TANZANIA WELLS

PROJECT

CONTENT.

01 PROJECT OBJECTIVE.

02 PROJECT OVERVIEW

03 BACKGROUND.

04 DATA AND DISCOVERY.

05 MODELS USED.

06 MODEL PERFORMANCE.

07 RECOMMENDATION..

08 THANK YOU.

PROJECT OBJECTIVE.

This project develops a classifier to improve clean water access for Tanzania's population of over 59 million by predicting water well conditions.

The project aims to improve access to clean water and ensure a reliable supply by prioritizing maintenance and repairs.

PROJECT OVERVIEW

IN THE PROJECT, A MULTICLASS CLASSIFICATION PROBLEM WAS ADDRESSED, AND THE MODEL'S PERFORMANCE WAS EVALUATED USING PRECISION, RECALL, F1-SCORE, AND ACCURACY METRICS.

THE PROJECT FOCUSED ON THREE CLASSES: CLASS 0, CLASS 1, AND CLASS 2.

BACKGROUND.

- Tanzania is a country in East Africa with a population of over 64 million people
- Over 16 million Tanzanians live below the poverty line, unable to meet their basic needs. That is 25.7 percent of Tanzanians.

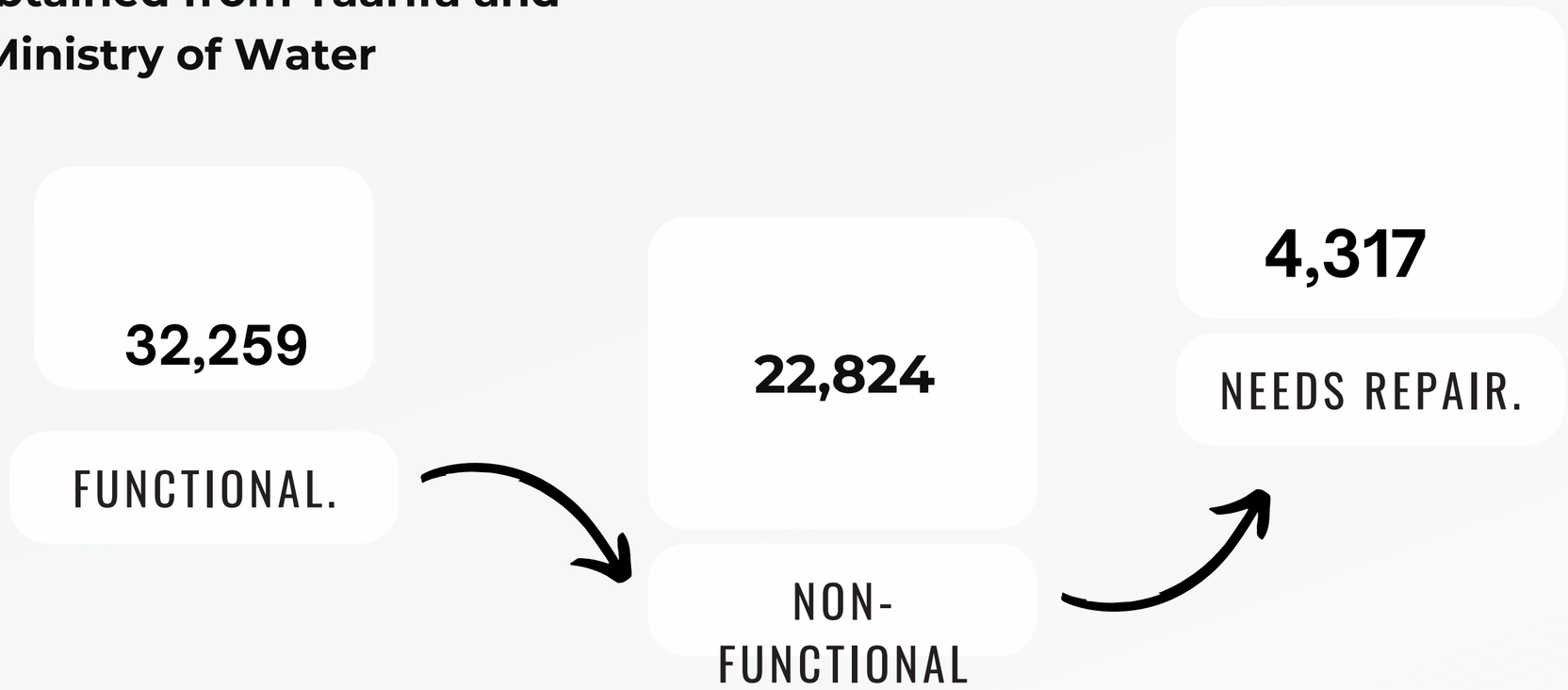

• 39% of Tanzanians do not have access to improved water supply
• 68% of Tanzanians do not have access to improved sanitation.
• 52% of Tanzanians do not have access to basic hygiene.
• These numbers are based on data from the World Bank. The data is from 2020.

