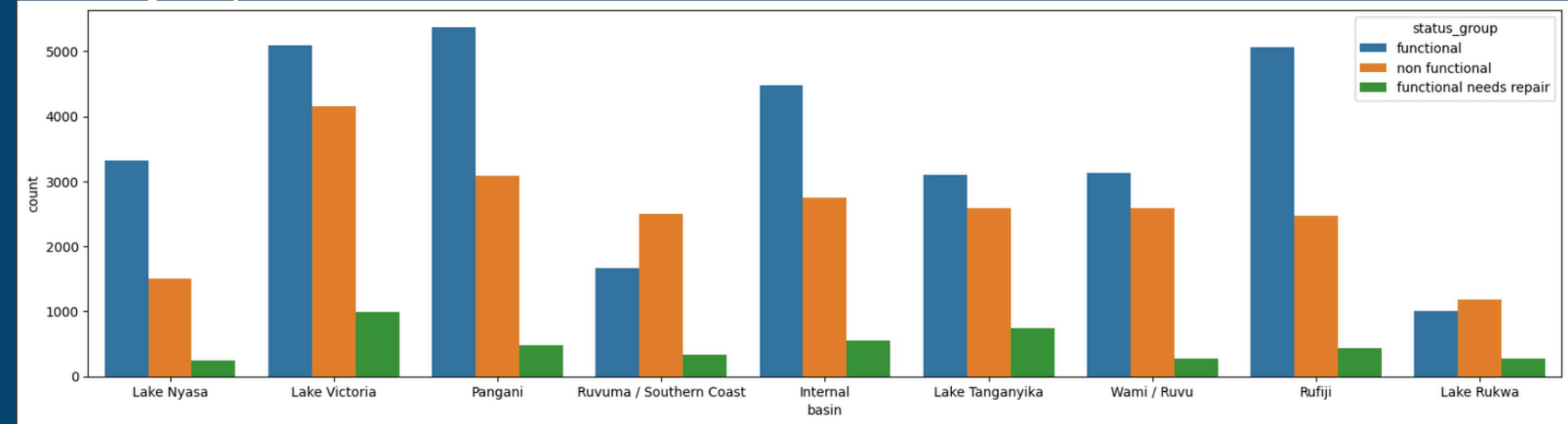


# Visualizations

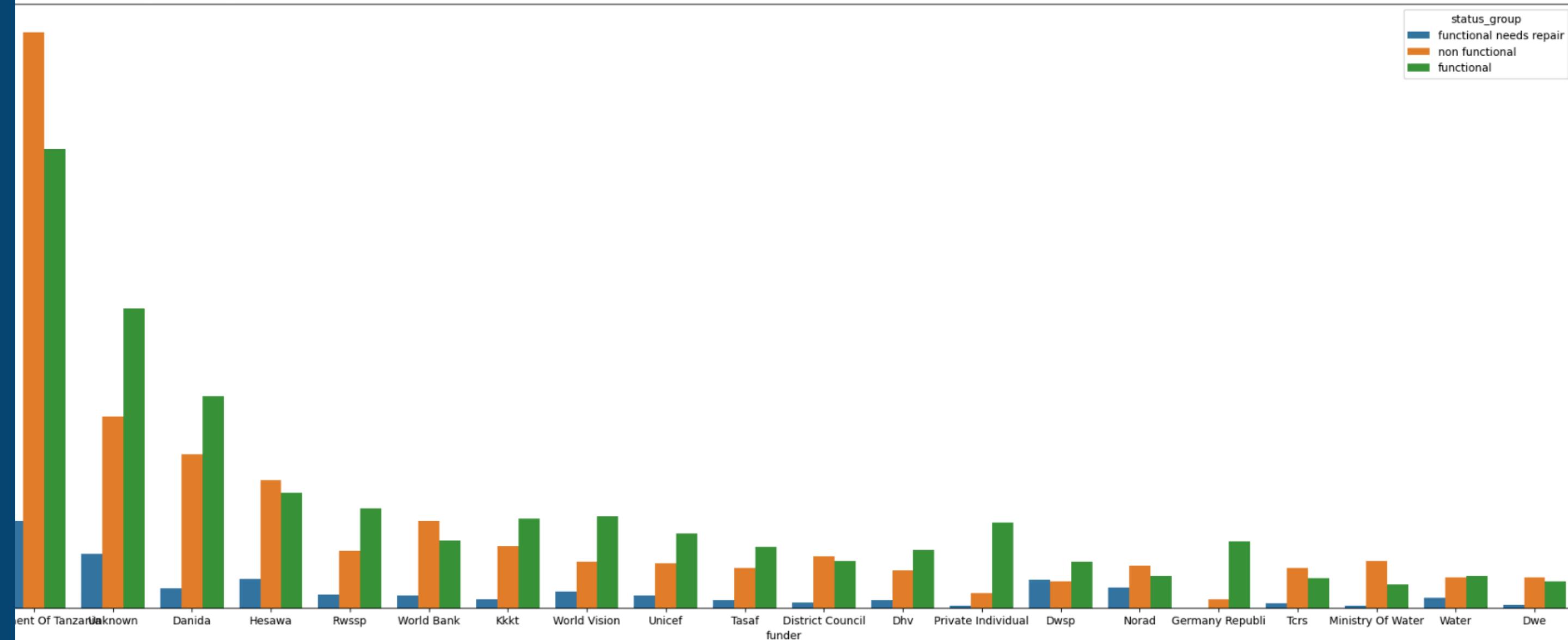
We also observed the relationships of the different status groups vs features such as basins, funders, installers, gps height among others.

