

# PREDICTING OPERATING CONDITION OF WATER WELLS IN TANZANIA

This project involves building a data-driven model to predict the condition of water wells in Tanzania.

By Grace Rotich



## BUSINESS UNDERSTANDING

The project's main goal is to create a model for forecasting operational conditions of water points in Tanzania.

With accurate predictions showing whether a water point will be working, broken, or under repair, the Tanzanian government will improve maintenance decision-making procedures, thereby making sure that communities have sustainable access to drinking water.

## PROBLEM STATEMENT

According to Water.org(2024), Tanzania faces a significant water and sanitation crisis, out of its population of 65 million people, 58 million people (88% of the population) lack access to safe water.

# OBJECTIVES

1. To understand the data and the various factors that affect the condition of water wells in Tanzania.
2. To create a machine learning model that can reliably forecast the operational state of water wells Using past data.
3. To give advice on how to allocate funds and resources more effectively for water wells maintenance and repair.



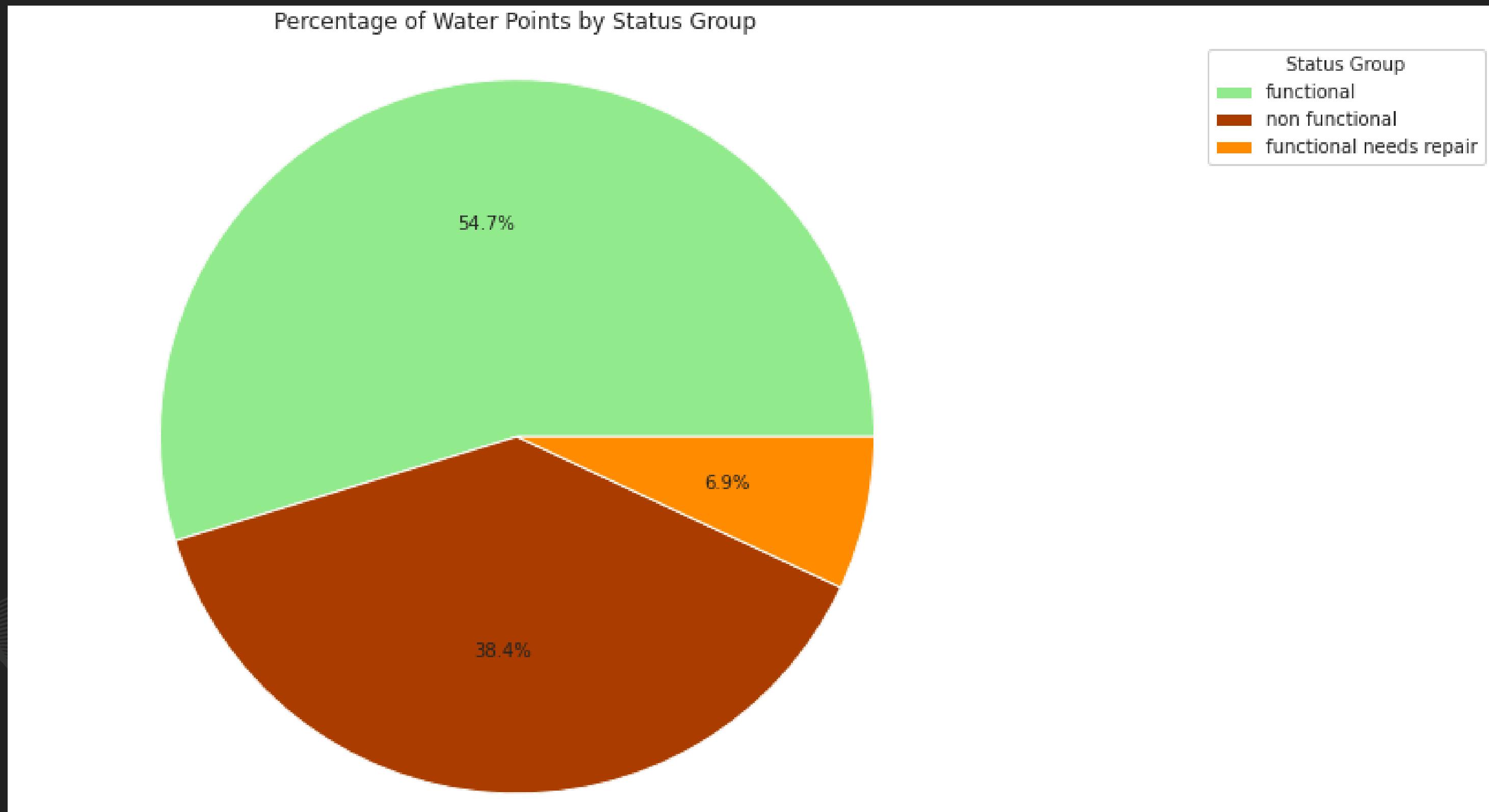
# DATA UNDERSTANDING

- The data was sourced from from Taarifa (<http://taarifa.org/>) and the Tanzanian Ministry of Water(<http://maji.go.tz/>).
- The data had 59400 individual data points and 41 columns