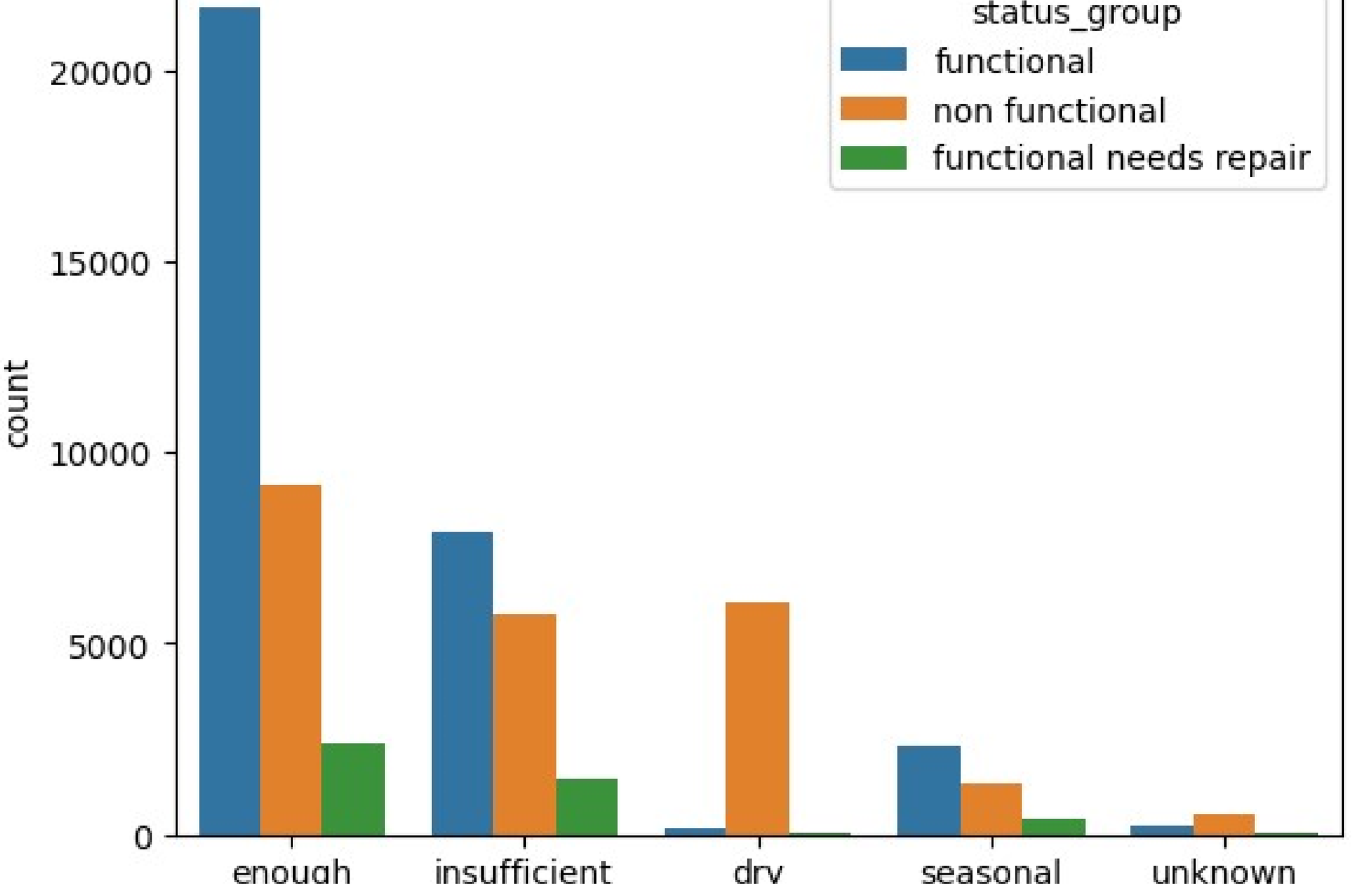
Business Problem

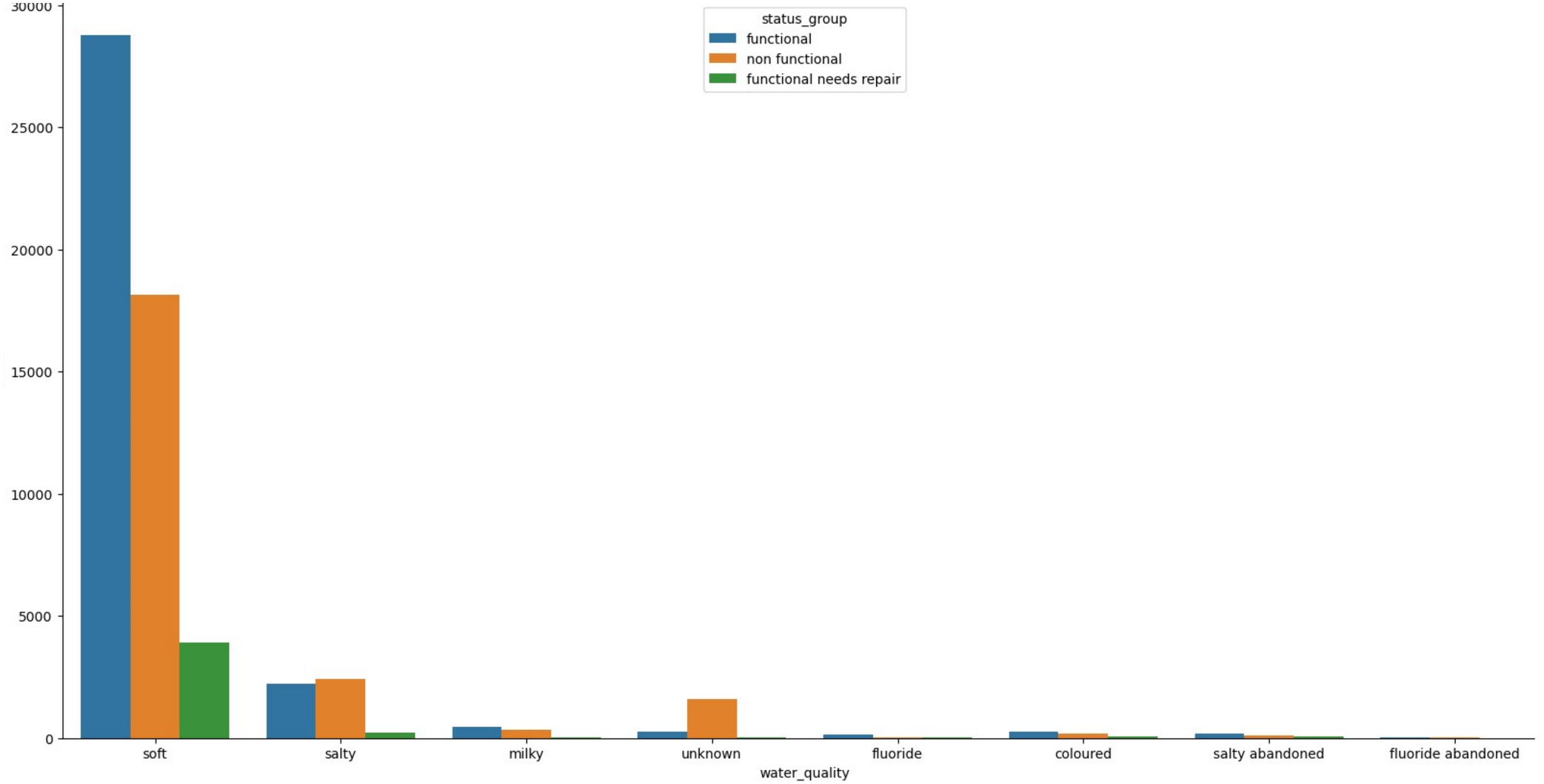
Limited access to clean water in Tanzania requires targeted maintenance and repair efforts for waterpoints. The project aims to develop a classifier to predict water well conditions, ensuring a more reliable water supply.