The next slides display some visualizations that show this:

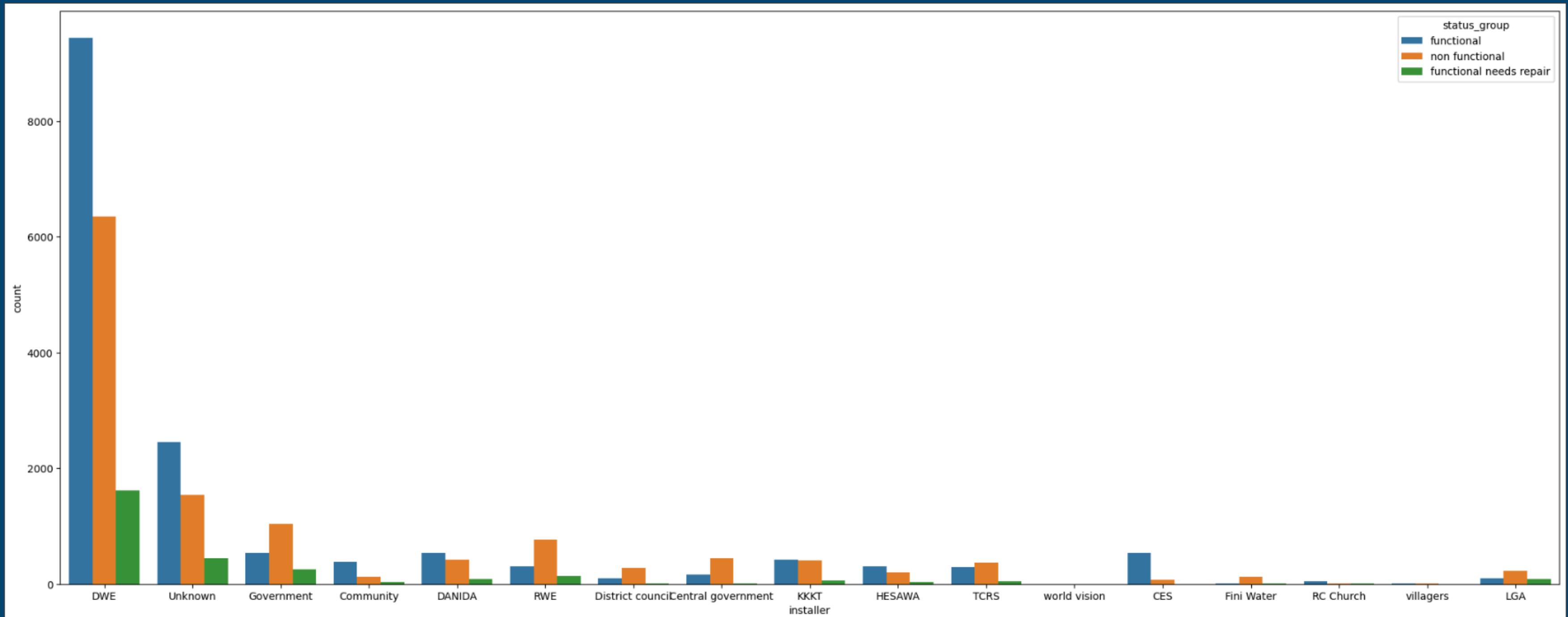
## status group vs basin



## status group vs funder



## status group vs installer





# Modelling

We decided to use a logistic regression model, a confusion matrix and a decision tree classifier.

The logistic regression model acted as our baseline model.





# Model Evaluation

- **Different metrics were used in evaluating the chosen models. These are accuracy score, precision and recall.**
- **We established that the best mode to use was the decision Tree Classifier because it had a higher accuracy than the Logistic Regression Model.**



# Conclusion and Recommendations

## Conclusion

- There are typically more functional wells in locations with higher population densities.
- Certain locations are more likely to have access to clean water, particularly if they are close to the appropriate basins.
- The majority of government-funded wells are inoperable.
- The majority of the water points established by the district council and central government are not working.
- The Logistic Regression model had an accuracy score of 62% while the Decision Tree Classifier model yielded an accuracy score of 64%.

## Recommendations

- The project executers should carefully consider the geographical locations of their wells since this is directly linked to the functionality/quantity of water that a well can yield.
- We recommend DWE as a potential partner(if needed) since the analysis has shown that most of the wells they have funded/installed are functional.
- The gps height should also be considered in order to determine the type of pumps to be used at the wells.
- Proximity to basins like Rufiji, Lake Victoria and Pangani is recommended for higher productivity.



# THANK YOU