Predicting Water Well Functionality in Tanzania

Business Value

- Tanzania is currently a 62million population country. And still doesn't have enough well water resources for some of peoples in the country.
- Water is the basic needs for human body. Tanzania Government is currently working for solve this problem by improving clean water sources.
- There are many water wells already established, but some of them are non-functional or needs repair.

Problem/Goal

In this model, our aim is the predict **functionality** of water points machine learning models. This will help Tanzania Government for future work.

If a water point needs repair or why is not functional and what features affect functionality. With this model, we can help the Tanzanian authorities how to use water sources in a productive way.

Data Understanding

We have data from 1960 to 2013 with different funders at 21 different regions in Tanzania.

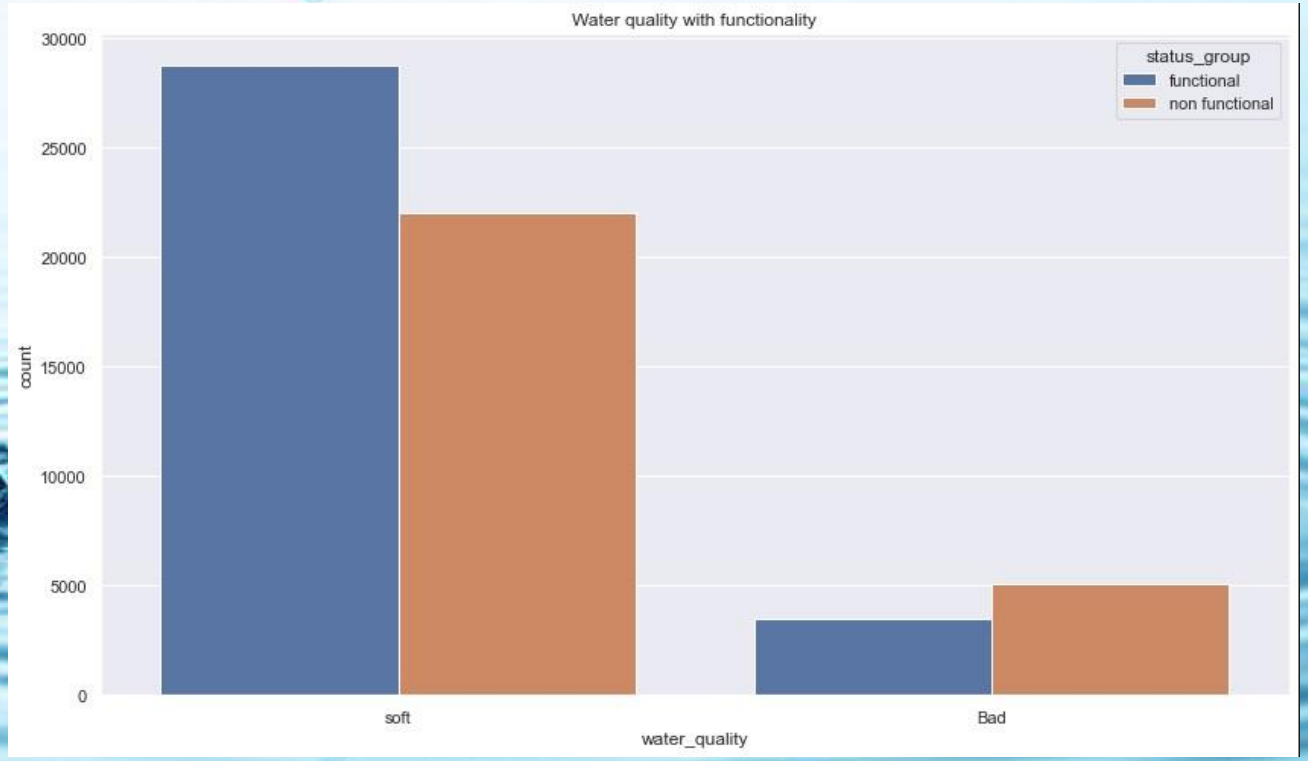
Also it has water quality as ; soft and bad with waterpoint type as; communal pipe, hand pump etc.

