May 2020

Tanzania Water Wells Project

By Vidya Menon

Points to Consider

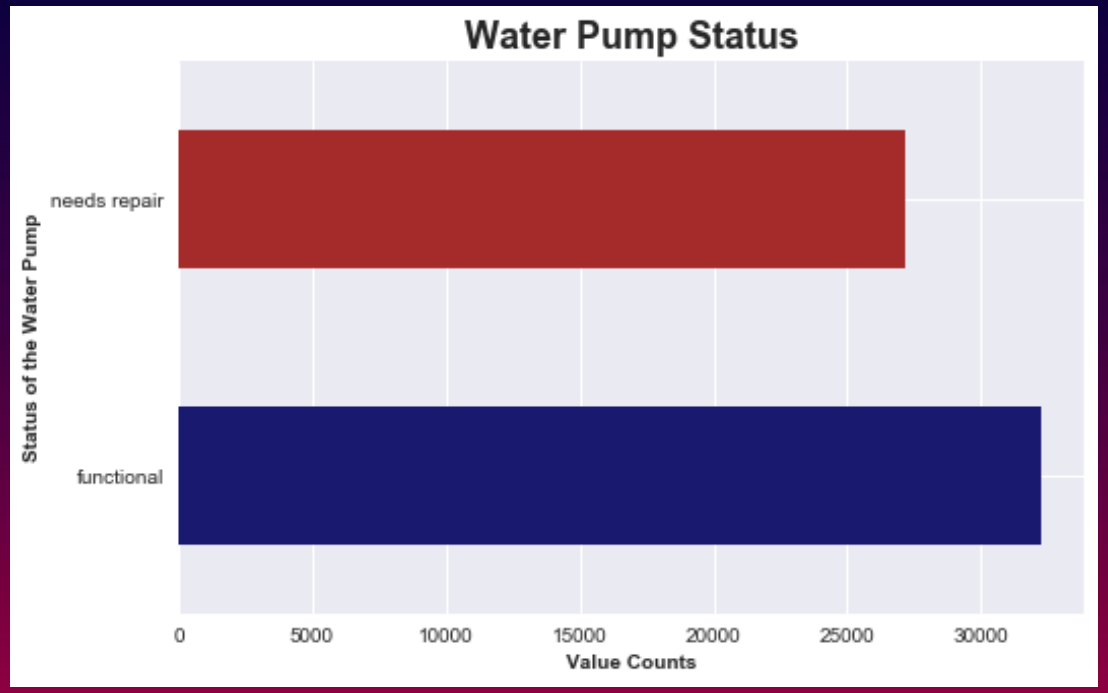
- Does source of water influence the functionality of the wells?
- Does Age of the wells effect the condition of the wells?
- What is the best mode of payment for maintenance of wells?
- Does having Public Meetings help to keep the wells functional?

Methodology

- **ROSEMED Methodology** - This is the one of the most straightforward of the Data Science processes.
  R - Research O - Obtain S - Scrub E - Explore M - Model E - Evaluate D - Deploy

- **Research** - Find out about various models and see which model works best for our data.
- **Obtain** - This step involves understanding stakeholder requirements, gathering information on the problem, and finally, sourcing data that we think will be necessary for solving this problem.
- **Scrub**- During this stage, we focus on preprocessing our data. Important steps such as identifying and removing null values, dealing with outliers, normalizing data, and feature engineering/feature selection are handled around this stage.
- **Explore** - During this step, we create visualizations to really get a feel of the dataset. We focus on things such as understanding the distribution of different columns, checking for multicollinearity, and other tasks like that.
- **Model** - It consists of building and tuning models using all the tools we have in our data science toolbox. In practice, this often means defining a threshold for success, selecting machine learning algorithms to test on the project, and tuning the ones that show promise to try and increase our results.
- **Evaluate** - Here we interpret the results of models and communicate results to stakeholders. If the results are satisfactory to all stakeholders involved, we can go from this stage right into putting our model into production and automating processes necessary to support it.
- **Deploy** - Deploying the Model.

