

Data Modeling Strategy for Tanzania Water Wells

for MajiSafi Aid and Central Government of Tanzania

By Nesphory Mwakale

Project Analysis Overview

- Descriptive analysis and model construction of status of water points in Tanzania. This analysis can be used to
 - Identify functional, non-functional wells and functional wells that need repair in Tanzania.
 - To prioritize which wells to fix depending on it's condition and it's impact on society.
 - Plan resources and budget accordingly to ensure communities have easier access t o clean and reliable data.

Business Understanding

- MajiSafi Aid and Central Government of Tanzania wants to identify the status of all wells in Tanzania without spending too much resources in physically visiting them.
- This way, they can prioritize and focus their limited resources on non-functional and functional wells that need repair. This all depends on the impact the well has on the community around it.

Data Understanding

- The sources of our data is [DrivenData](#) in collaboration with [Taarifa](#) and [Ministry of Water in Tanzania](#).
- The latest addition into the dataset was in 2015.

DRIVEN**DATA**

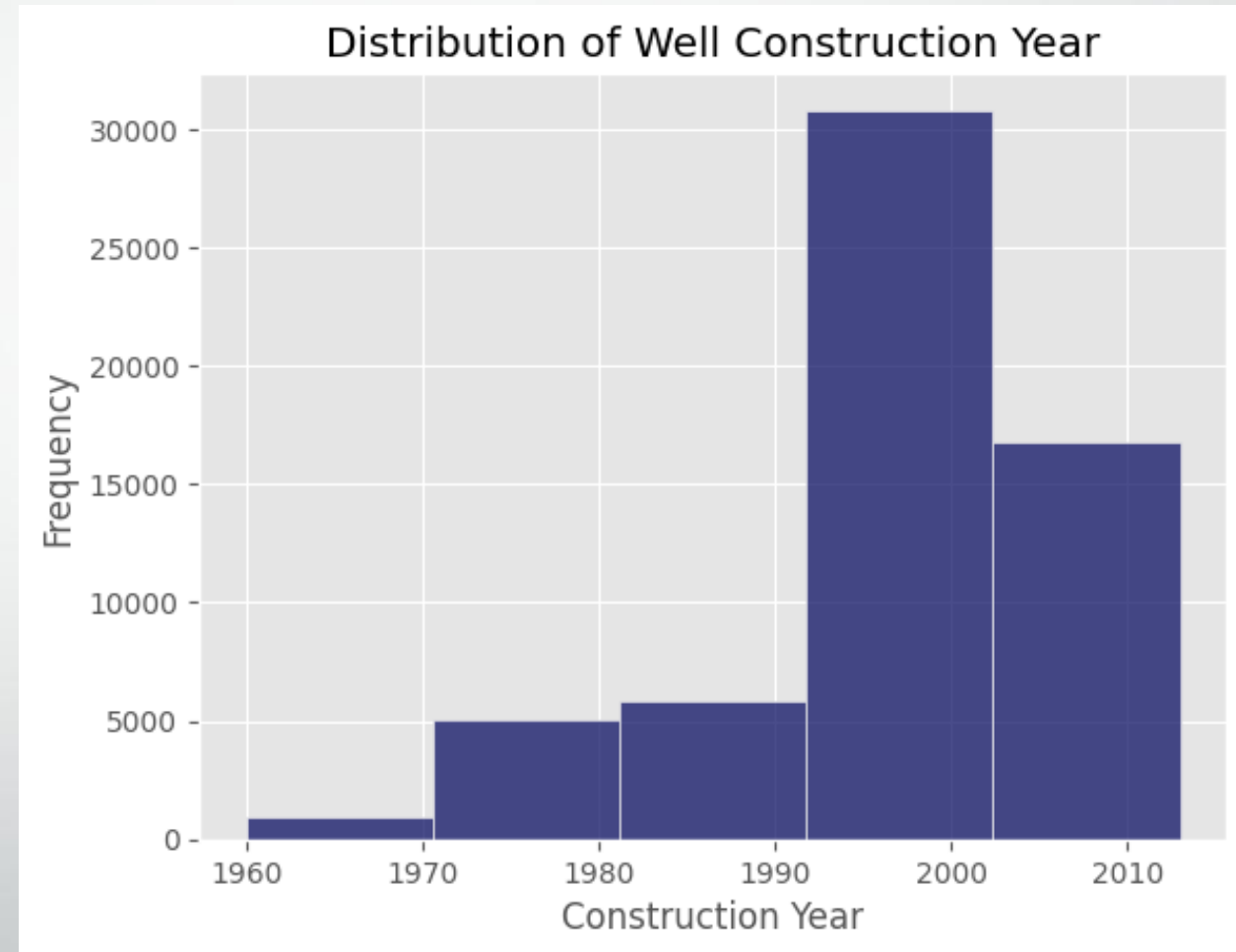




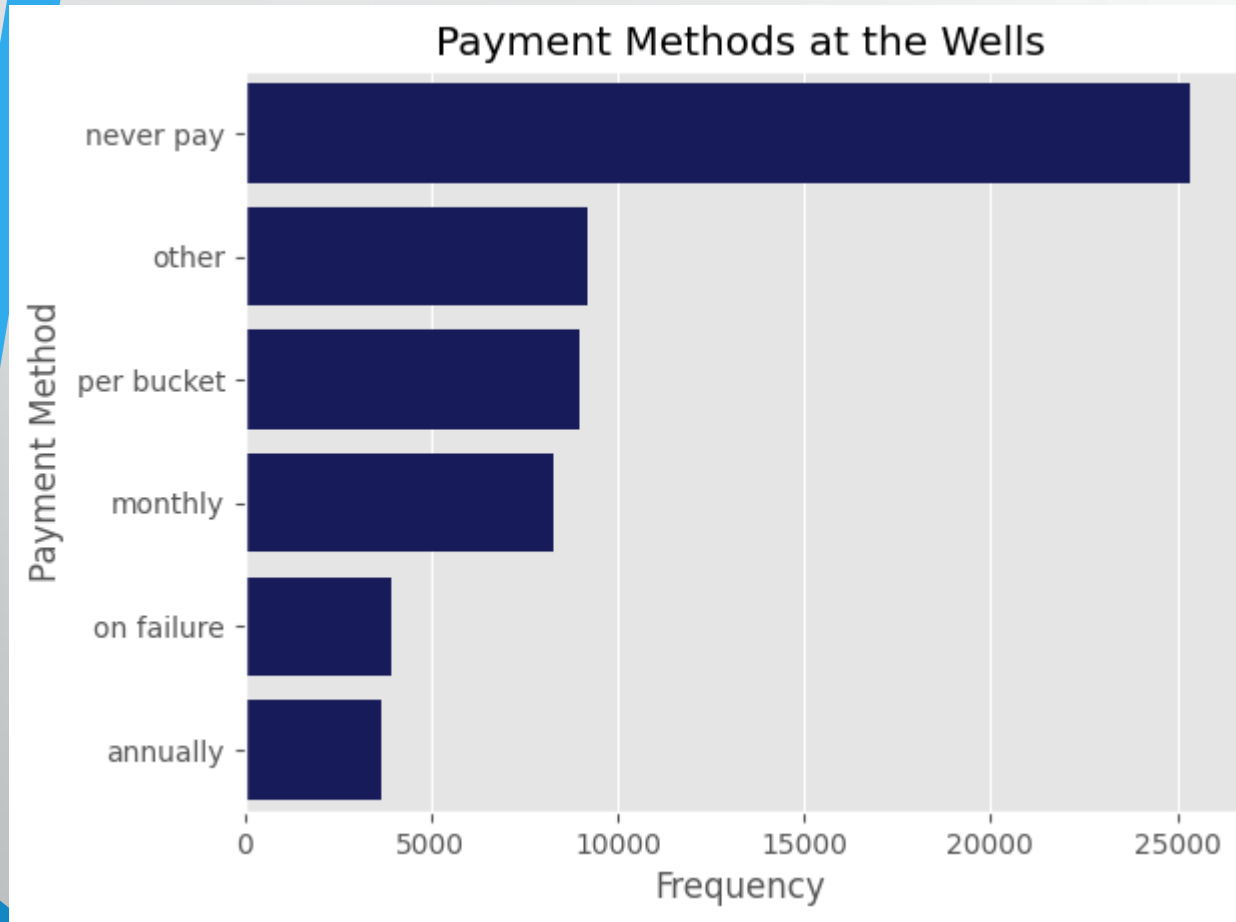
Data Analysis Report

Data Analysis

- Over 77% of the wells were constructed after 1992.
- The oldest well in the dataset was constructed in 1960.



