

The background of the slide is a light blue gradient. It is decorated with several realistic water droplets of various sizes, some with highlights and shadows, giving them a 3D appearance. The droplets are scattered across the slide, with a higher concentration in the top left and bottom right corners.

GROUP 12 MEMBERS:

- 1. VIVIAN MAIYO**
- 2. GEORGE TIDO**
- 3. CAROLINE KISAULU**
- 4. LEWIS GITARI**
- 5. CAROLINE NGABU**
- 6. FAITH GITAU**

THE TANZANIAN WATER WELLS

IS
THIS WELL
FUNCTIONAL
?





OVERVIEW

- The objective of this project is to develop a machine learning classification model that accurately predicts the functionality of water wells in tanzania.
- By doing so, we aim to improve access to clean ad reliable water sources by identifying wells that are functional, non functional and those that need repair.
- In our modeling, we shall use tanzanian wells data from drivendata labs that we shall use to train and validate our model to predict functionality status of water wells.
- We shall make data driven recommendations to stakeholders and policy makers on where attention is needed to ensure water crisis in tanzania is resolved.

OUTLINE

Business Problem

Data and Methods

Findings

Conclusion

Recommendations

BUSINESS PROBLEM

- 39% of Tanzanians do not have access to basic water supply.
- Lack of reliable infrastructure for monitoring and maintaining water wells has led to inefficiencies in resource allocation and delays in addressing non functional wells.
- Traditional approaches in management of wells have been ineffective as installed water wells cease to function over time while others remain in dilapidated conditions needing repair for a long time.
- This project aims to come up with a predictive classification model that identifies functionality status of wells in Tanzania and directs stakeholders and policy makers attention to where attention is needed in a proactive approach.

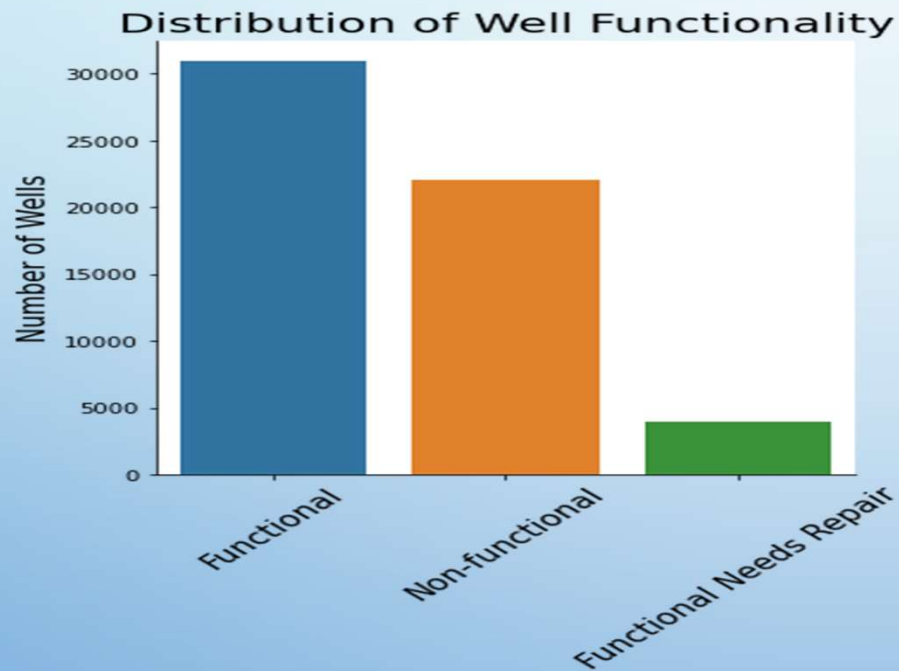
DATA AND METHODS

- Tanzanian water wells data set.
- Contains 39 features describing wells such as location, quality, quantity, funder, installer, management e.t.c
- Data cleaning to address data issues such as missing values, data types e.t.c
- Exploratory data analysis
- Feature engineering
- Classification modeling

The background of the slide features a light blue to white gradient. It is decorated with several realistic water droplets of varying sizes, some with highlights and shadows, positioned in the top-left, top-right, and bottom-right corners.