Functioning status of the wells

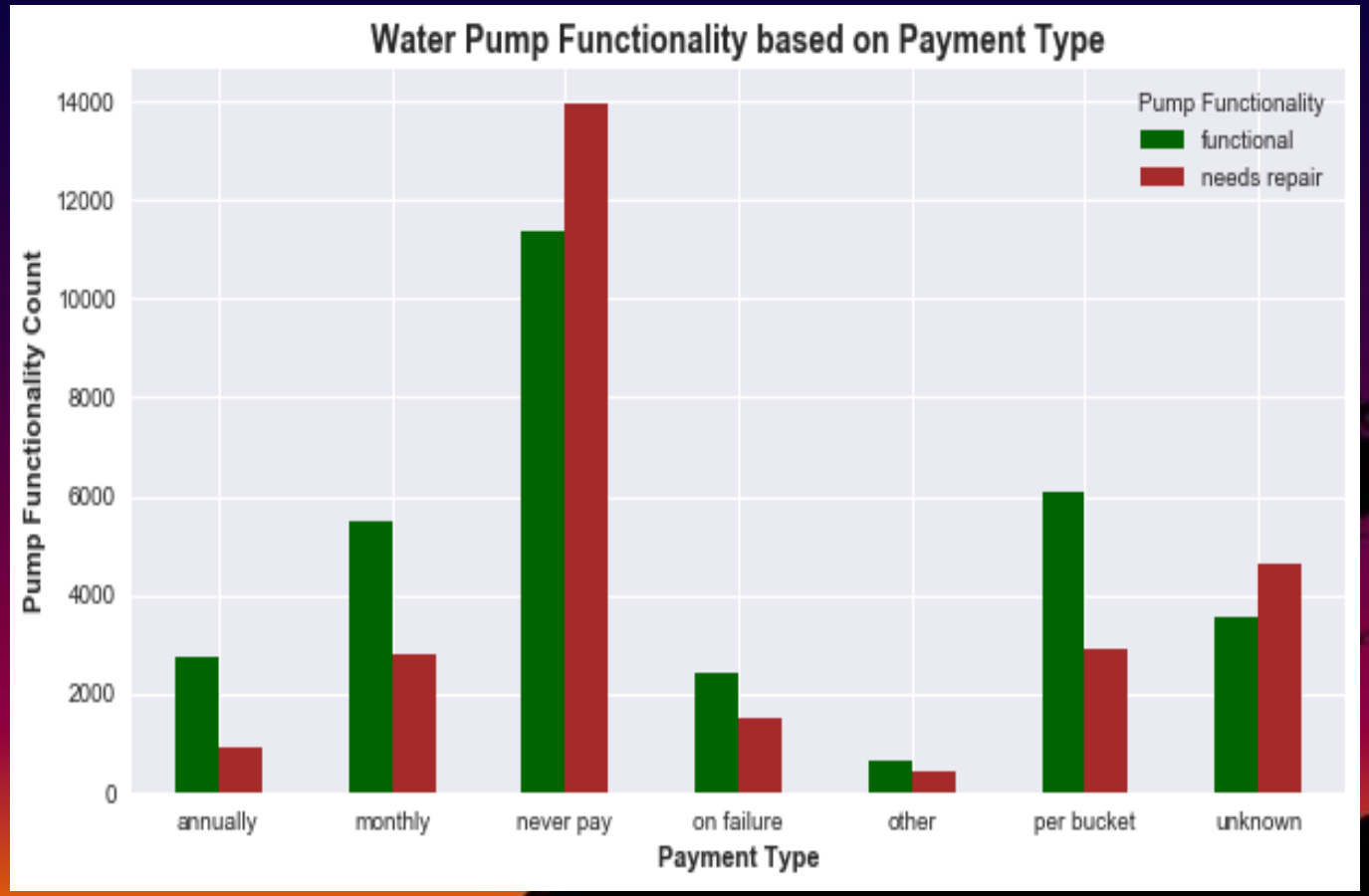


- Around 54% of the wells are functional.
- But 46% of the wells are not functioning or needs repair which is a concerning factor.

