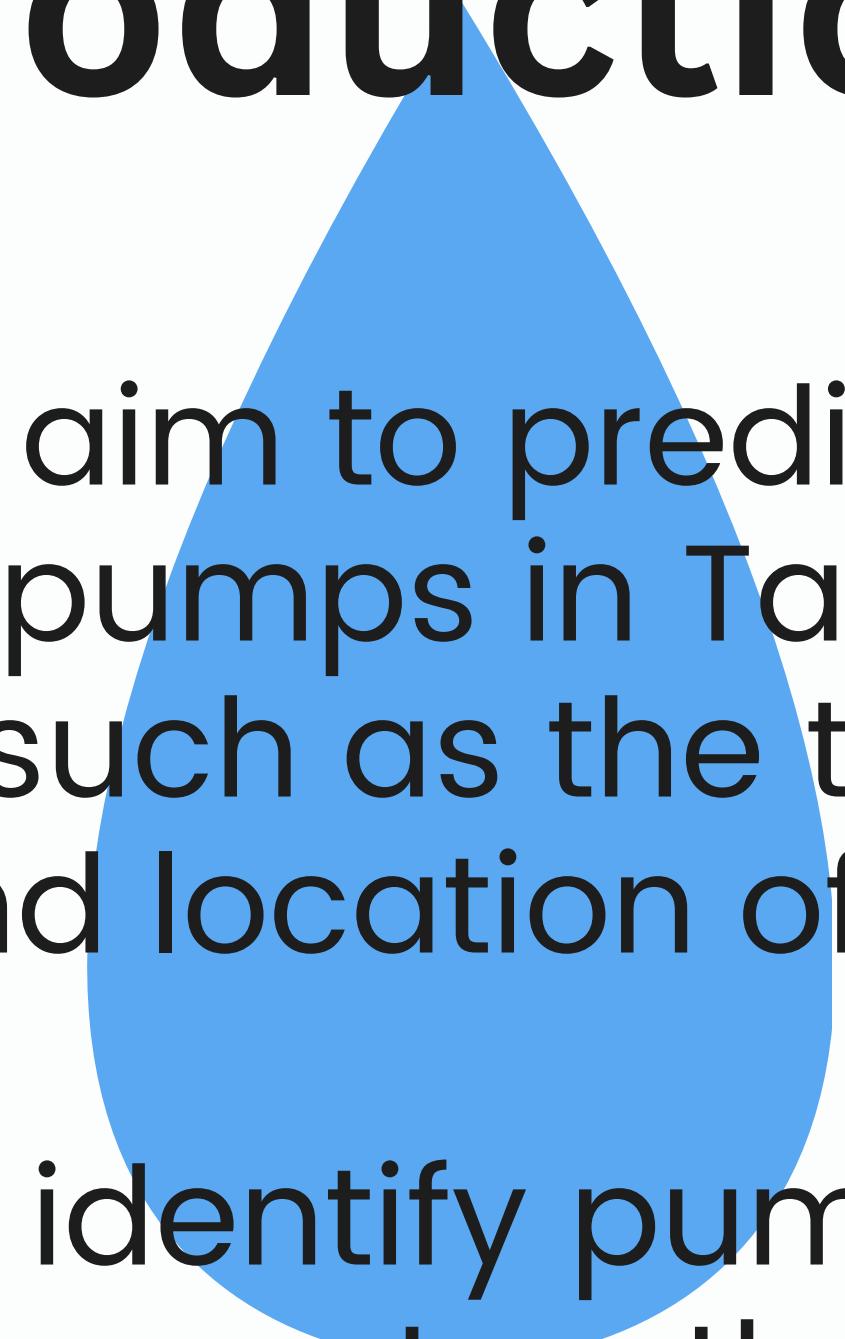


WATER WELLS FUNCTIONALITY IN TANZANIA

Introduction



In this project, we aim to predict the functionality status of water pumps in Tanzania based on various attributes such as the type of pump, water quality, and location of the pump.

The goal is to help identify pumps that are in need of repair or replacement so that the communities relying on them have access to clean and safe water.

PROBLEM STATEMENT

With a population of over 57 million, it's important that Tanzania's water wells are functional and provide safe drinking water.

Still, MSABI, one of the largest WASH NGOs in the country, is struggling to improve access to clean water as it needs adequate information about the condition of the wells, making it difficult to prioritize its efforts and allocate resources effectively.