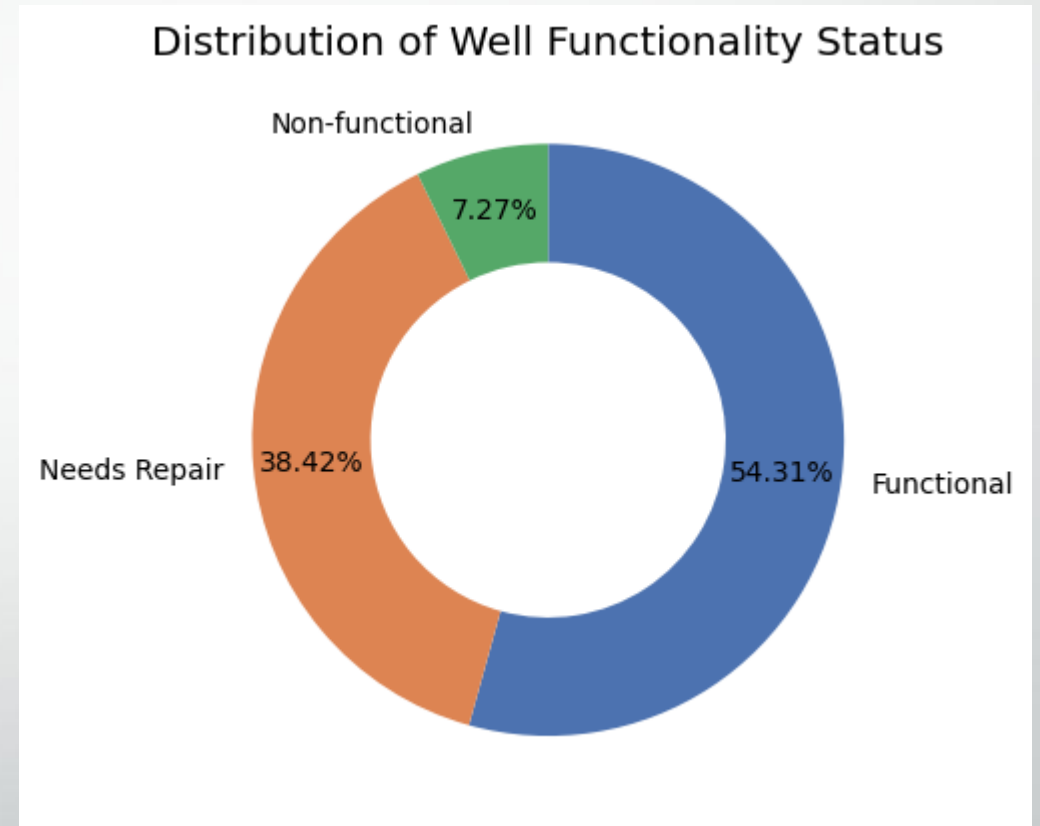
Data Analysis



- The water in almost 25,000 wells is free for public use. No payment is needed to access the water in these wells.

