Brian Chabari Nyagah.

TANZANIA WATERWELLS PROJECT.

PROJECT OVERVIEW.

- Access to clean and reliable water is crucial for the well-being and development of communities. In Tanzania, like many other countries, water wells play a vital role in providing this essential resource. However, ensuring the functionality and sustainability of these wells can be challenging due to various factors such as geographic location, infrastructure maintenance, and water quality.
- The Tanzania Water Wells project aims to leverage data science techniques to predict the functionality of water wells across different regions in Tanzania. By analyzing historical data and identifying patterns, we can develop a predictive model that will aid in optimizing maintenance efforts, resource allocation, and decision-making processes related to water well management.

OUTLINE.

- Data Understanding.
- Data Preparation.
- Exploratory Data Analysis.
- Feature Engineering.
- Model Selection and Training.
- Model Evaluation.
- Deployment and Recommendations.

PROBLEM STATEMENT.

The main challenge faced by water authorities and organizations in Tanzania is the high number of non-functional or partially functional water wells. These dysfunctional wells result in limited access to clean water for communities, leading to health and sanitation issues, reduced agricultural productivity, and hindered economic growth.

OBJECTIVE.

The objective of the Tanzania Water Wells project is to develop a predictive model that accurately determines the functionality of water wells in Tanzania. By leveraging historical data and applying data science techniques, the project aims to address the challenge of non-functional or partially functional water wells, ultimately improving access to clean water for communities.

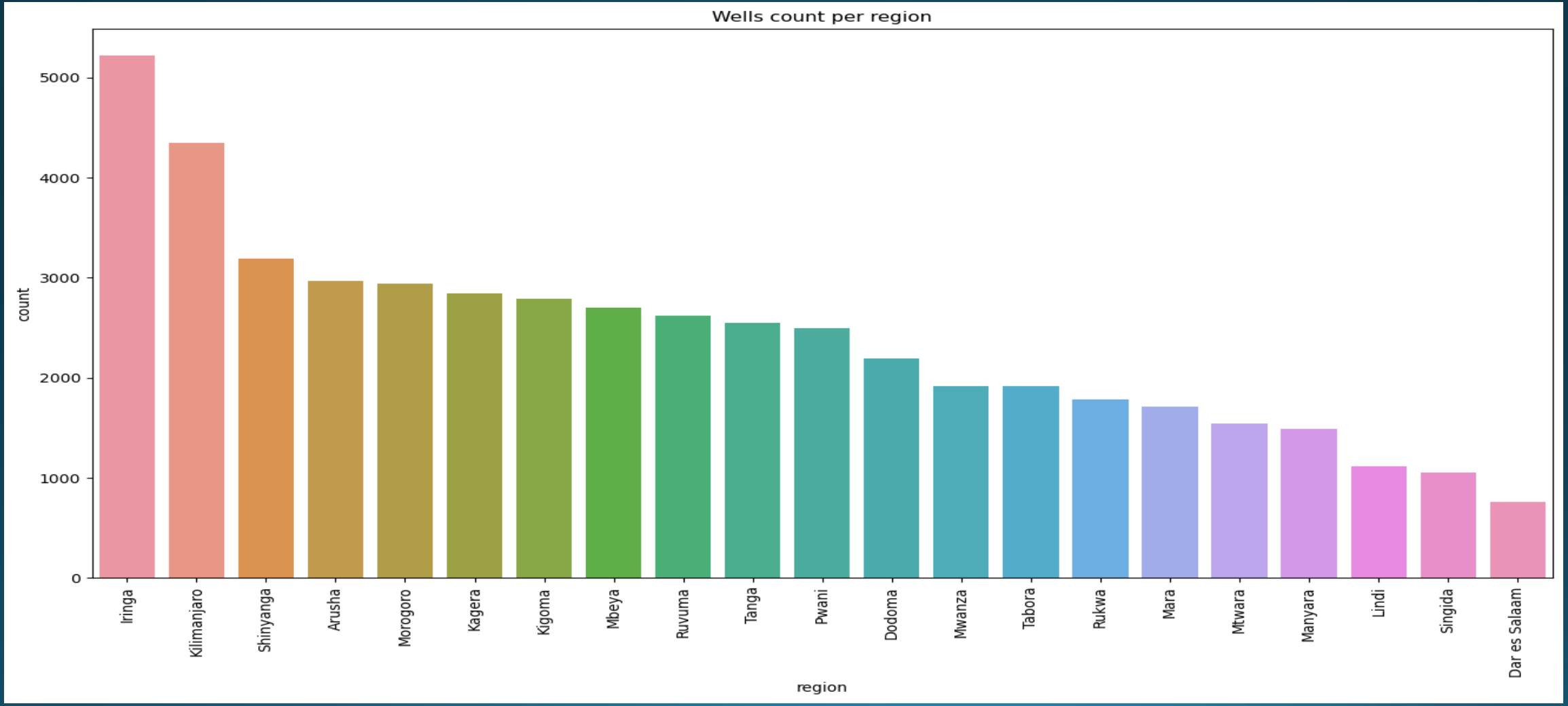
DATA UNDERSTANDING.