MAIN OBJECTIVE

To build a classifier that can predict the condition of water wells in Tanzania and improve access to clean water for communities in the country.



SPECIFIC OBJECTIVES

- Clean and pre-process the data to remove missing values and outliers.
- Conduct exploratory data analysis to identify patterns and relationships in the data.
- Build and train a machine learning model .
- Evaluate the model using appropriate metrics and techniques.
- Optimize the model to achieve the best possible accuracy.



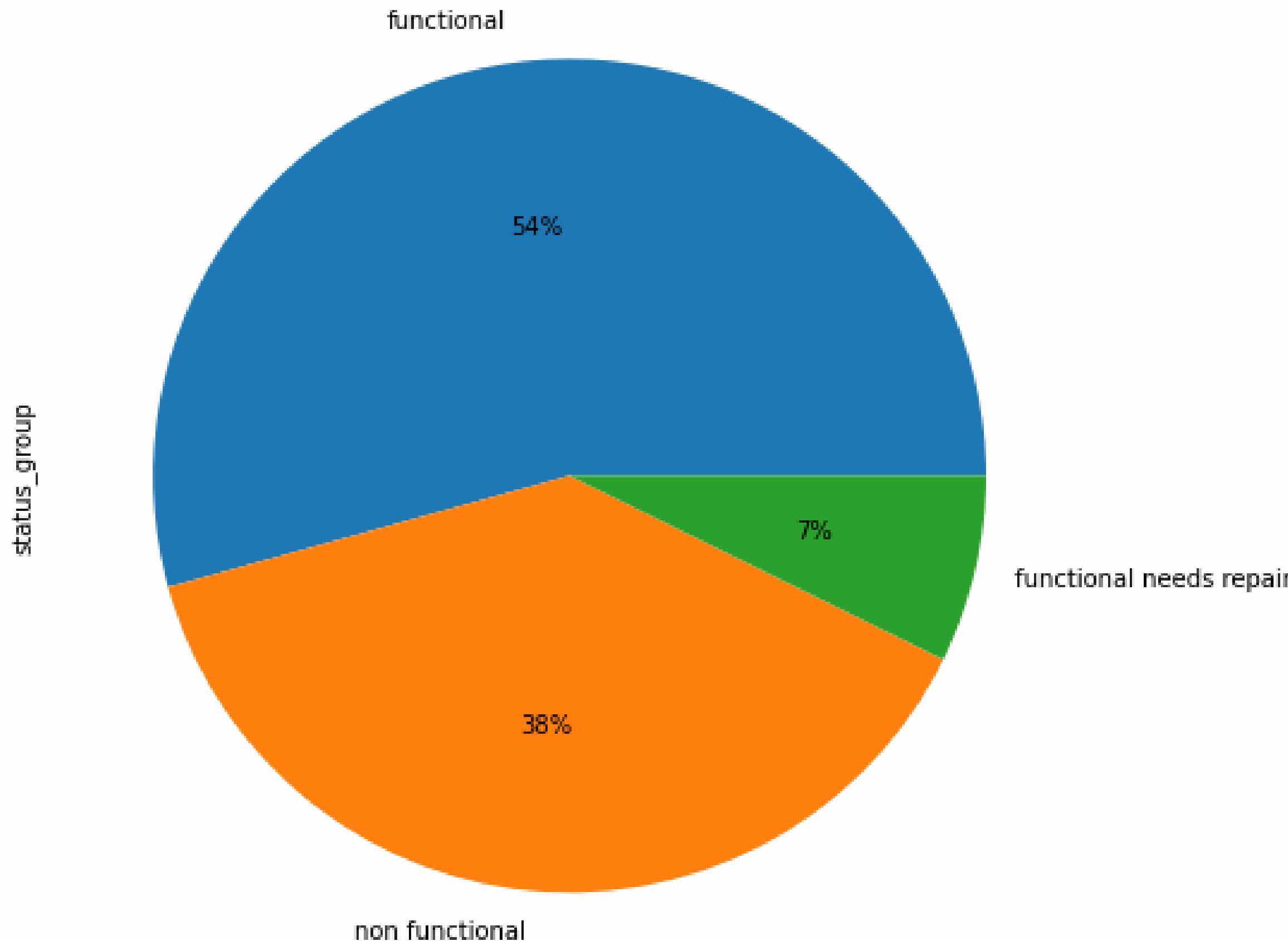
Data Pre-processing

The data for this project was sourced from Taarifa and the Tanzanian Ministry of Water.