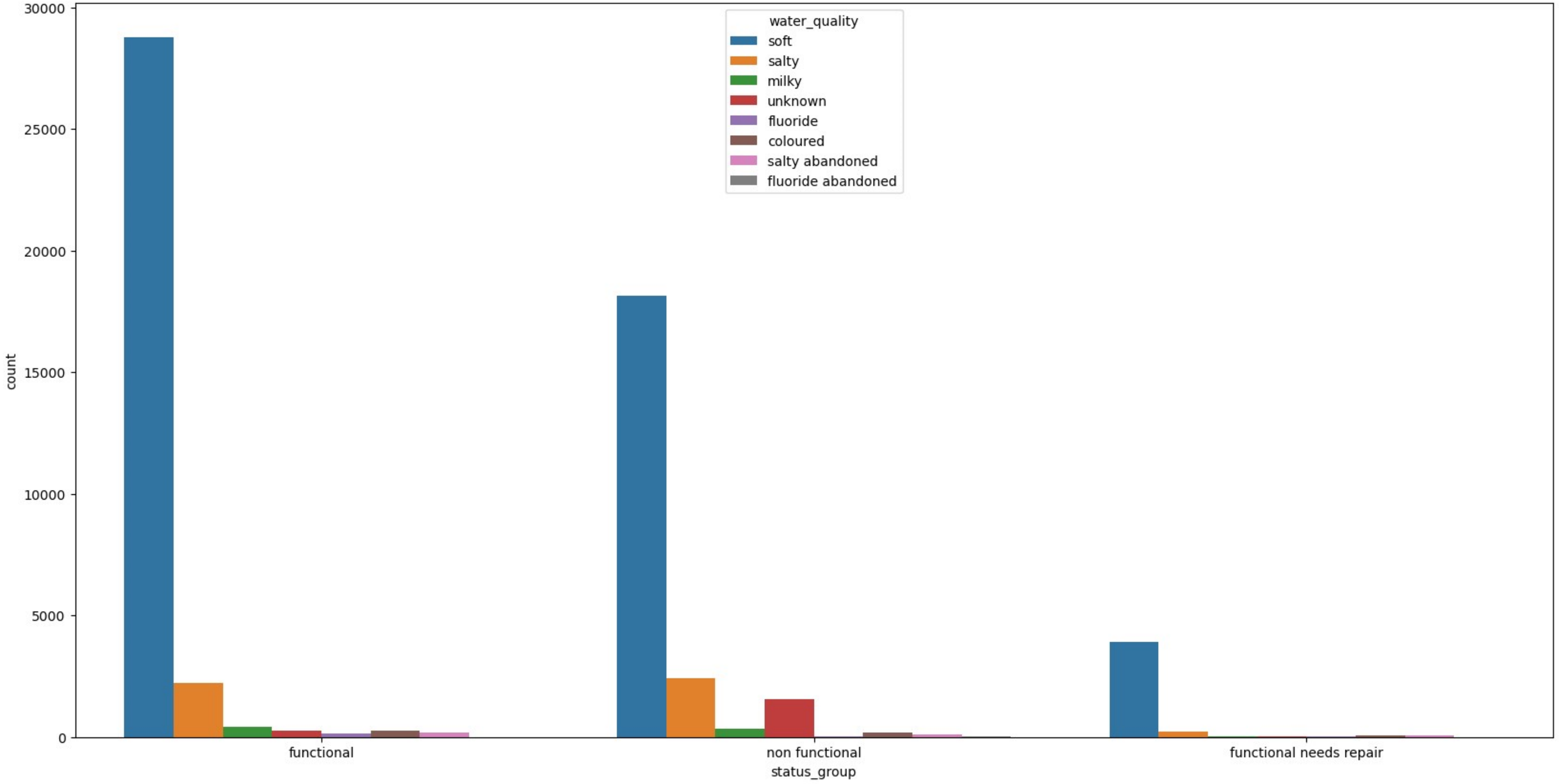
DATA.