The original data was obtained from the Driven Data 'Pump it Up: Data Mining the Water Table' competition. Basically, there are 4 different data sets; submission format, training set, test set and train labels set which contains status of wells. With given training set and labels set, competitors are wanted to build predictive model and apply it to test set to determine status of the wells and submit. In this project, we used train set and train label set which have 59400 water points data with 40 features.

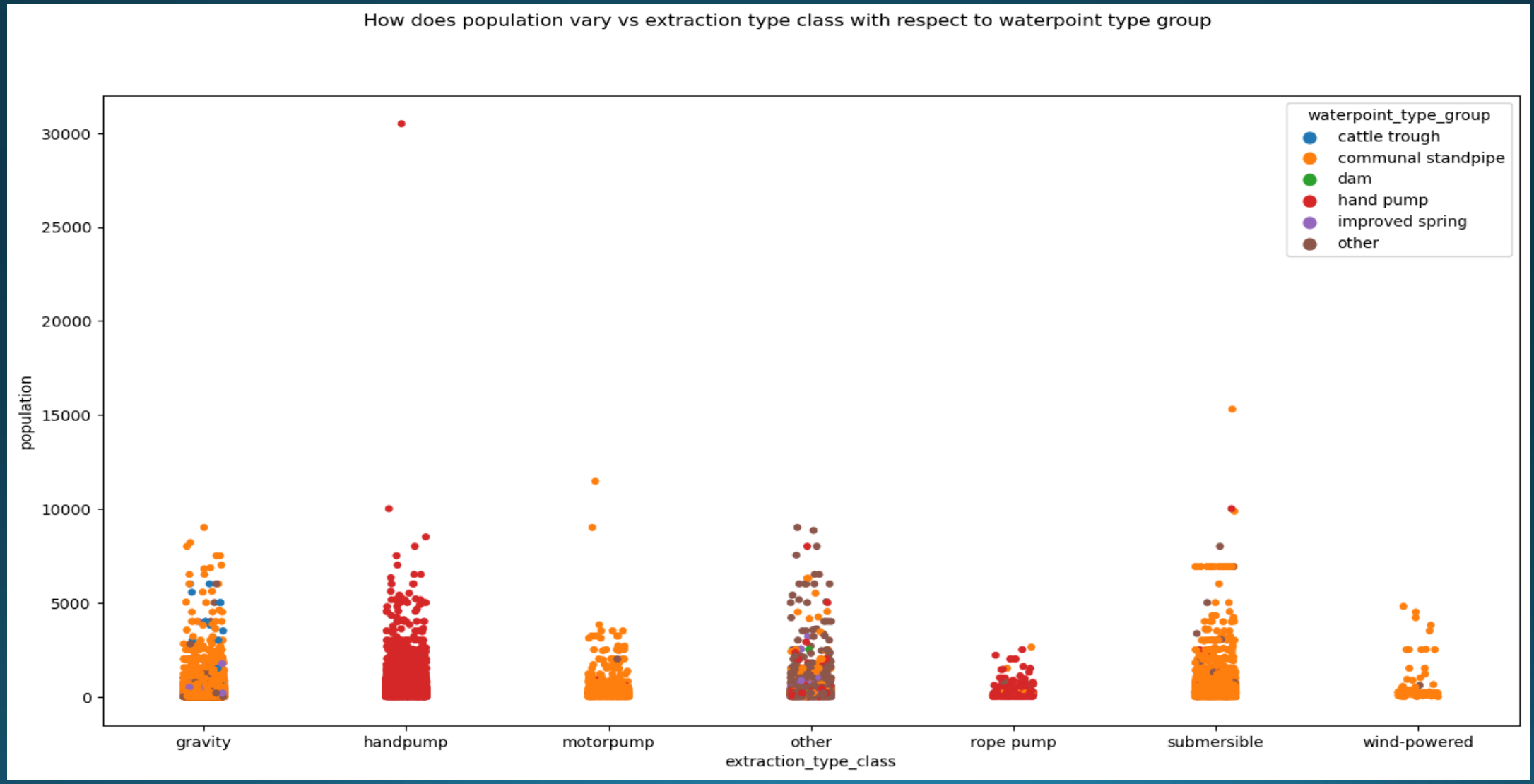
EXPLORATORY DATA ANALYSIS.

In the Exploratory Data Analysis (EDA) phase, we delve into understanding the dataset and gaining insights from its structure and content. Through visualizations, statistical summaries, and data exploration techniques, we aim to uncover patterns, relationships, and potential outliers within the data.