The initial dataset was pre-processed to handle missing values and convert categorical variables into numerical data.

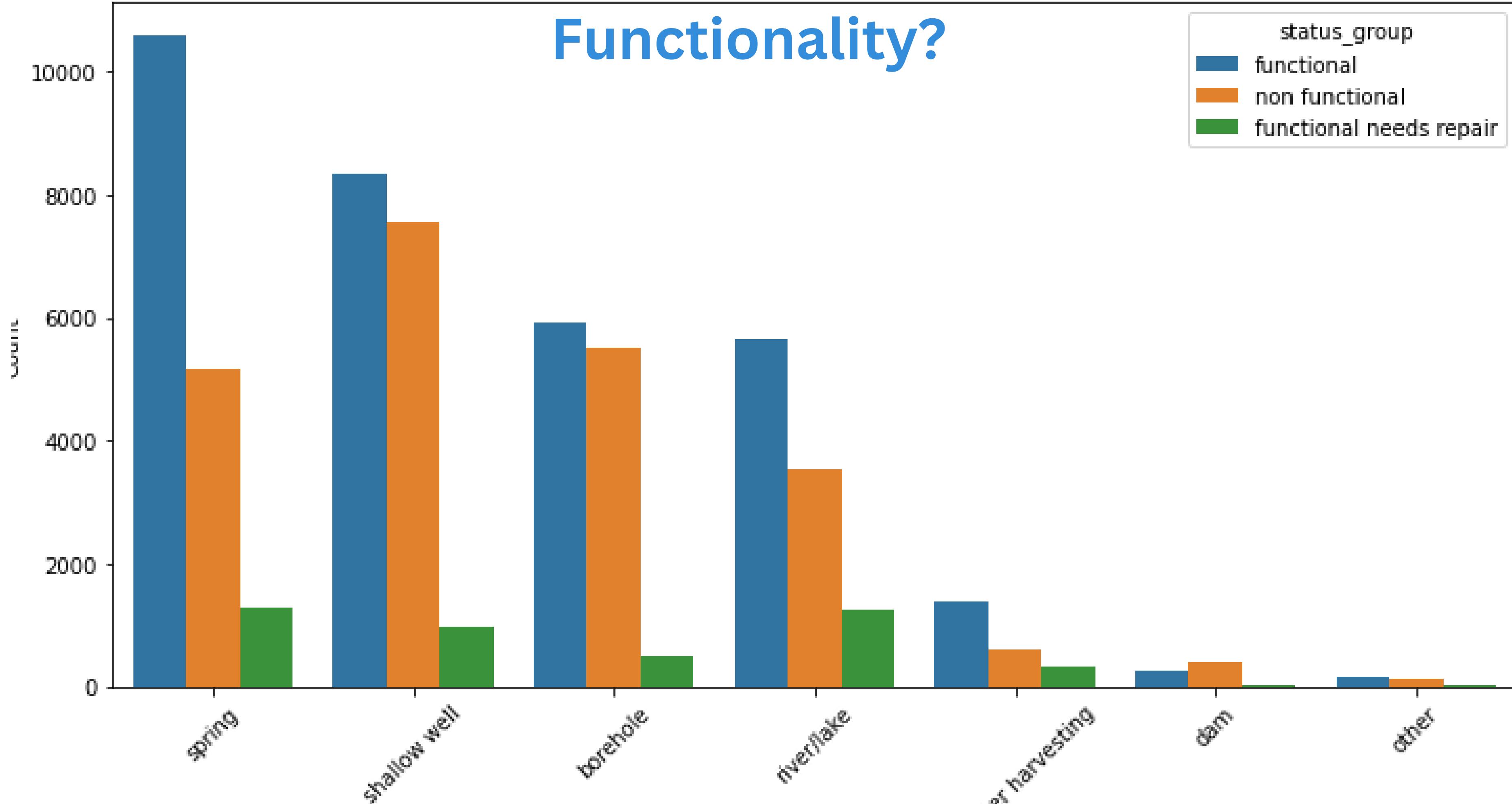
Exploratory Data Analysis