- **DrivenData Competition**
- **Data set obtained from Taarifa and Tanzania Ministry of Water**




status_group
functional
non functional
functional needs repair
count
20000
15000
10000
5000
0
enough
insufficient
dry
seasonal
unknown

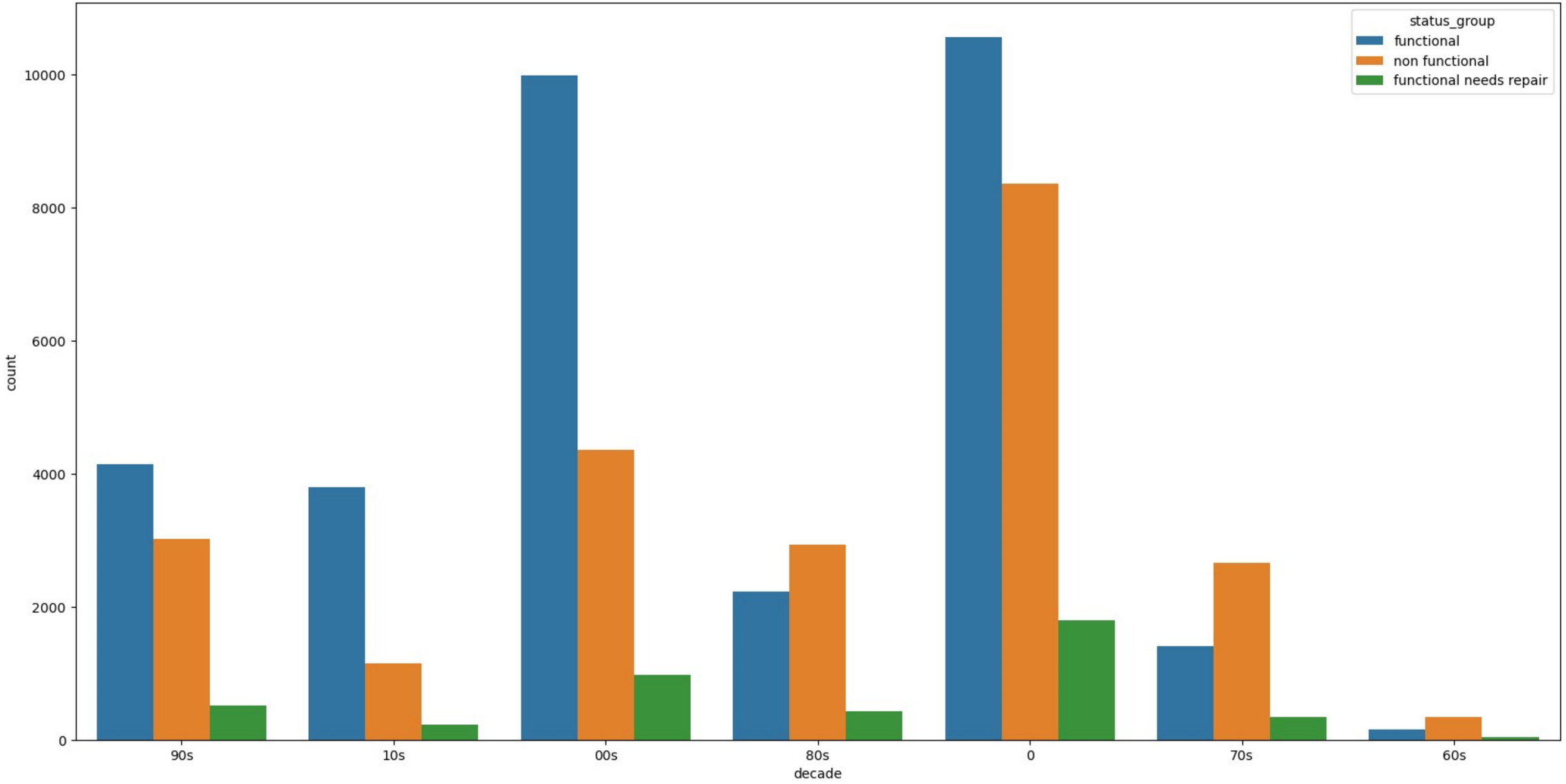

status_group
functional
non functional
functional needs repair
30000
25000
20000
15000
10000
5000
0
soft
salty
milky
unknown
fluoride
coloured
salty abandoned
fluoride abandoned
water_quality