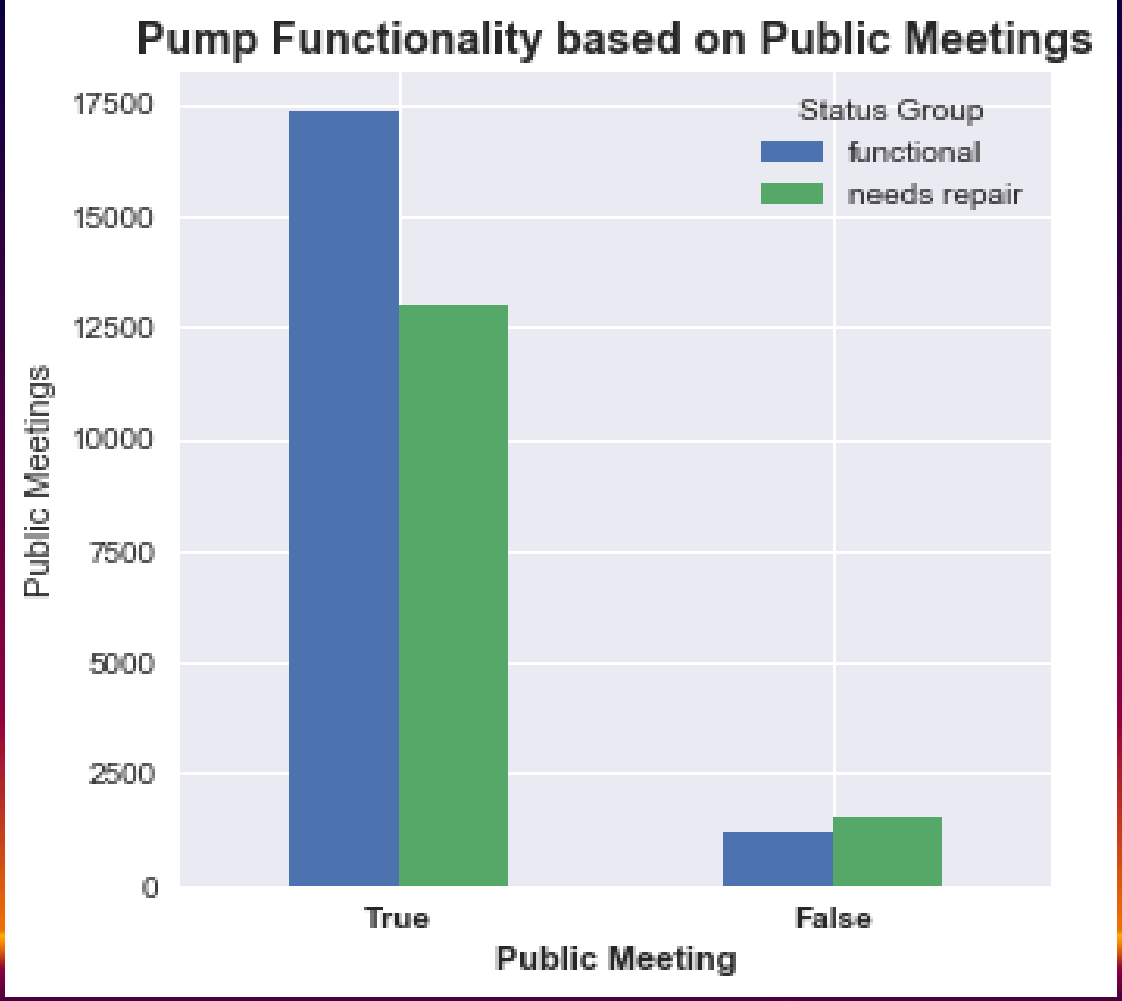
Payment Type

- Water pumps where there is a payment method are positively correlated with working water pumps, whereas when there are no payments, or it is not known tends to characterize not working/needs repair water pumps.

- The various payment options are per bucket, monthly or paying annually.

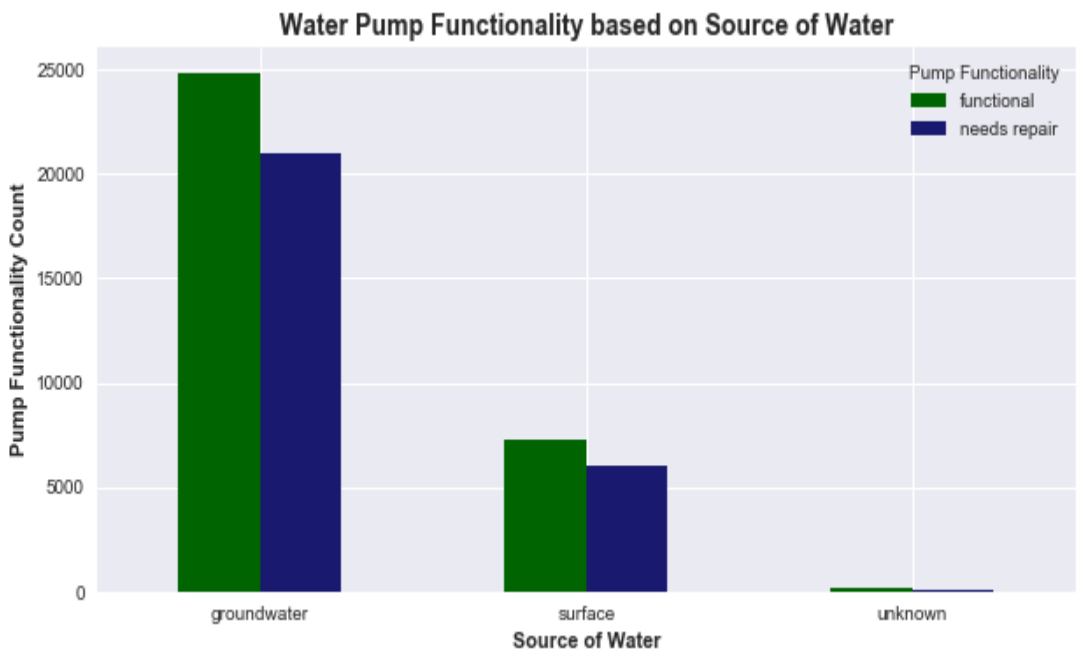


Public Meetings helping functionality of wells?