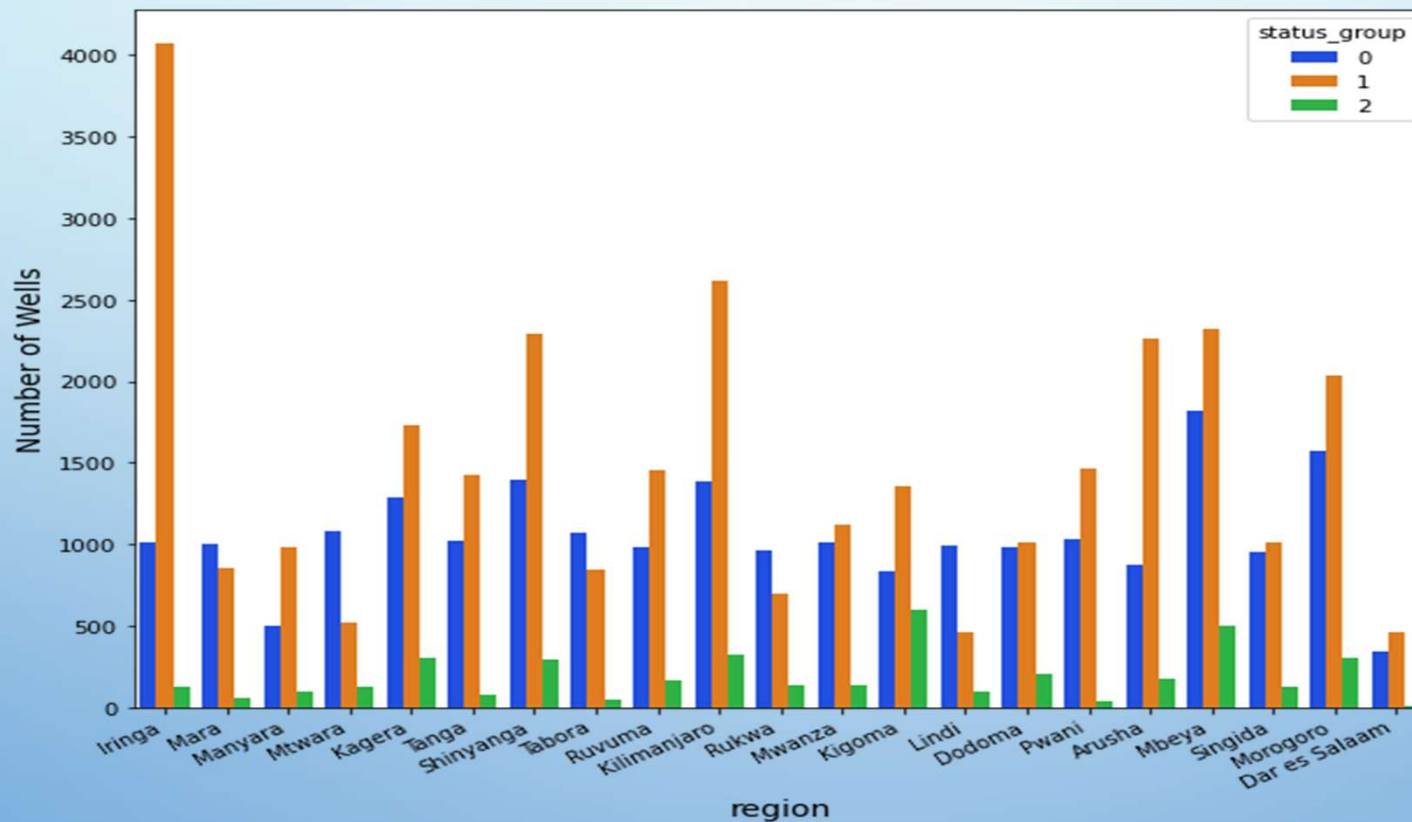
FINDINGS

OVERALL WELL FUNCTIONALITY



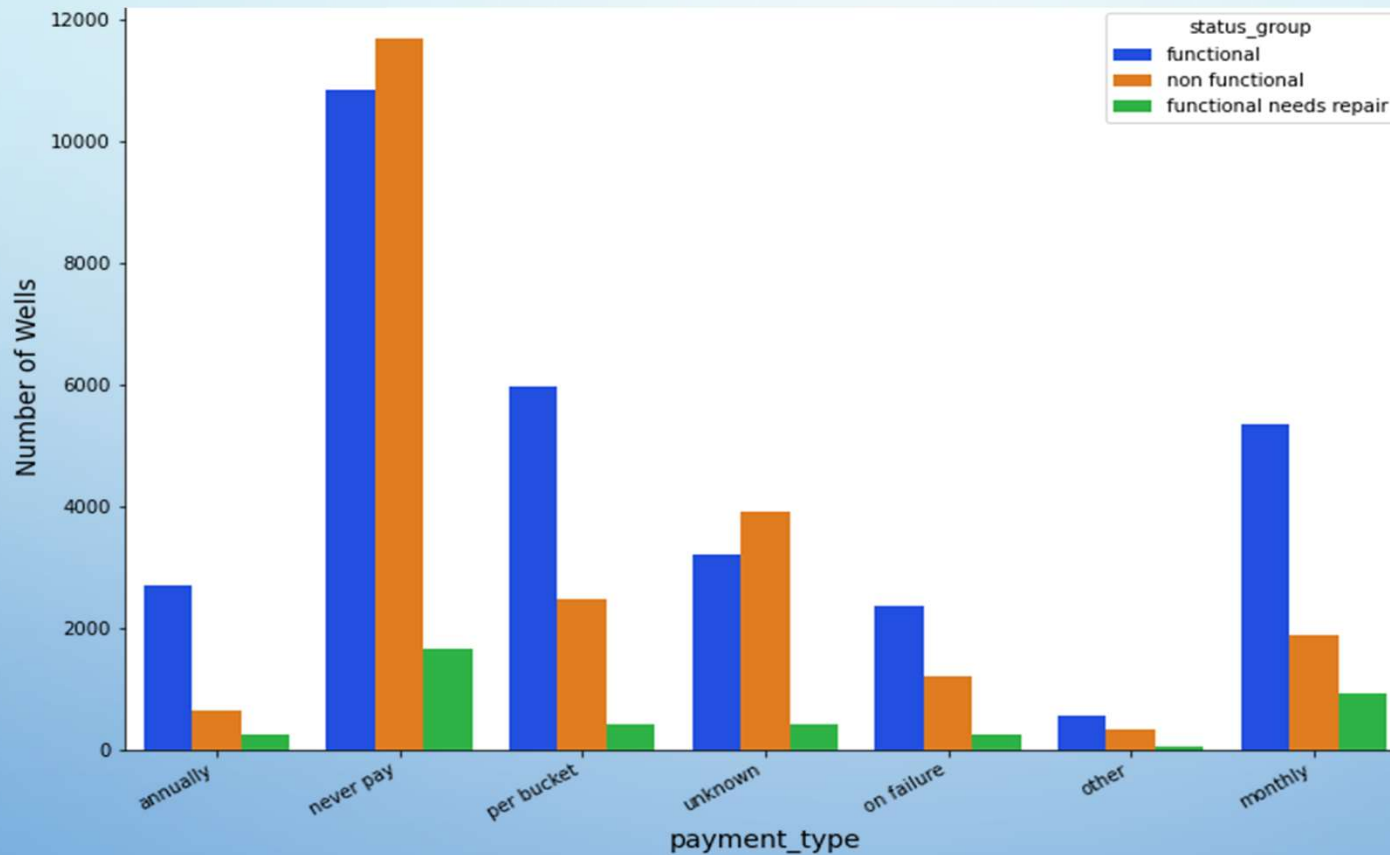
- More than half of the wells are functional.
- Minority of the wells are functional but need repair.
- A significant proportion of wells are non-functional.