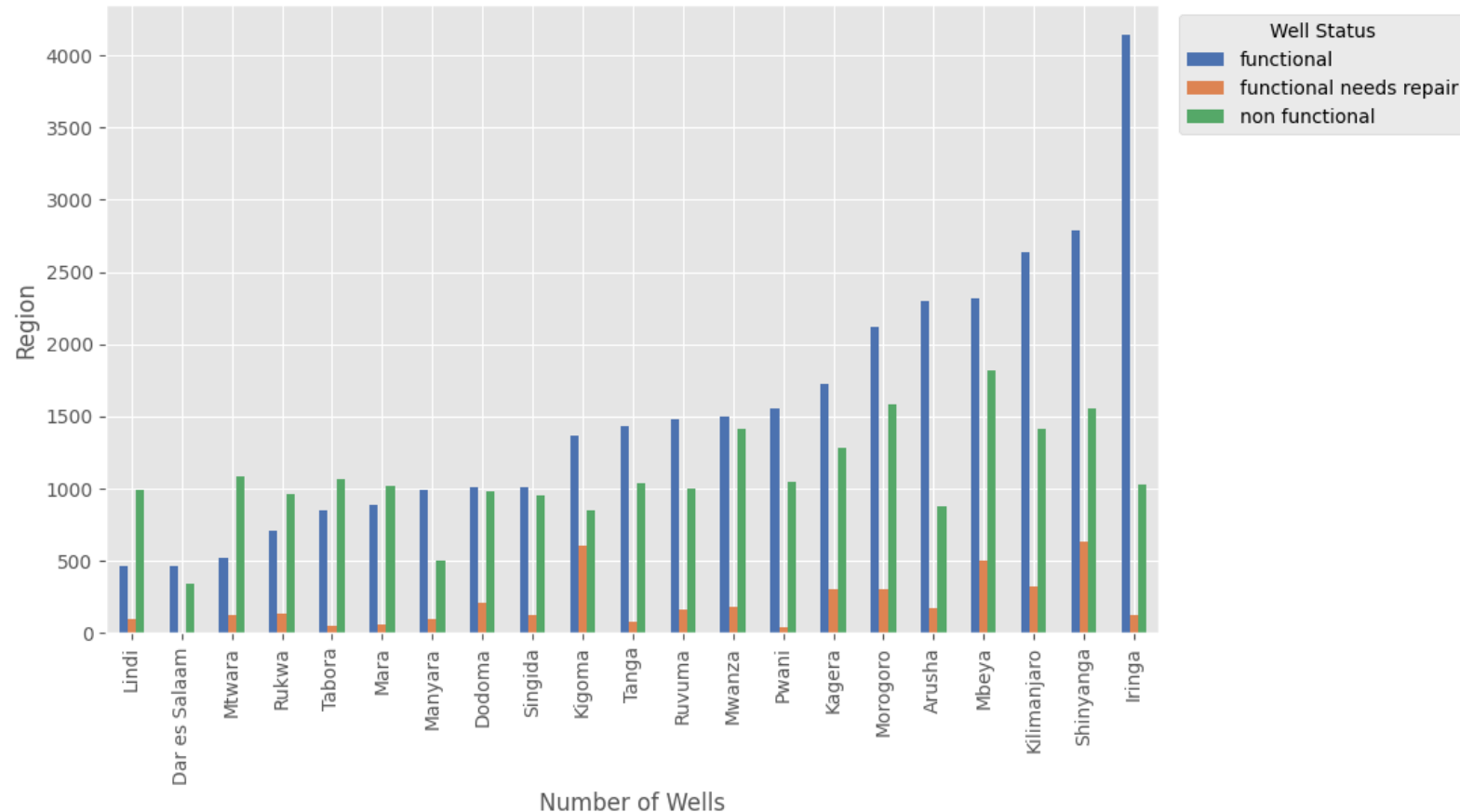
Data Analysis

- 92% of the wells in Tanzania are functional but 38% of them need repair or servicing.
- The other 7% are non-functional.