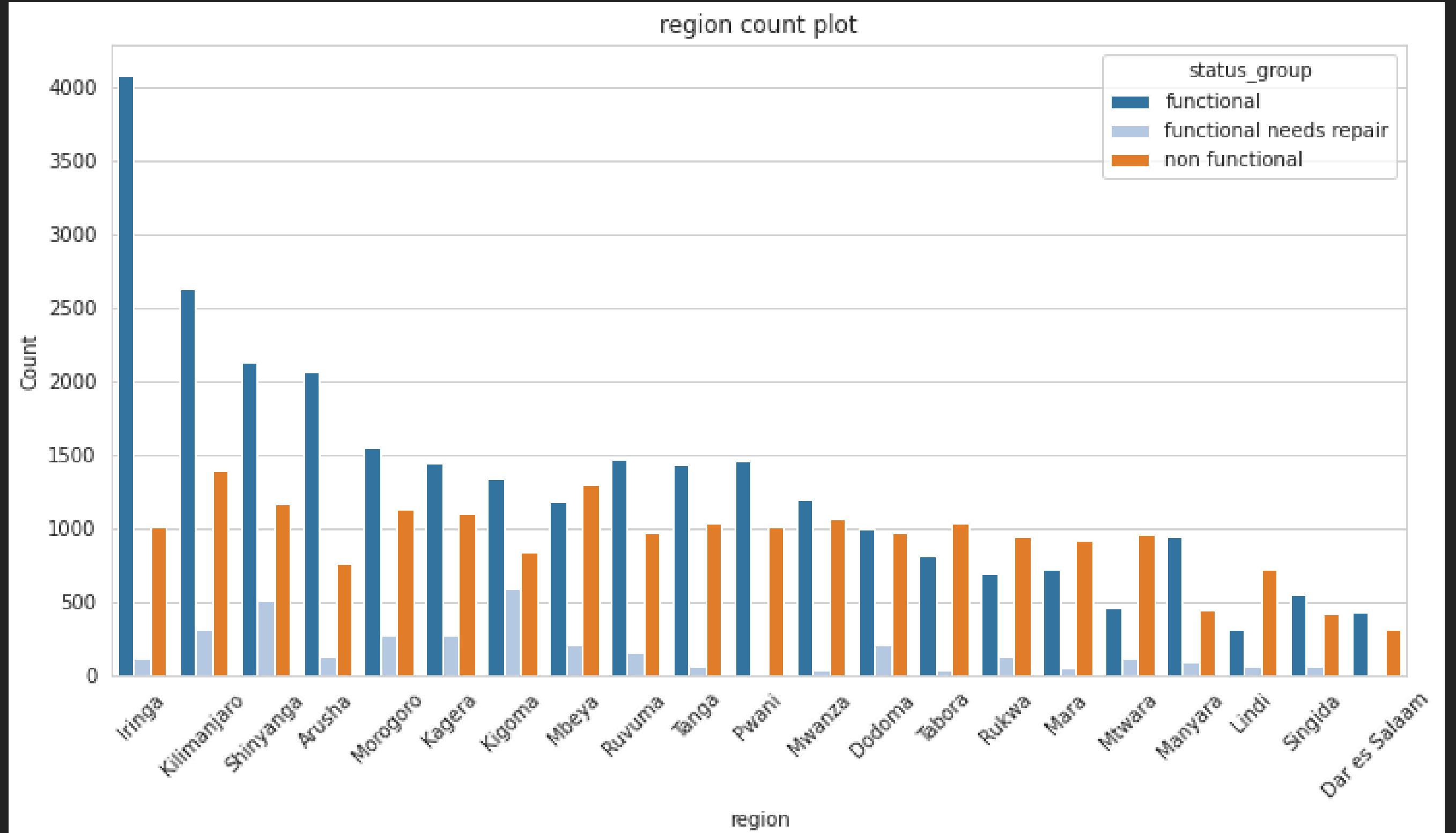
# EXPLORATORY DATA ANALYSIS

The pie chart shows the distribution of wells according to functionality



# EXPLORATORY DATA ANALYSIS

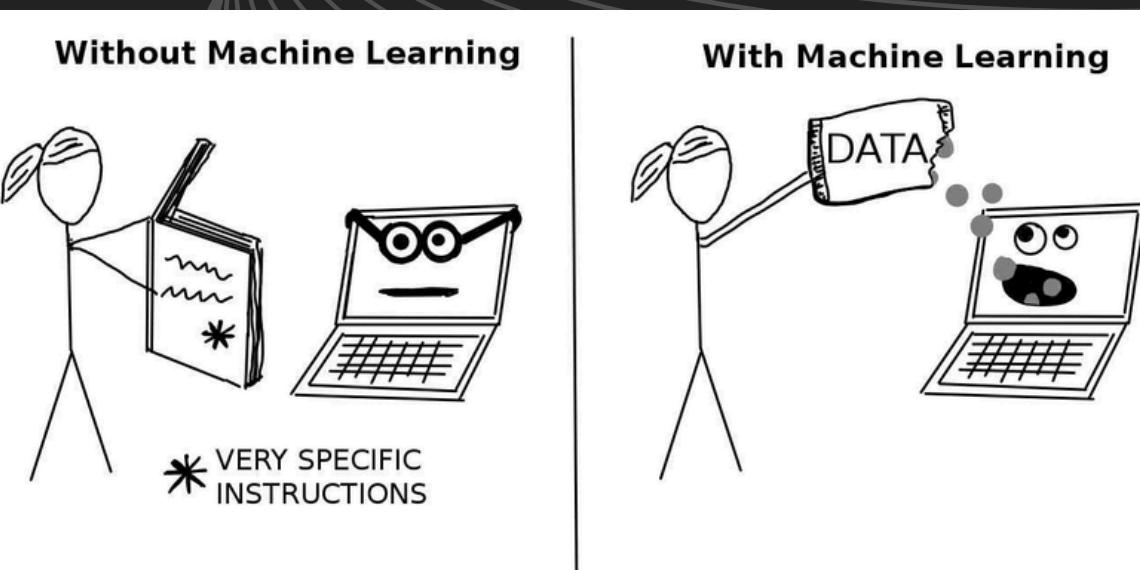
- The graph is showing the counts of water points from each region in TZ.