WELL FUNCTIONALITY PER REGION



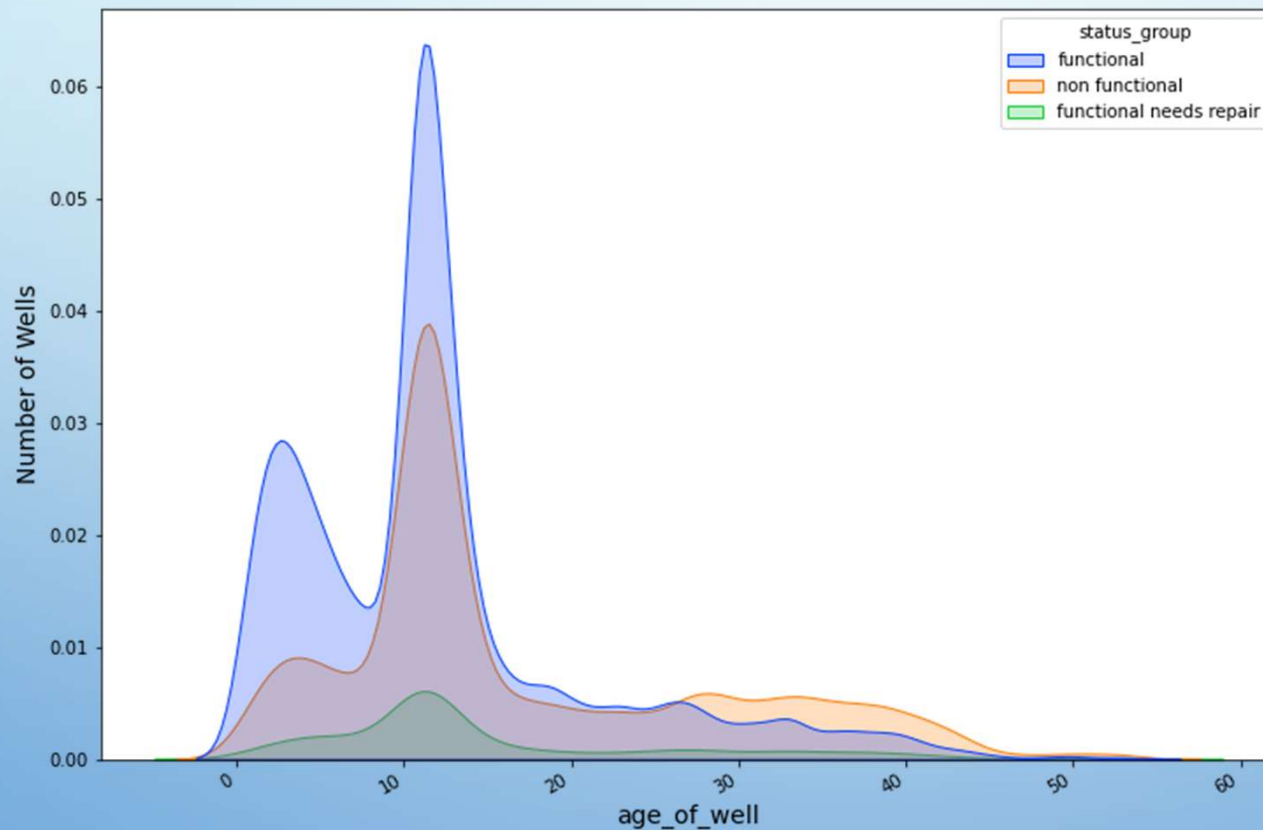
- Mtwara region and lindi region stand out for having a higher proportion of non functional wells.