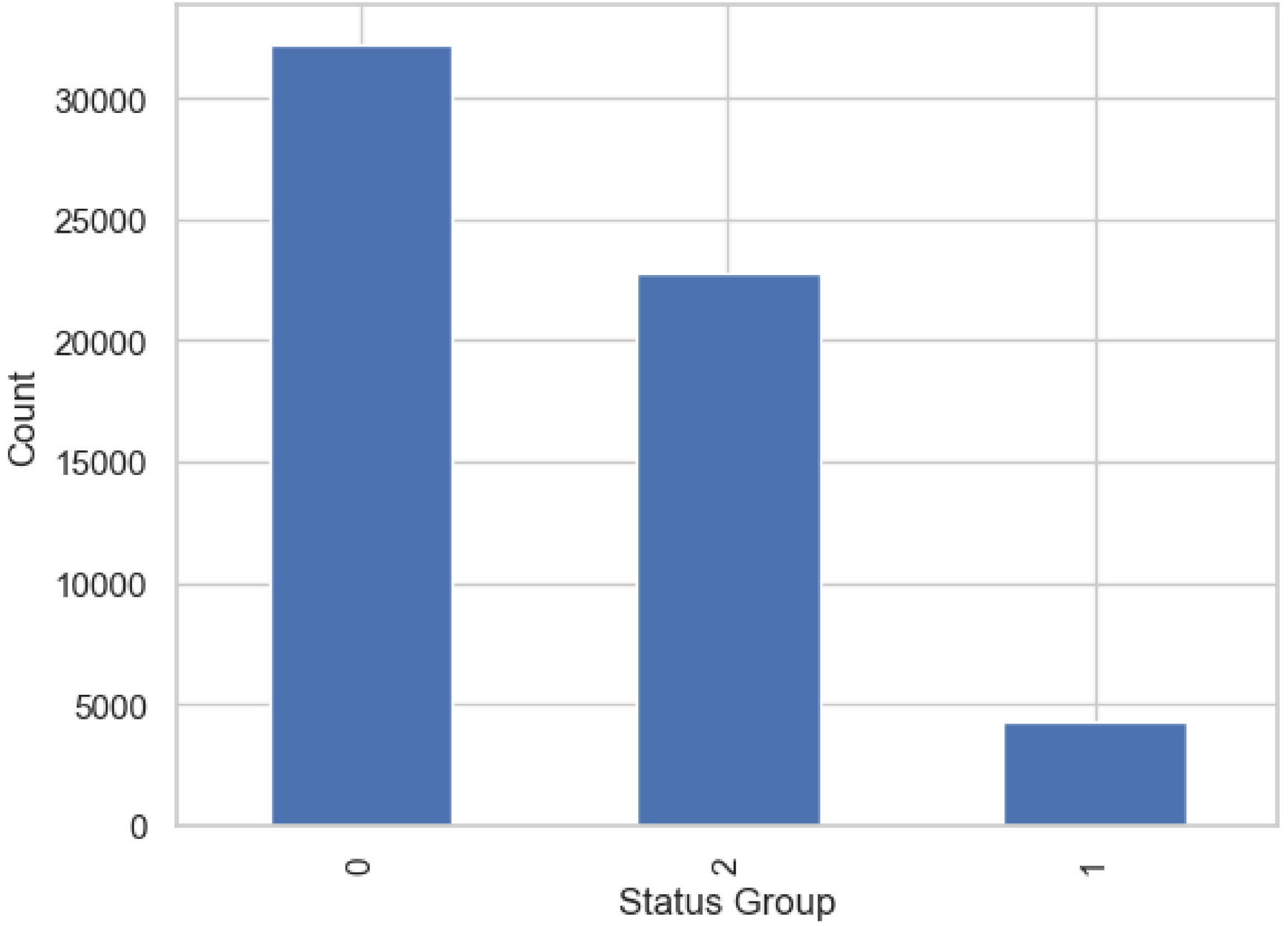

water_quality
soft
salty
milky
unknown
fluoride
coloured
salty abandoned
fluoride abandoned
30000
25000
20000
15000
10000
5000
0
count
functional
non functional
functional needs repair
status_group