Data Analysis

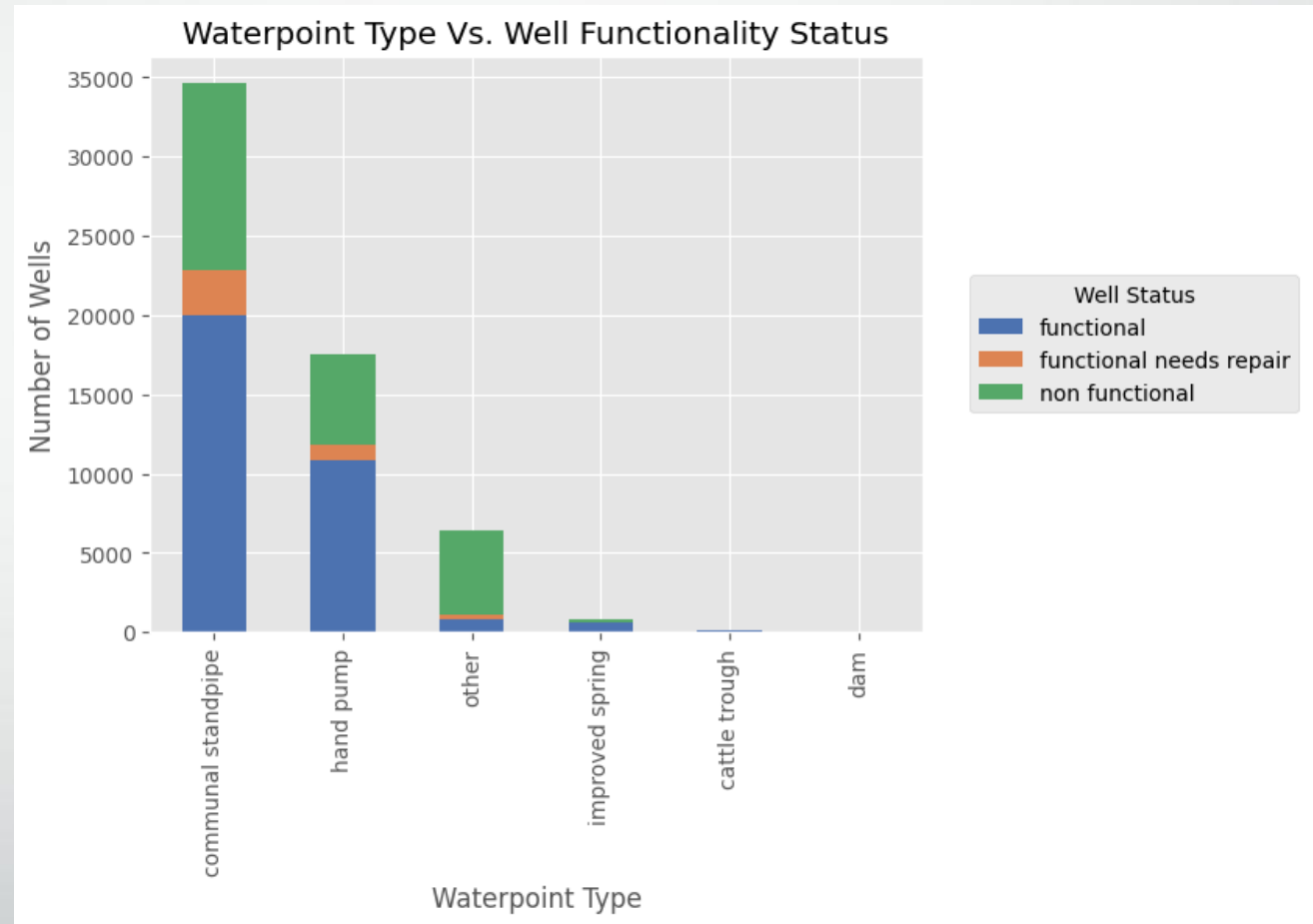
Region Vs. Well Functionality Status



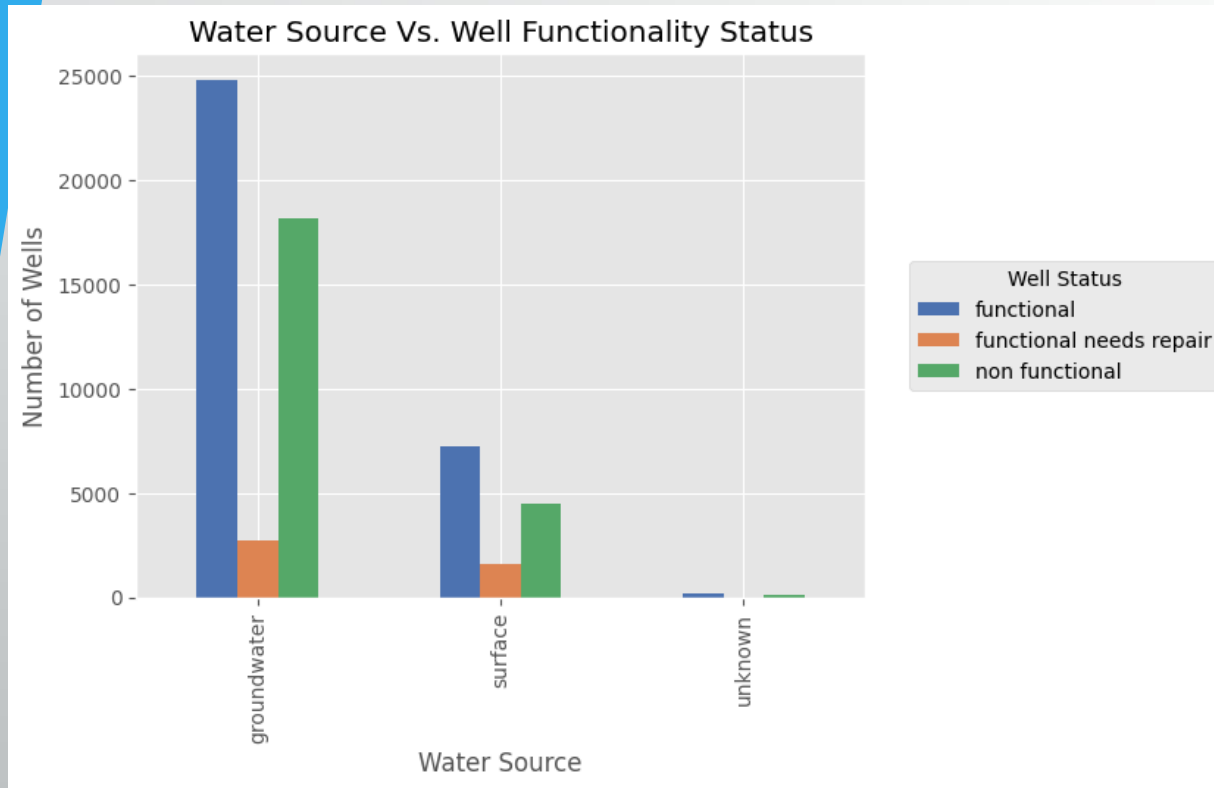
- Iringa region has the most functional wells with around 4,000 functional wells.
- Mbeya, Morogoro and Shinyanga regions have the highest number of non-functional wells.
- Shinyanga and Kigoma have the highest number of functional wells that need repair.

Data Analysis

- Communal standpipes are the most common type of water points with around 35,000 water points using them.
- Very few or no wells use dams and cattle troughs as their water point type.