- We can see that having a public meeting helps in functioning of the wells. More than 50% wells are functional when there is a public meeting held for the same.
- Thus, Public meeting is an important factor for the functioning of wells.

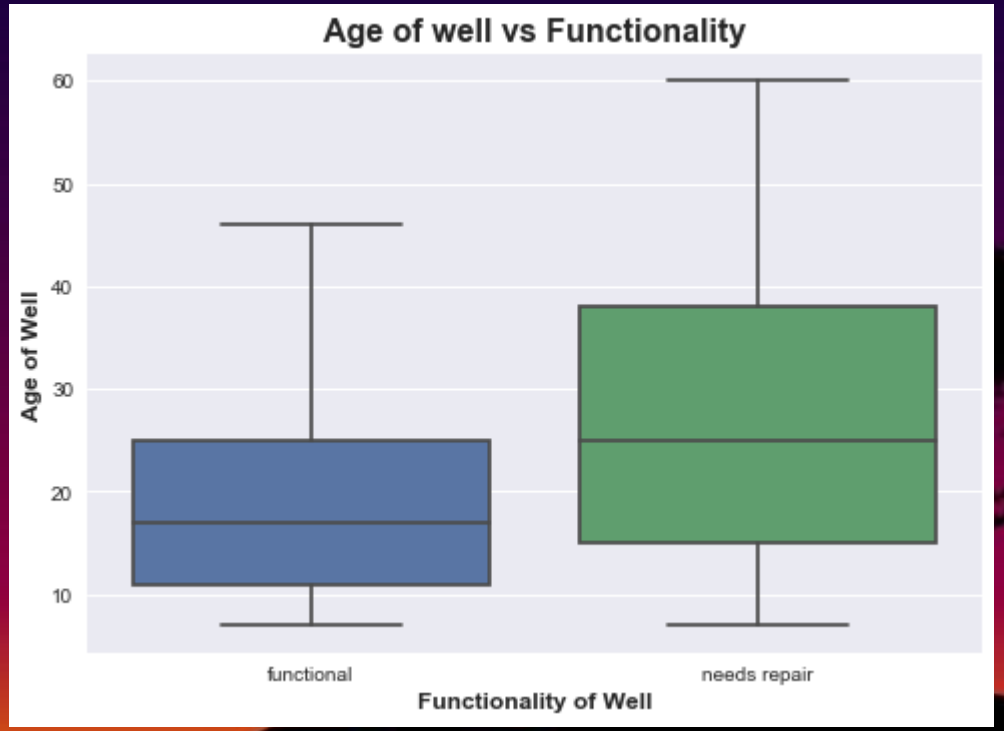
Source of Water



- Groundwater is very important to maintain the functionality of the wells. Almost, the entire water supply to the wells is dependent on groundwater.

Age of the Wells

- Water wells functioning declines over time.



Predictive Modelling

- A Model was built that can predict which pumps are functional and which need repair.
- We tried 3 Models – Linear SVC, XG Boost and Random Forest.

- XG Boost Model worked the best with a :

  Test Accuracy 81%

  Train Accuracy 83%

Some of our important factors for predictions were - Age of the well, Permit, Payment type, Scheme management, Lakes, GPS height.

Conclusion

Groundwater is very important to maintain the functionality of the wells. Almost, the entire water supply to the wells is dependent on groundwater. Hence, we should look into different methods such as rainwater harvesting and soil conservation which would also help sustain more water in the lakes.

We see that when we make payment based on bucket or monthly payment, the wells are maintained better. So payment is an important factor.

Definitely, age of the wells is an important factor to predict the functionality of the wells. The newer the wells are, more water they have or better is their condition.

We can see that having a public meeting helps in functioning of the wells. More than 50% wells are functional when there is a public meeting held for the same. Thus, Public meeting is an important factor for the functioning of wells.

Future Work

Need to examine the missing values or Zero values of Population and GPS height. Also, try binning the highly cardinal columns. Most static head values were 0. Need to examine those too.

Thank You