- Most numerical features have a positive skew.



# MODELLING

For this study, the following models were employed:

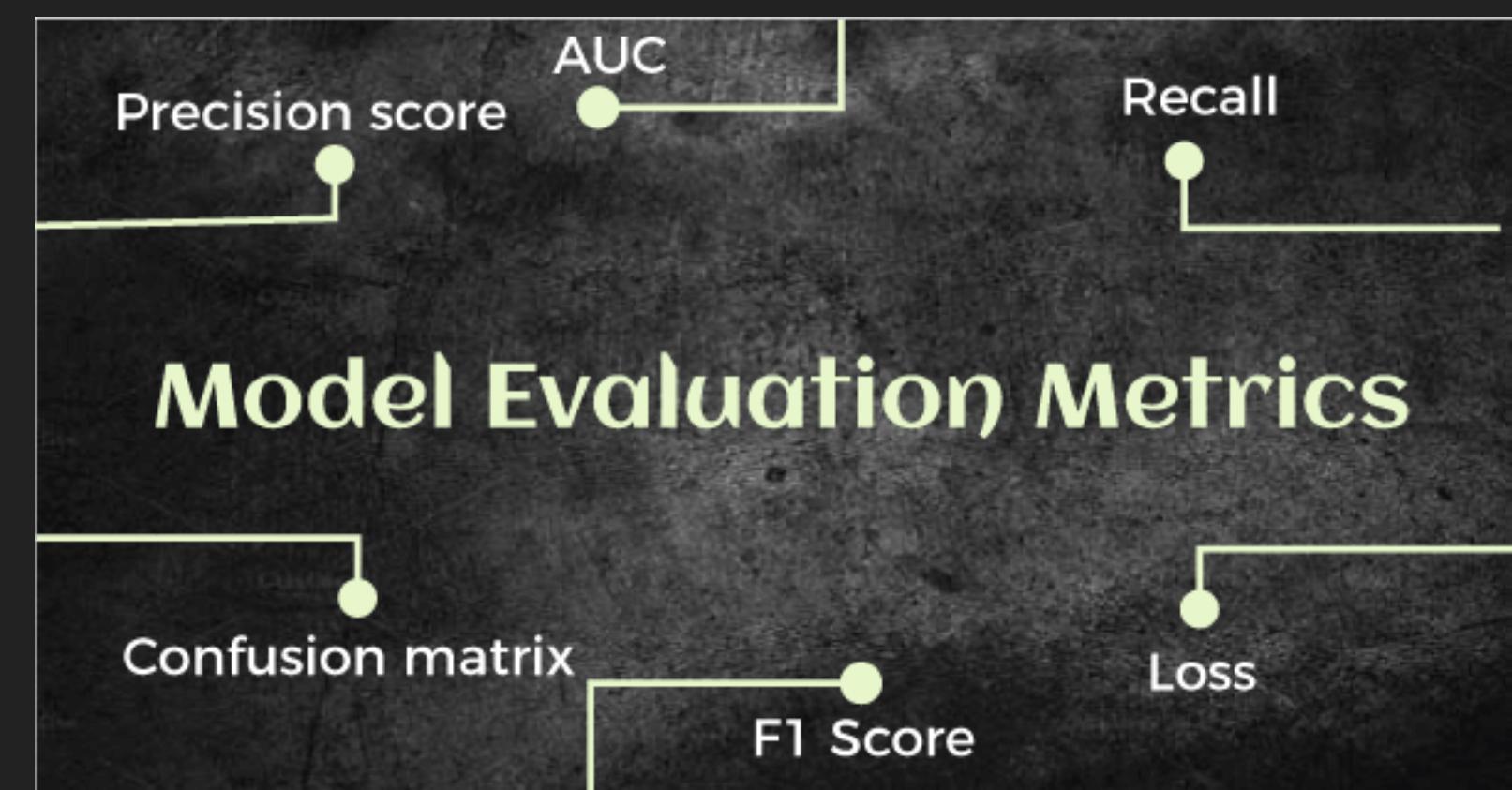
1. Decision Trees
2. Logistic Regression
3. K-Nearest Neighbors
4. Random Forest



