TANZANIA
WATER WELLS

TABLE OF CONTENTS

OVERVIEW

IN TANZANIA

16 MILLION PEOPLE DO NOT HAVE ACCESS TO SAFE DRINKING WATER

40 MILLION PEOPLE LACK ACCESS TO IMPROVED SOURCES OF DRINKING WATER

**60%** HOUSEHOLDS IN TANZANIA CURRENTLY HAVE ACCESS TO A BASIC WATER-SUPPLY

BUSINESS UNDERSTANDING

Despite efforts to improve water access in Tanzania, significant challenges persist in ensuring sustainable and reliable access to clean water for all communities. The lack of accurate predictive models for water wells hampers efficient planning and resource allocation, resulting in suboptimal drilling locations, unreliable well yields, and inadequate maintenance strategies. As a result, communities continue to face water scarcity, health risks from contaminated water sources, and economic hardships.

DATA UNDERSTANDING

In this project we shall use a dataset containing information about existing water wells in Tanzania sourced from an ongoing DrivenData competition.

The dataset contains 59,400 records and 40 columns. Of these columns, we identified 31 to be categorical, and 9 as numerical. We were able to further group the columns into the general features being captured.

Distribution of region by status_group
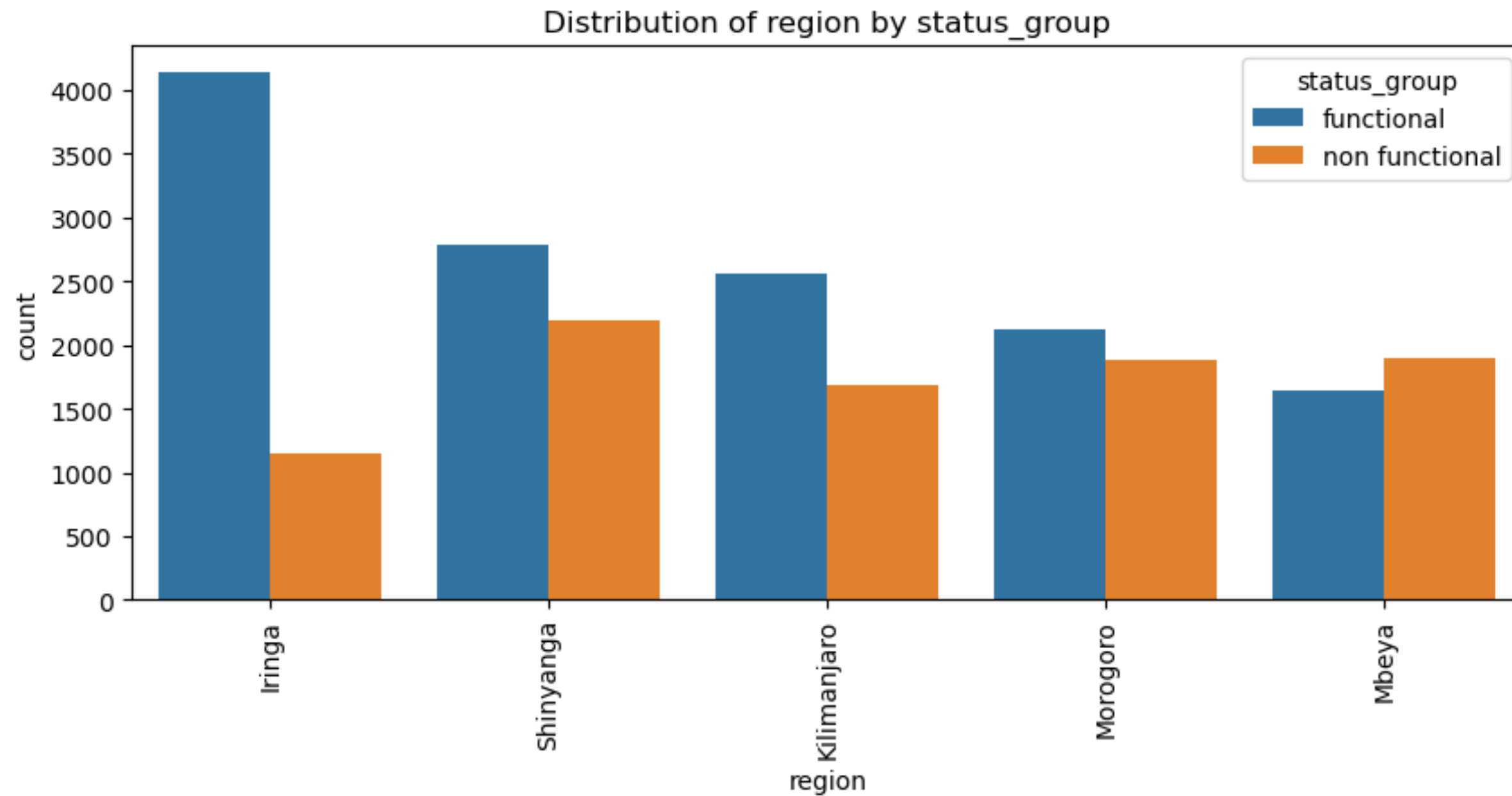
From the distribution above, Iringa region has the most functional wells compared to the rest of the regions followed by Shinyanga, Kilimanjaro, Morogoro and Mbeya.
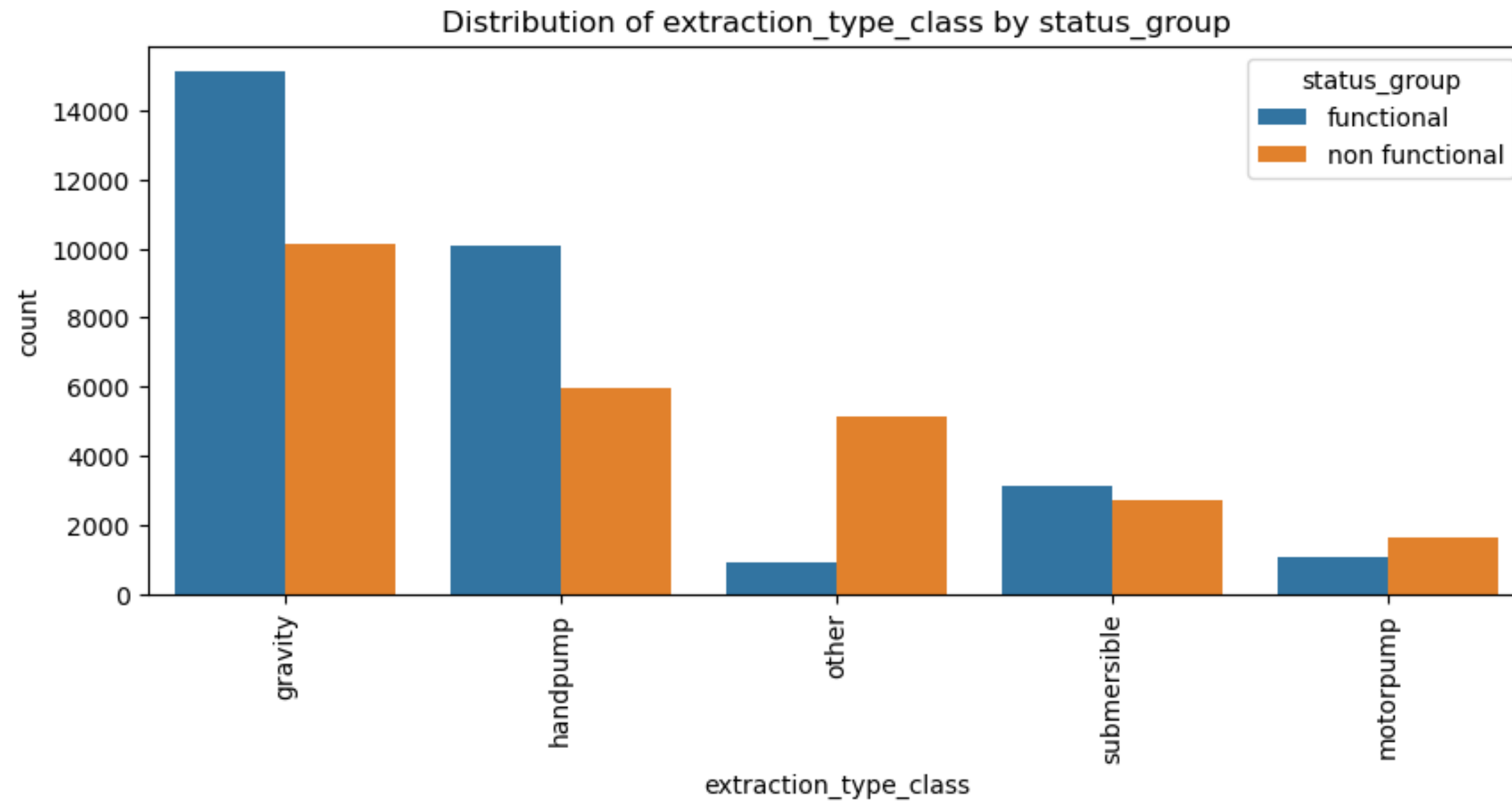
Njombe region has the most functional wells followed by Moshi Rural, Arusha Rural, Bariadi and Kilosa.