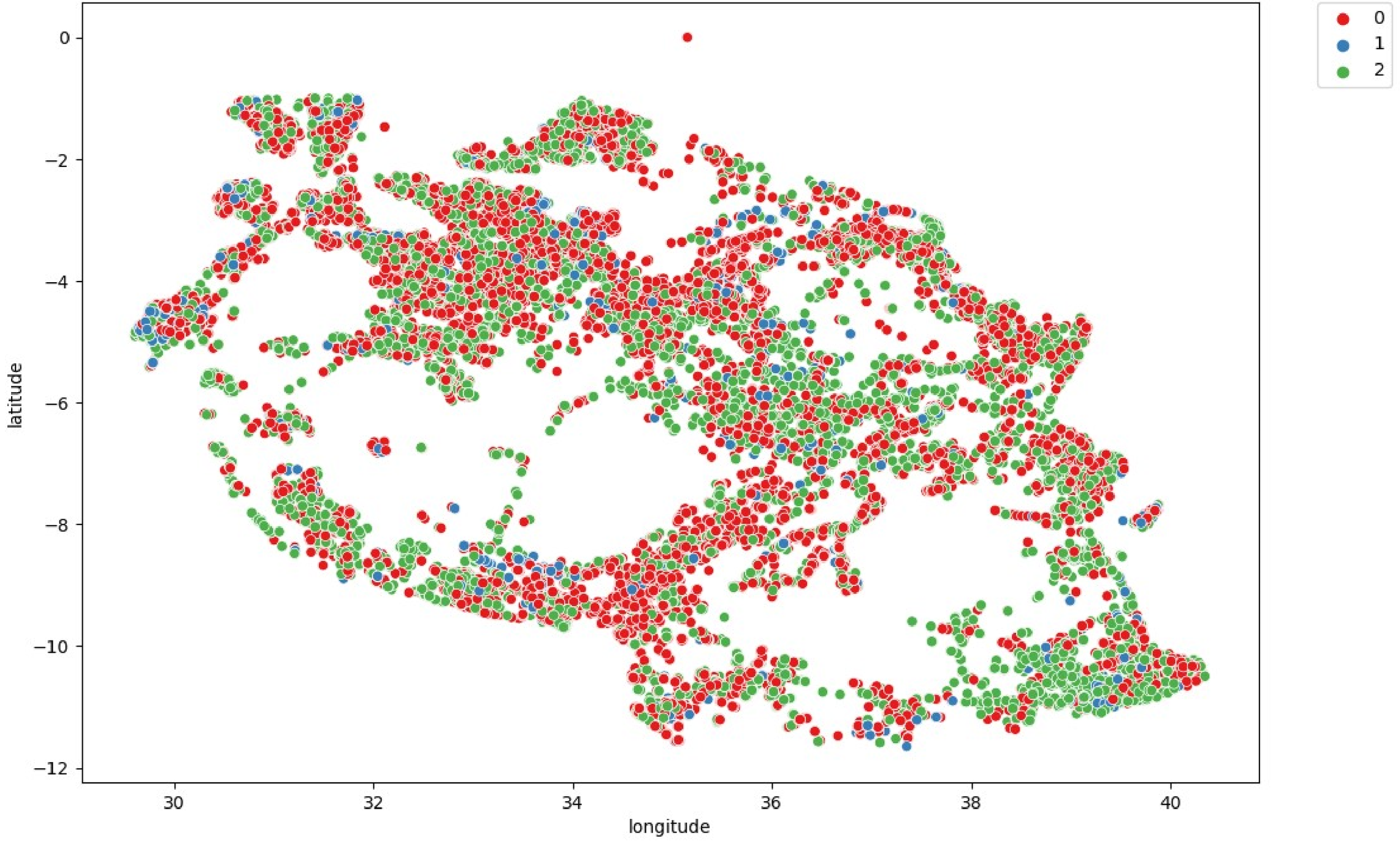
0 = Functional Water points

1 = Functional but needs repair

2 = Non-Functional Water points


status_group
functional
non functional
functional needs repair
10000
8000
6000
4000
2000
0
count
90s
10s
00s
80s
0
70s
60s
decade


30000
25000
20000
15000
10000
5000
0
Count
0
2
1
Status Group


0
1
2
0
−2
−4
−6
−8
−10
−12
latitude
30
32
34
36
38
40
longitude

Models used.

01

BASELINE.

The model had an efficiency of 75%.

02

LOGISTIC REGRESSION

It had an efficiency of 83%.

03

DECISION TREE.

It had an efficiency of 83%.

04

RANDOM FOREST.

It had precision of 76%.

MODEL PERFORMANCE.

- Among the classes, Class 0 showed the best overall performance, with high precision, recall, and F1-score

- Class 1 had more challenges in accurate predictions. Class 2 also performed well. Overall, the model achieved a moderate accuracy of 0.76 in predicting the classes.

RANDOM FOREST

The classification model used was Random Forest, and it achieved a moderate overall performance with an accuracy of 76%. However, there is still a need to enhance the model's ability to correctly classify the minority classes in order to make it more reliable and effective for practical applications.

RECOMMENDATIONS.

1.Improve Class 1: Explore reasons for lower precision, recall, and F1-score. Gather more data or apply targeted techniques to address challenges associated with this class.

2.Analyze Class 1: Conduct further analysis to understand patterns and features contributing to difficulty in accurate predictions. Gain insights for potential improvements.

3.We recommned people to use our model to predict the clssses of the wells in Tanzania.

THANK YOU.