# EVALUATION

For this project, the following evaluation metrics were used to rate the performance of the models:

1. Accuracy
2. Precision
3. Recall
4. F1 Score



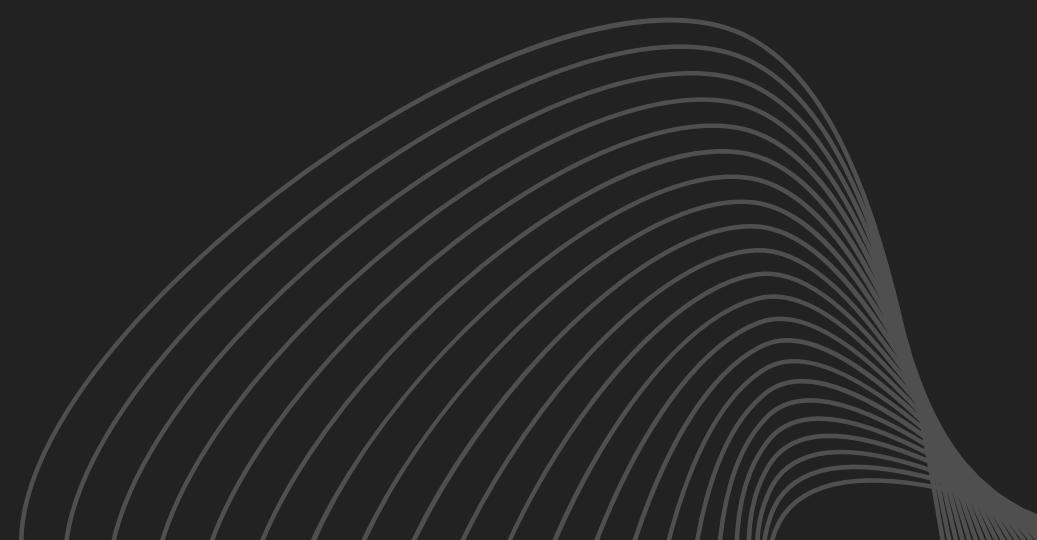
# MODEL PERFORMANCE

The table below shows the metrics of performance observed for the four models used.

	Model	Accuracy	Precision	Recall	F1_Score
0	Decision_Tree	63%	64%	63%	59%
1	Logistic_Regression	59%	56%	59%	53%
2	KNN	62%	60%	61%	60%
3	Random_Forest	66%	66%	66%	62%

# RECOMMENDATIONS

- The random forest model being the best performing model based on the performance metrics should be adopted.
- The Tanzanian government should prioritize drawing water from springs.
- Wells with enough water should be closely monitored and maintained, as they are frequently put to use by the population and hence need more maintenance.





**THANK YOU!**



Questions  
Comments  
Feedback

END