Distribution of source_type by status_group
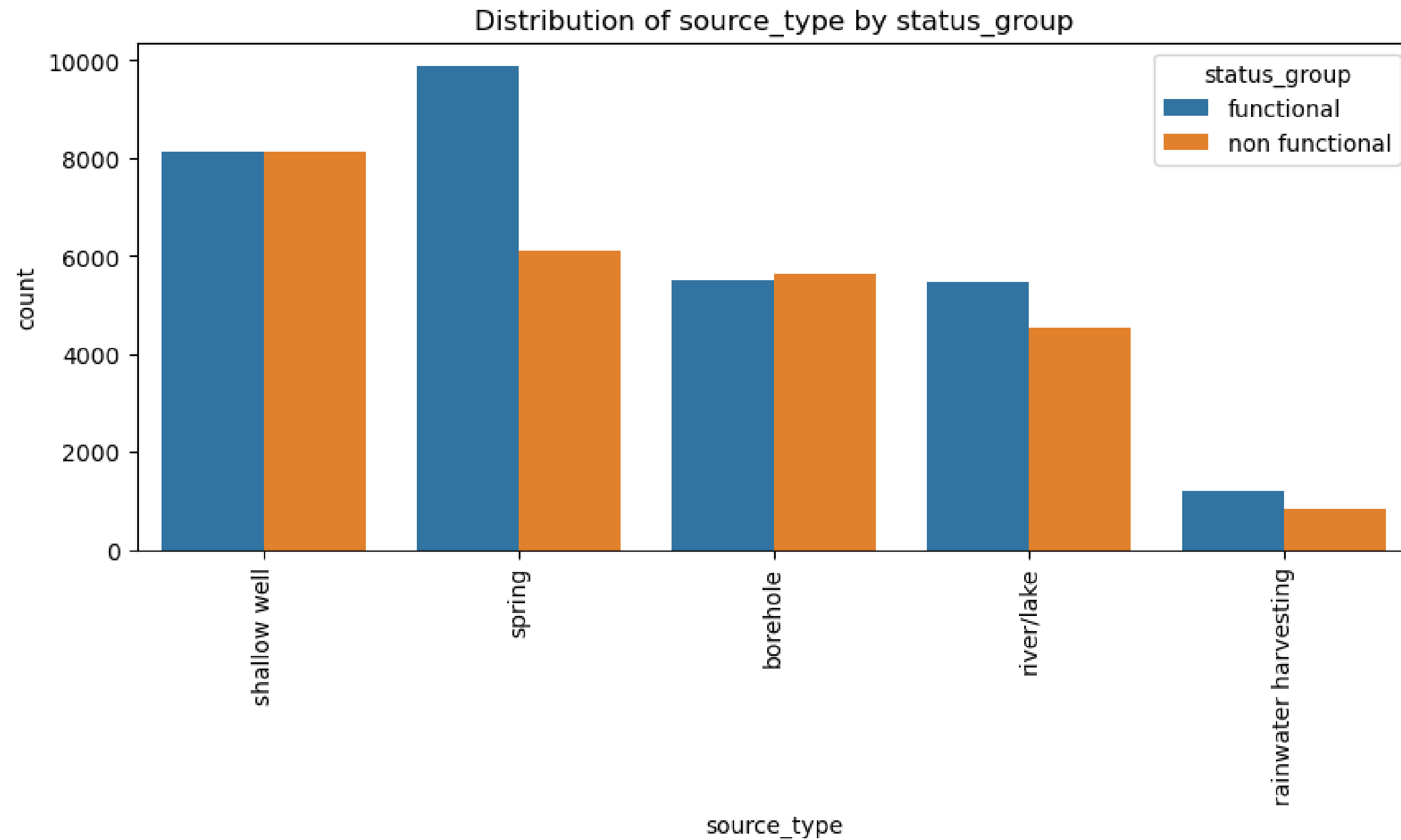
Springs are the most common source of functional wells.

MODELLING

MODELS PERFORMED
- DUMMY CLASSIFIER
- DECISION TREES
- RANDOM TREES CLASSIFIER
- SVM
- K- NEAREST- NEIGHBORS
- DECISION TREE WITH GRID SEARCH
- K- NEAREST NEIGHBORS GRID SEARCH
- RANDOM FOREST WITH GRID SEARCH

THE BEST PERFORMING MODEL WAS THE RANDOM FOREST CLASSIFIER WITH AN ACCURACY OF 78% AND KNN WITH GRID SEARCH WITH AN ACCURACY OF 77%

CONCLUSION

The accuracy of my random forest classifier and K-Nearest Neighbors Classifier was 78% and 77% respectively. While it is still a good predictive model, I would like to undertake further feature engineering to boost this accuracy score if I had more time. I achieved my objectives to be able to predict the functional wells and had a conclusive accuracy score.

RECOMMENDATION

The Lake Rukwa basin area, where there are considerably more non-functional wells than functional ones, may be worth considering for our stakeholders when they decide to build more wells in Tanzania. There are more non-functional wells than functioning ones in the Dodoma region; this situation has to be investigated. Over the course of time, wells with operating permission typically become more viable and useful than those without. Due to possible public misuse, unpaid wells frequently become inoperable; perhaps creating a reasonable payment plan will help stop this. Because wells without permissions also have a higher likelihood of becoming inoperable, our stakeholder must verify that they have permits to ensure that they are acceptable.