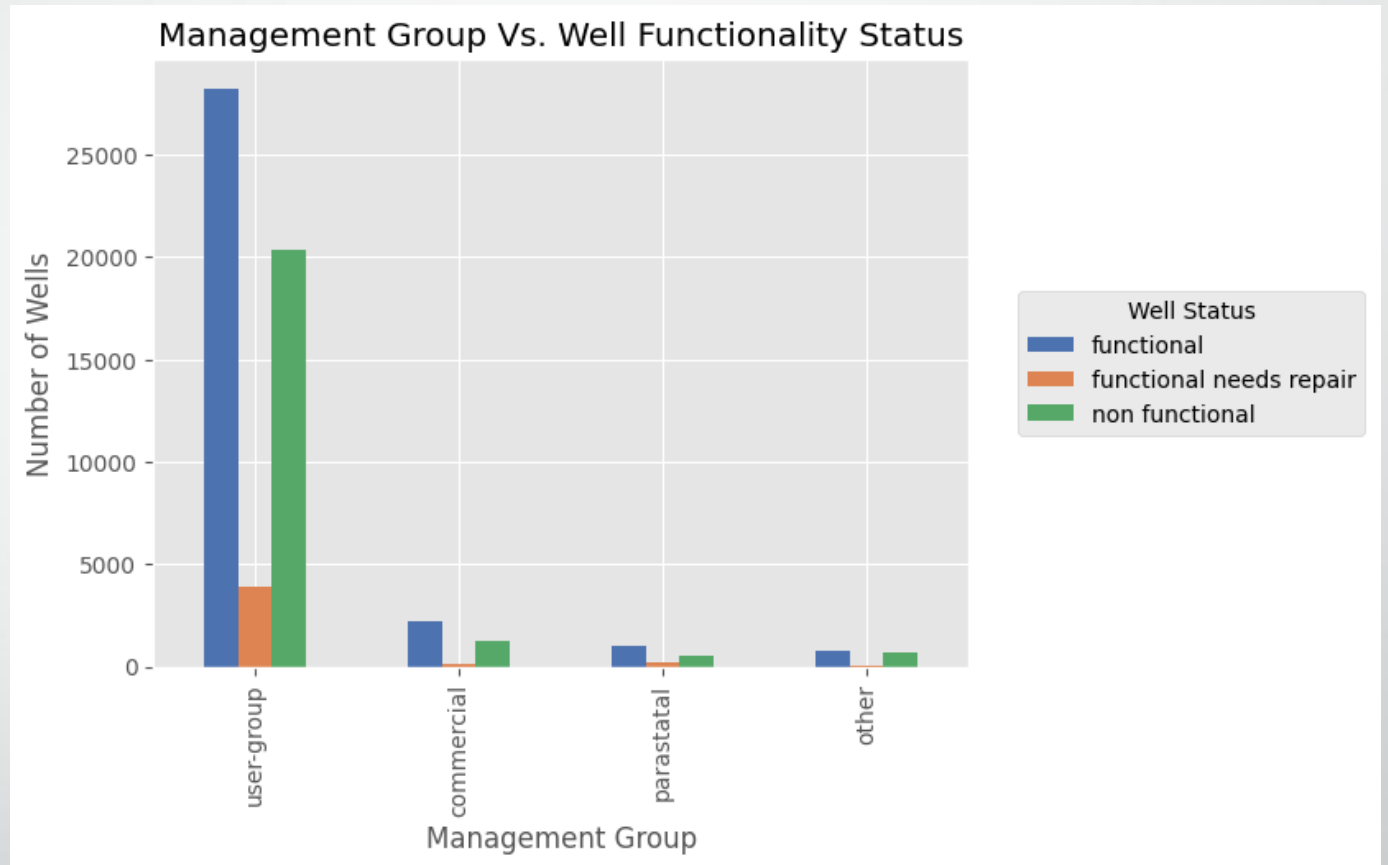
Data Analysis



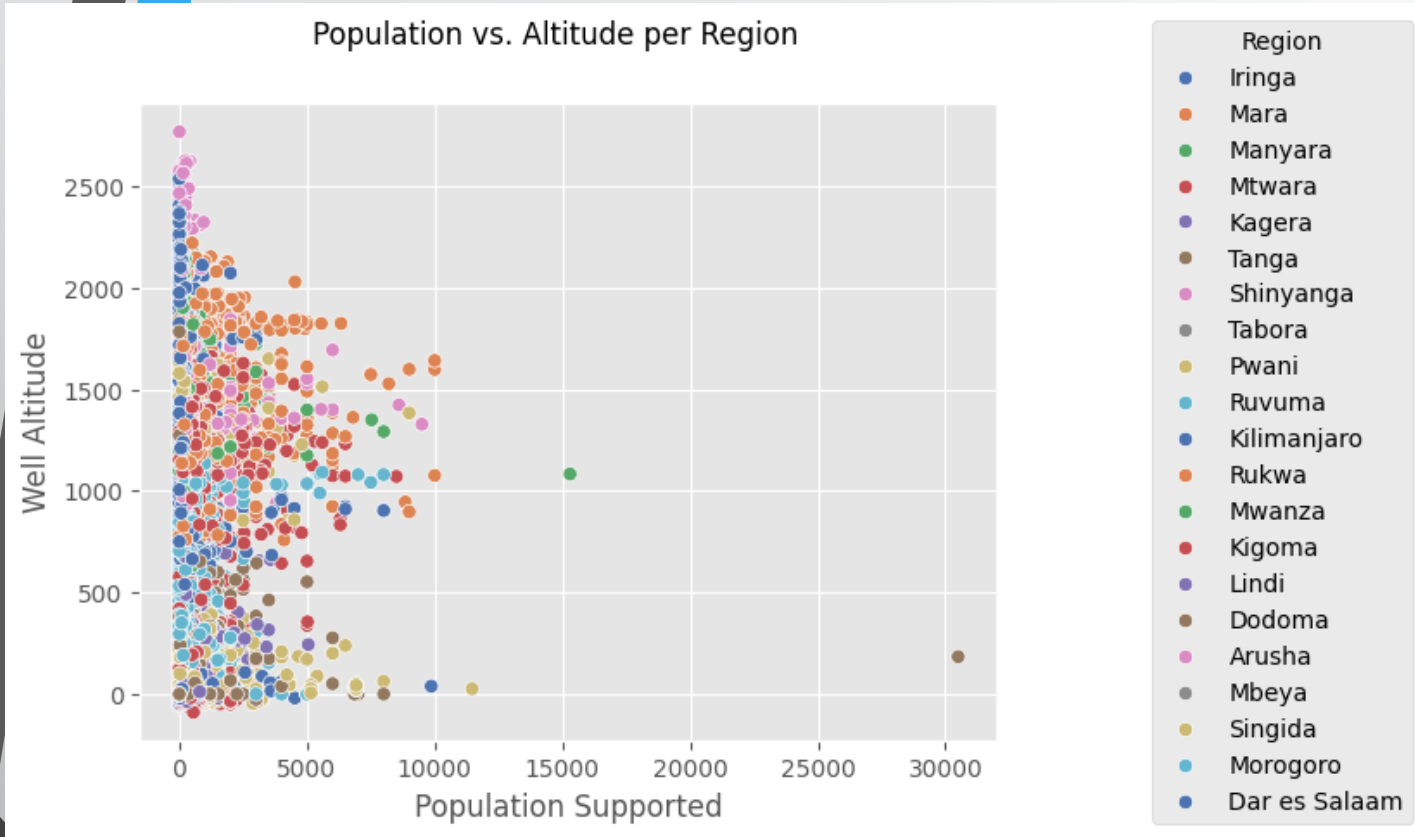
- 76% of all wells depend on groundwater as their water sources.
- Groundwater sources are springs, shallow wells and boreholes.

