TANZANIA WATER PROJECT

Sally Kinyanjui.

PROJECT OUTLINE

DATA UNDERSTANDING

DATA PREPARATION

EXPLORATORY DATA ANALYSIS

MODEL SELECTION AND TRAINING

MODEL EVALUATION

RECOMMENDATIONS AND CONCLUSION

BUSINESS UNDERSTANDING

Tanzania is known to be a well developing nation with a population of approximaltely 60 milion people.Millions of people in Tanzania, struggle to get clean water and are forced to cover very long distances in search of clean water. Wells are the main source of water for most Tanzanians,however many are broken or in bad shape. This shortage of water is a major issue as water is an essential in various sectors,leading to poor health,slow economic growth and hindered productivity.

This projects with the use of machine learning tools aims at identifying why some wells fail,predict whether the new wells will work .

t

PROBLEM STATEMENT

Water shortage is a crucial problem in Tanzania, and this arises mainly from the maintenance of wells. Maintaining these water sources is also a challenge, maybe because of water quality, geographical location, infrastructure, and many other factors. Some are non-functional, while others are partially functional. This may limit the local citizens from accessing clean water. Challenges faced include health and sanitation issues, poor economic growth, and hindered productivity.

Our project here aims to help those who are in charge of water, be it governmental, non-governmental organizations, and maybe policymakers, to make improved or better decisions that would help in the improvement of the functionality and maintenance of water wells in the country.

OBJECTIVES

Main objective: come up with a prdeictive model that will determine functionality of water wells

Specific objectives:

- Identifying key factors that may contribute to functionality of waterwells in Tanzania.
- Developing a predictive model that can predict functionality of water wells based on attributes such as geographical location,infrastracture and water quality.
- Enable data-driven decision making.
- provide recommendations that can boost on the efforts for maintenance

DATA UNDERSTANDING

HTTPS://WWW.DRIVENDATAORG/COMPETITIONS/7/PUMP-IT-UP-DATA-MINING-THE-WATER-TABLE/DATA/

HAS 41 COLUMNS AND 59400 ENTRIES

OUR CSVS ARE: 1. TEST SET VALUES
2. TRAINING SET LABELS
3. TRAINING SET VALUES

THE LABELS IN THIS DATASET ARE SIMPLE. THERE ARE THREE POSSIBLE VALUES:

1. FUNCTIONAL - THE WATERPOINT IS OPERATIONAL AND THERE ARE NO REPAIRS NEEDED

2. FUNCTIONAL NEEDS REPAIR - THE WATERPOINT IS OPERATIONAL, BUT NEEDS REPAIRS
3. NON FUNCTIONAL

DATA PREPARATION AND CLEANING

1.Checked for duplicates

2.dropped duplicates

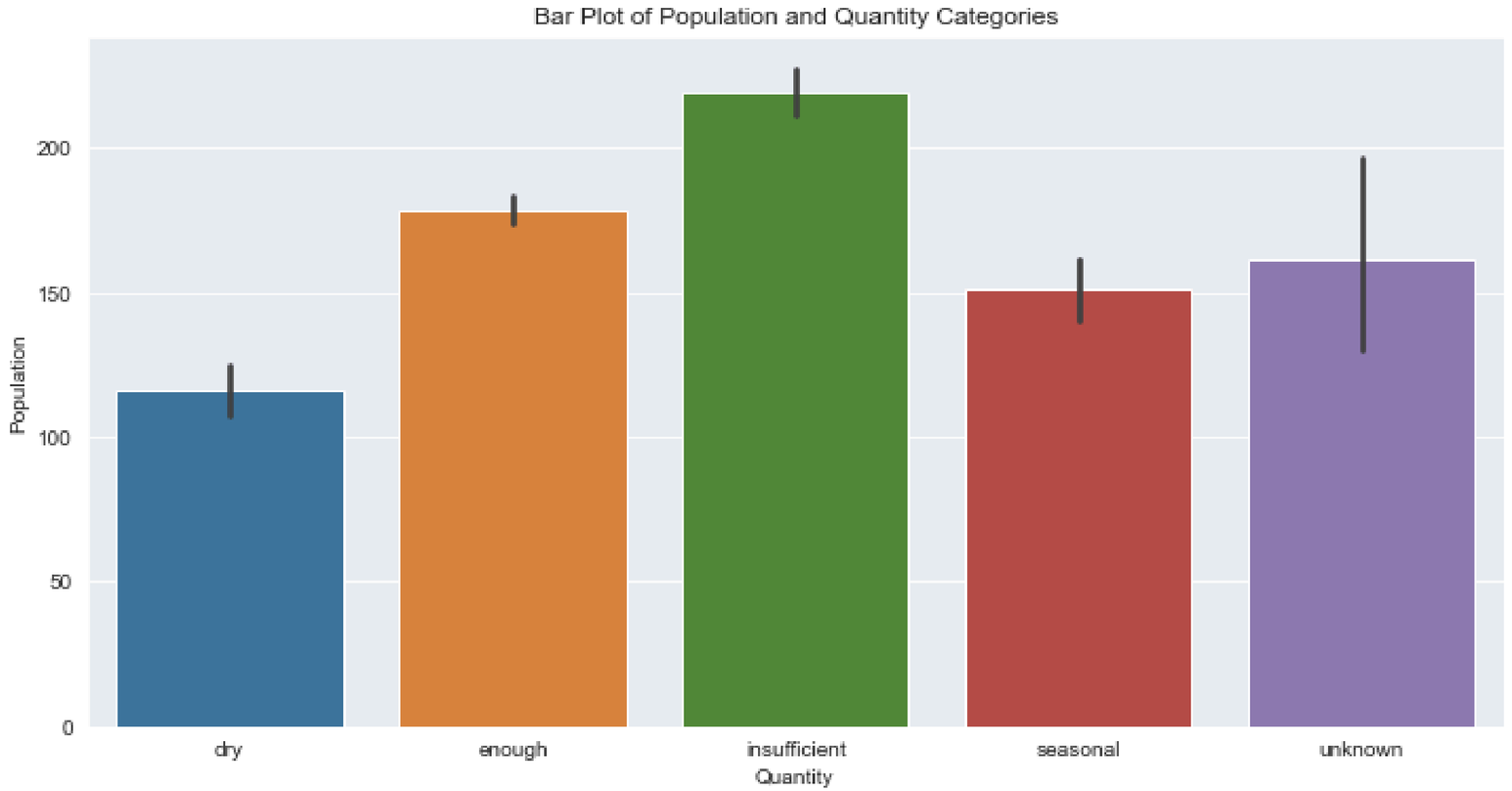
3.checked missing values

4.replaced missing values with mode

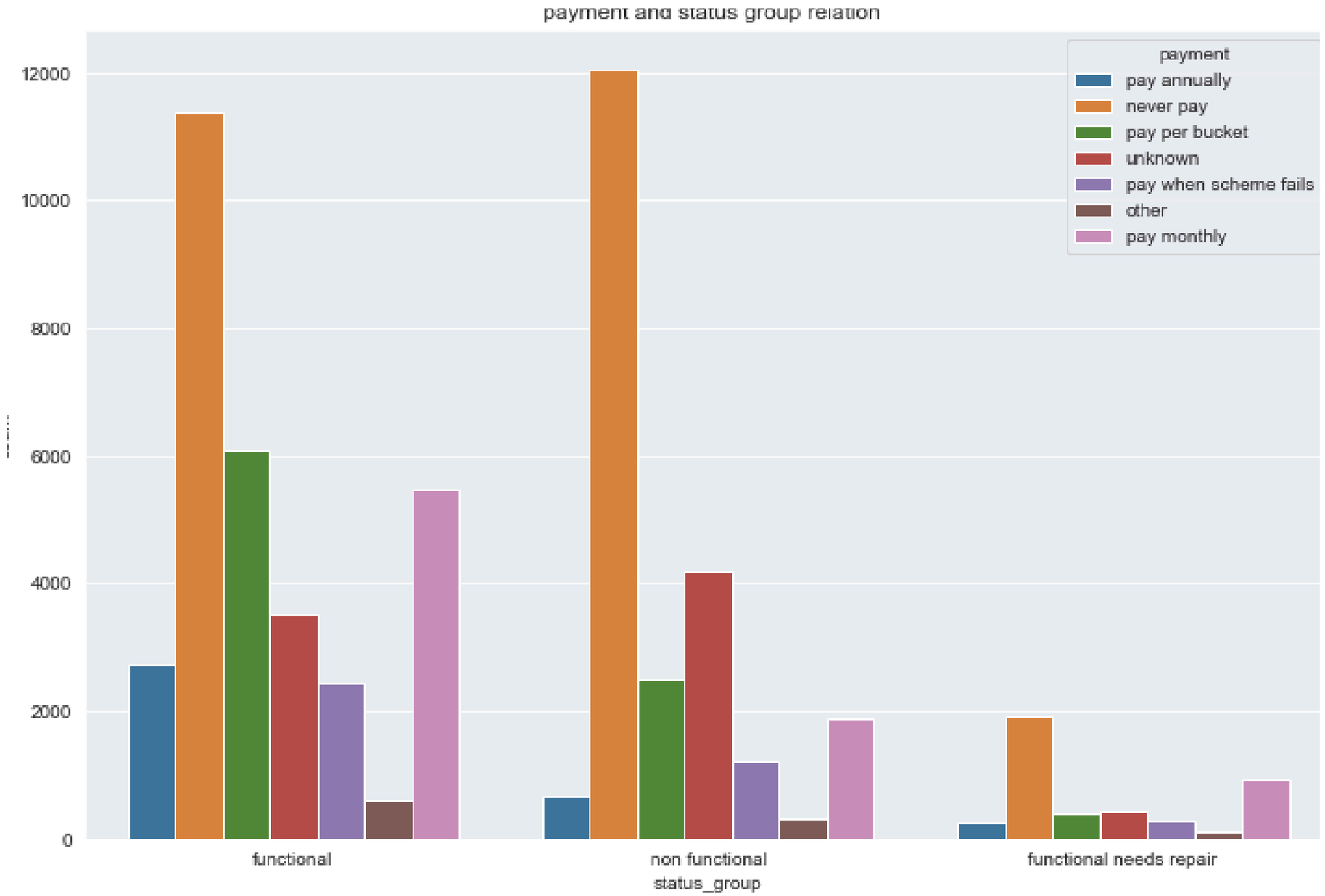
5.dropped columns

EXPLORATORY DATA ANALYSIS

Relationship between population and quantity



Relationship between payment and status_group.



From the previous graphical reprsentations ,the first one shows that most wells are financed by the goverment of tanzania. The second one shows that most wells that are non-functional are not paid for.

MODELING.

We used the following models:

1. KNearest Neighbours
2. Random Forest
3. Logistic Regression
4. Decision Tree

EVALUATION

The other models seemed to be overfitting and underfitting. Gradient Boosting was the best model compared to the others as it showed balancing. It had a training accuracy of 63.02%.
It was successful since it met our success metrics.

FINDINGS

1. Iringa is observed to be the region with the highest number of functional water points.
2. Many functional water points seem to be providing insufficent amounts of water.
3. A lot of non-functional water sources are not paid for.
4. Factors that contribute functionality of the water sources include:Region, source type and water quality.
5. The functional water points wer the most populated.
6. Many wells are found in areas that are highly populated.

RECOMMENDATIONS.

1.Top financiers should be approached to fund for repairs and also add more water wells.

2.Government ofTanzania should impose a rule where it will be mandatory for them to pay.

THANK YOU.