WELL FUNCTIONALITY VS PAYMENT



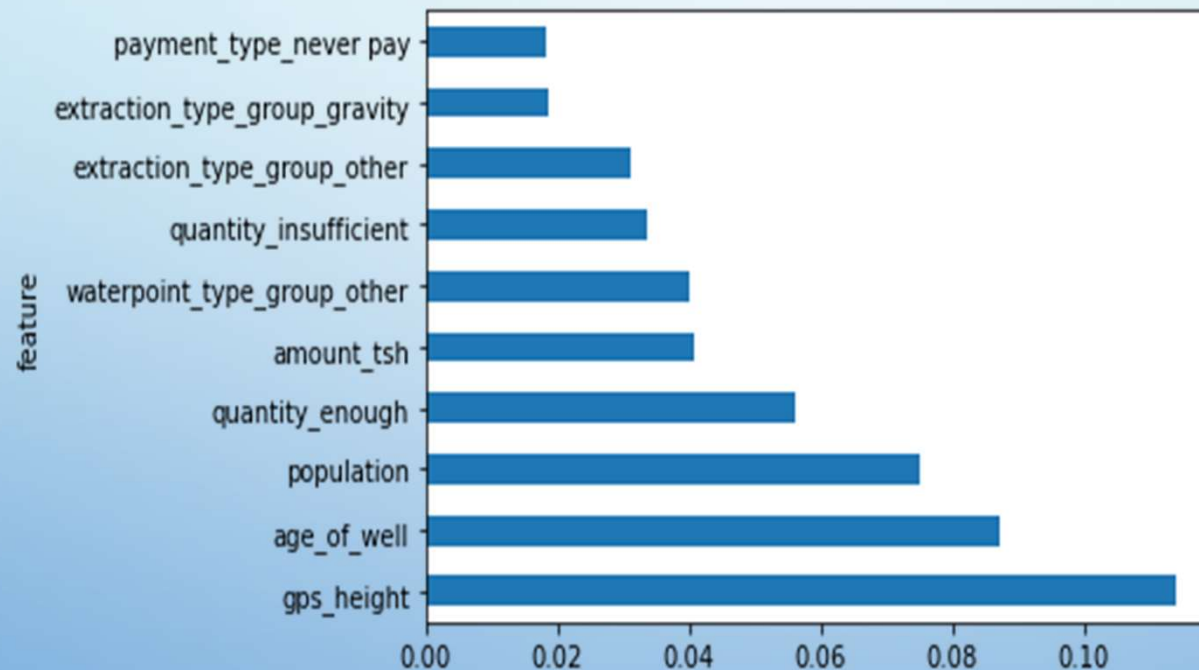
- There is high proportion of non-functioning wells where there is no payment compared to wells where there is payment

FUNCTIONALITY BY AGE



- There is a higher proportion of functional wells between 0 and 15 years.
- The proportion of non functional wells significantly increase from 30 years of age onwards.

FEATURE IMPORTANCE



- These are the top ten important features in predicting well functionality.
- Altitude of the well influences the functionality of the well.

RESULTS

- There is a high proportion of non functional wells accounting for 45% of the total wells.
- Well projects funded by the Government of Tanzania and the Ministry of Water are more likely to be non functional.
- Mtwara, Lindi, Ruvuma/Southern Coast and L. Rukwa basin regions have a high proportion of non functional water wells which can be prioritized by stakeholders for repair.
- Free well projects and where public meetings are not held, have a higher proportion of non functional wells compared to the functional wells.

RESULTS

- Well projects that are older have a higher proportion of non functional wells than new ones(0 - 15 years). Older wells should be monitored more closely as they are more likely to fail than newly installed wells.
- Dry wells, wells with salty water and wells with dam and lake as a source are more likely to be non-functional.

CONCLUSION

- One of the main causes of water crisis faced by Tanzanian population has majorly been occasioned by the high number of non functional/dilapidated wells across the regions.
- Stakeholders should adopt our predictive model to monitor the wells being built to ensure they remain functional and prompt quick action incase of a failure.

RECOMMENDATIONS

- Stakeholders should focus to repair the high number of non functional wells and those that need repair to improve water access by the people of Tanzania.
- Government authorities and policy makers should focus on the top ten features highlighted when installing new wells to improve functionality of wells.
- Stakeholders should adopt our model as a proactive approach to predict the functionality status of the wells for faster resolution as opposed to traditional methods of examination and reporting of wells.

FURTHER ANALYSIS

1. **Further hyperparameter tuning:** Hyperparameter tuning can be explored further using Gridsearch to further improve on the models used in prediction.
2. **Exploring more model types** such as XGboost, Adaboosting, naive bayes, svm and model stacking that were not used here can be explored to find a better performing model.
3. **More information on data features** some symbols and abbreviations used in the data are not universal. More information on the data can enhance the data cleaning process especially in the funders and installers features to enhance the cleaning process and consequently the model outputs.

The background of the slide features a light blue to white gradient. It is decorated with several realistic water droplets of varying sizes, some with highlights and shadows, positioned in the corners and along the bottom edge.

THANK YOU.

Q&A