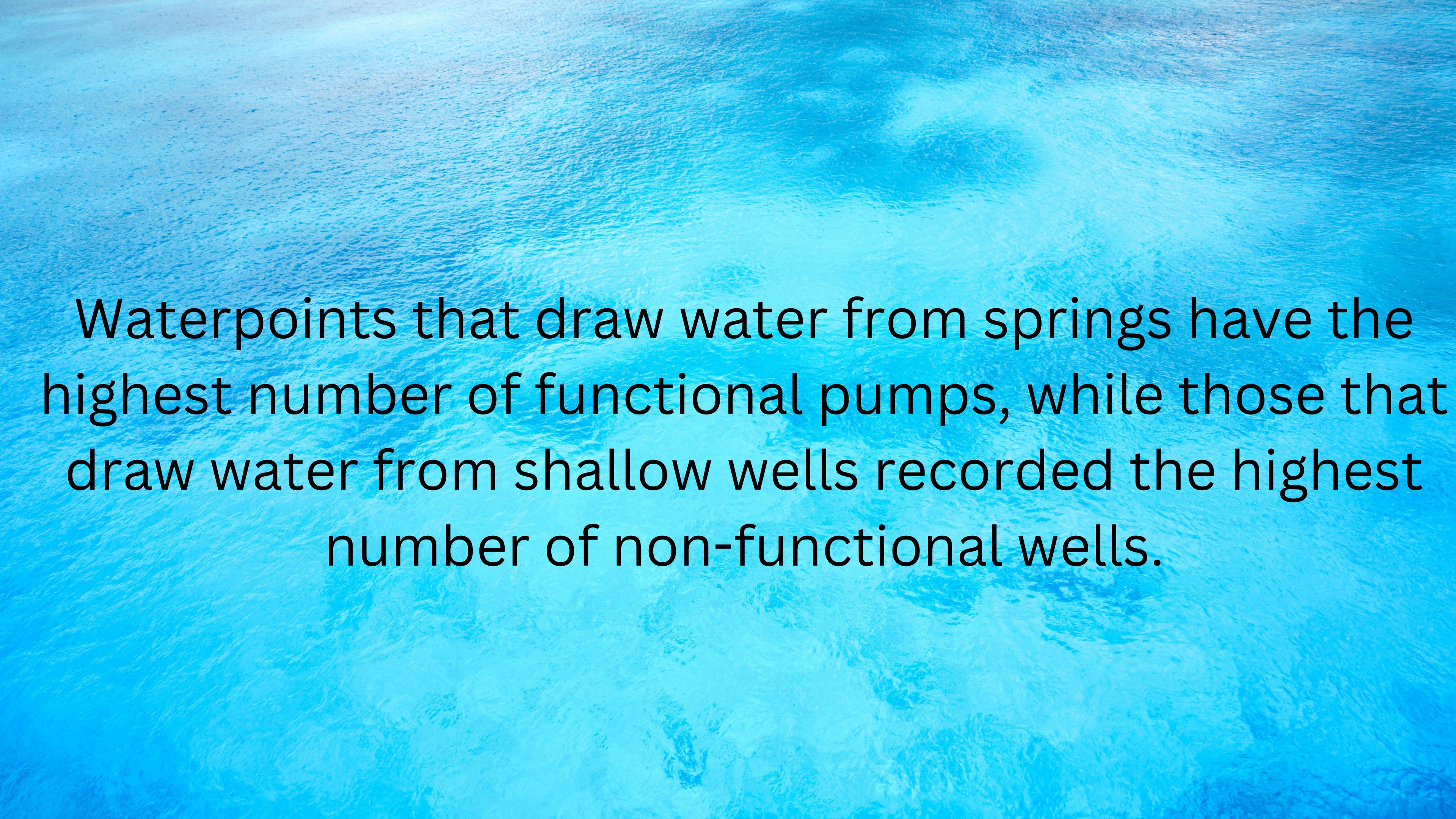
WaterPoint Status Groups



55% of waterpoints are functional, 38% are non-functional, and 7% are functional needs repair.