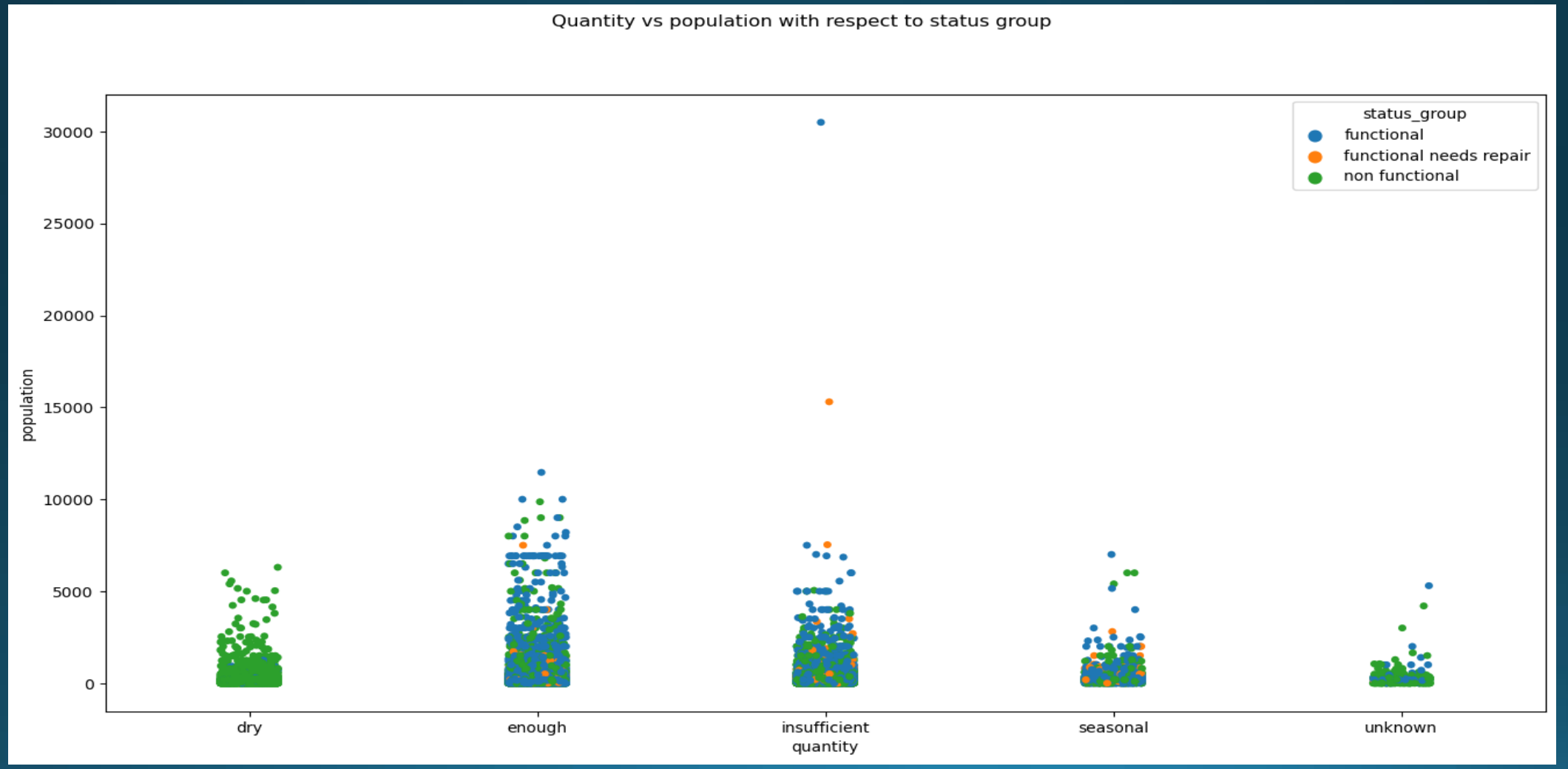
Distribution of Waterwells by region.



Relationship between extraction_type_class and population with respect to waterpoint_type_group



Relationship between quantity vs population with respect to Status_group.



MODELLING.

- Models used:
- Logistic regression
- Random Forest
- Knearest Neighbours
- Xg boost

EVALUATION.

- The best overall model is the Kneighbors Classifier baseline model. It achieved an accuracy of 71% on the testing data, indicating that it correctly predicts the functional class around 71% of the time.
- The reason for selecting this model is that it demonstrates a balanced performance between the training and validation accuracy scores.

FINDINGS.

- Functional water wells were able to provide enough water to the population.
- A majority of wells found in the most populated areas need repairs, probably due to scrambling and over usage.
- Majority of the functional wells provided enough water. However, there is an alarming number of non functional wells.
- The functional water points were the most populated.
- Most water points that are not paid for are non functional.
- Wind powered and rope pump have the least amount of water available.
- The most common extraction type is gravity followed by hand pumps.

RECOMMENDATIONS.

- The Government of Tanzania should put proper measure to ensure the maintenance and functionality of their water points as a big number of them are non functional or need repairs.
- The Government of Tanzania should make effort to ensure availability of water to the sparsely populated areas too as most of these sparsely populated areas did not seem to have enough water.
- The Government should channel more funds towards communal standpipe as most of them are non functional.