Data Analysis

- User Groups manage around 87% of all the wells in Tanzania with Village Water Committee (VWC) managing around 60% of those wells.



Data Analysis



- Most water points serve relatively small populations.
- Very few wells support more than 10,000 people.
- There is a weak correlation between altitude and population as the altitude of the well does not affect the population supported by the well.

Model Evaluation

- A classification model was selected for this as our target was to categorize our data into:
 - 0 – Functional wells
 - 1 – Functional wells that need repair
 - 2 – Non-functional wells
- XGBClassifier was our chosen classification model. It gave us the highest accuracy of 74% and correctly identified 69% of the wells.

Conclusion

- 54% of the wells in Tanzania are functional, 7% of them are non-functional and 38% of them are functional but need repair.
- Most wells in Tanzania are found along the coastline as almost 22,500 wells have an altitude of zero.
- Over 30,000 wells produce enough water for the population that they support.
- Gravity is the most common extraction method being used in just over 16,000 functional wells, around 8,000 non-functional wells, and almost 3,000 functional wells that need repair. Wind-powered extraction methods are not very common in the country.
- Around 82% of all wells contain soft water.

Recommendations

- MajiSafi Aid can focus on regions like Lindi, Mtwara, Rukwa, Mara, Mwanza, Singida and Tabora should be focus regions as they have more non-functional wells than functional or almost the same number of functional and non-functional wells.
- Some wells support a population of zero. The Tanzanian government and MajiSafi Aid can focus on building a few deep wells in every region and reservoirs and focus on water transportation pipelines to reduce distance travelled by villagers to get water.
- Water cleaning technology like reverse osmosis and filtration can be established in areas with milky, salty, colored and fluoride water to ensure that everyone gets clean water.

Future Steps

- Our best model was able to correctly identify 69% of all wells. Most models however failed to identify the models that need repair. More data is necessary to help train the data and correctly predict more wells accurately. Especially data on wells that need repair.
- More models can be implemented on this model, especially because that this model was full of categorical data. Further tuning can also be done to improve the accuracy of the models. More experience, of course, is also needed to know where to improve and to increase domain knowledge in the sector.
- Investing in data validation during the data collection stage can help reduce the data errors that were observed in the installer and funder columns of this data set.



Questions



Thank You!

Nesphory Mwakale

LinkedIn: <https://www.linkedin.com/in/nesphory-mwakale/>

GitHub link: https://github.com/M-Nesphory/Tanzania_Water_Wells_Modeling_Project.git