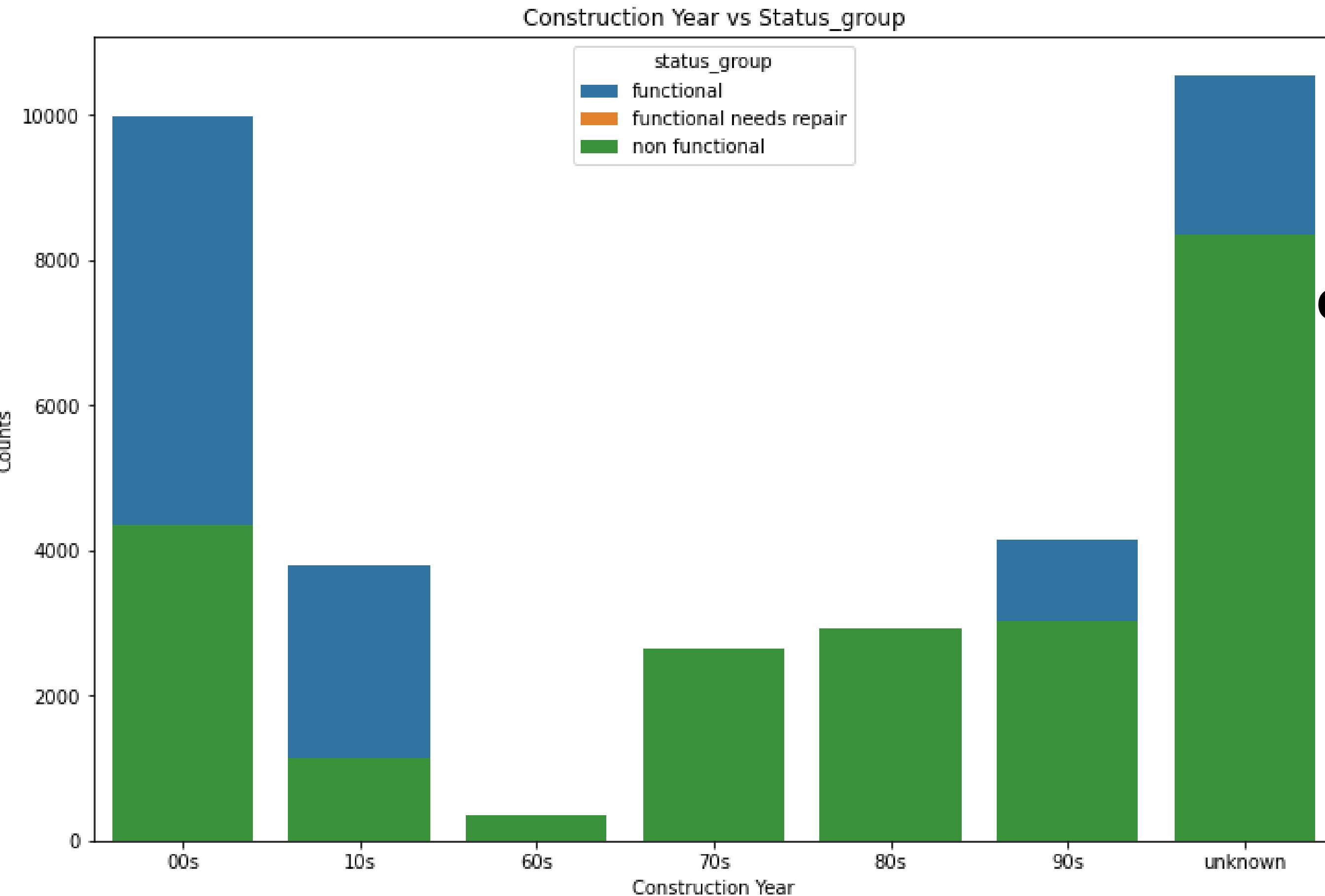
Do Different water sources Affect Waterpoint Pump





Waterpoints that draw water from springs have the highest number of functional pumps, while those that draw water from shallow wells recorded the highest number of non-functional wells.

Does the Construction Year affect Well Status?



Most of the wells constructed between the 60s and 80s are mostly non-functional.

Modelling

In order to build a predictive model for the water pump functional status, I used the Gradient Boosting Classifier and Random Forest Classifier algorithms.

I first split the available data into training and validation sets to assess the performance of our models.

Then, I performed a grid search over different hyperparameters for the Random Forest Classifier to find the best set of hyperparameters that would give us the highest validation accuracy.

Results

The best model I found using the grid search had a validation accuracy of 0.796. The best hyperparameters found for the Random Forest Classifier were: 'min_samples_split': 8, 'n_estimators': 1000.

Conclusion

The model is able to predict the functionality status of a water pump with an accuracy of 0.796, which is a relatively good performance. This can be used to aid in decision-making and maintenance planning for the water pumps in Tanzania.

However, it is important to note that this is a limited sample, and there is room for improvement with more data or additional feature engineering.

Recommendations

- Further feature engineering could be done to improve the model performance.
- The model could be tested on additional data to ensure its generalizability.
- The model could be improved with the use of other algorithms or a combination of multiple algorithms.

Any
Question



THANK
YOU