We have the quantity of the water source as enough, insufficient,dry and seasonal,

with source type as spring, shallow well, etc.

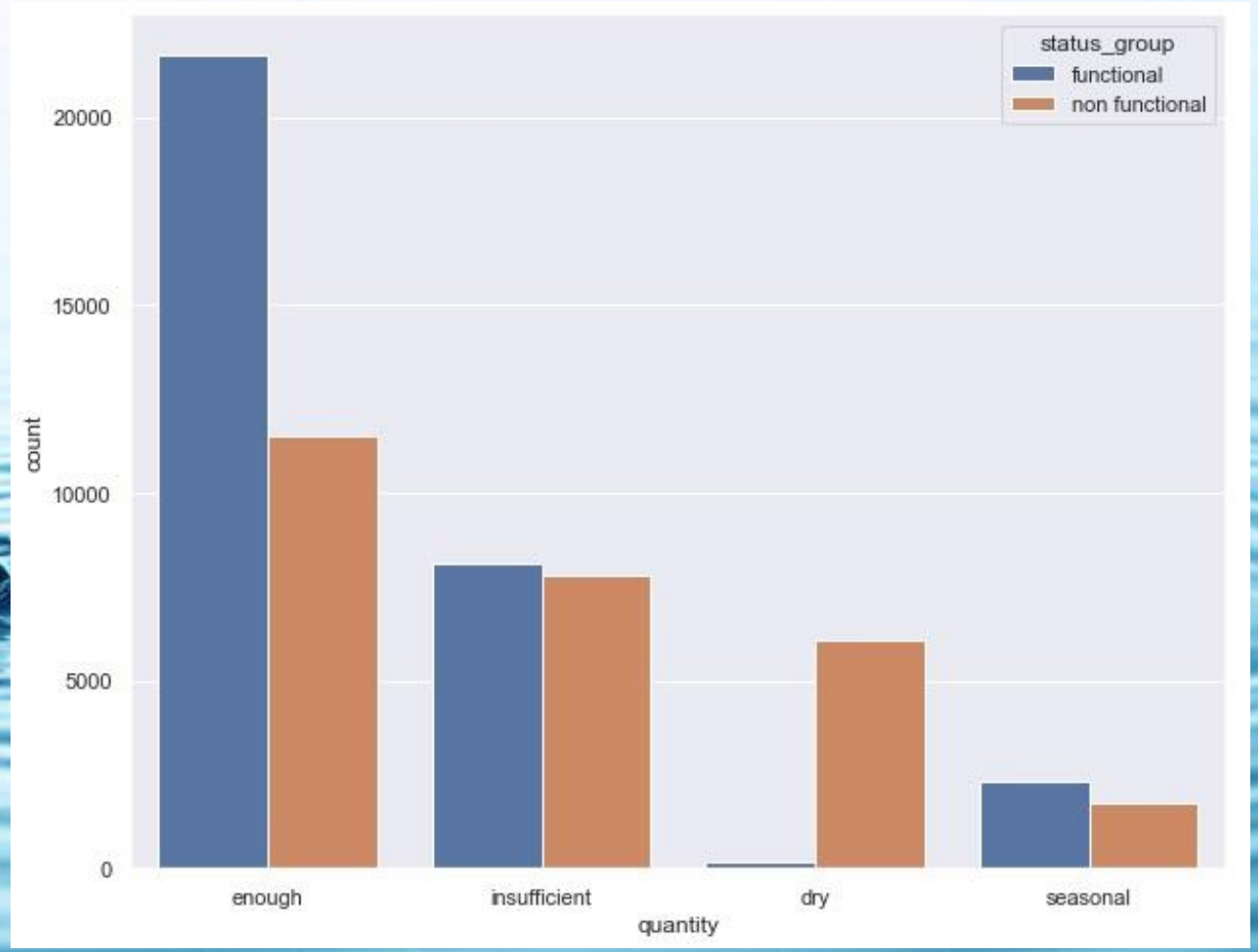
Water Quality with Functionality

As we processed our water quality values as soft(convenient) and bad(not usable). It seems convenient water quality points needs improvement, because there is a big portion of non-functional water points.

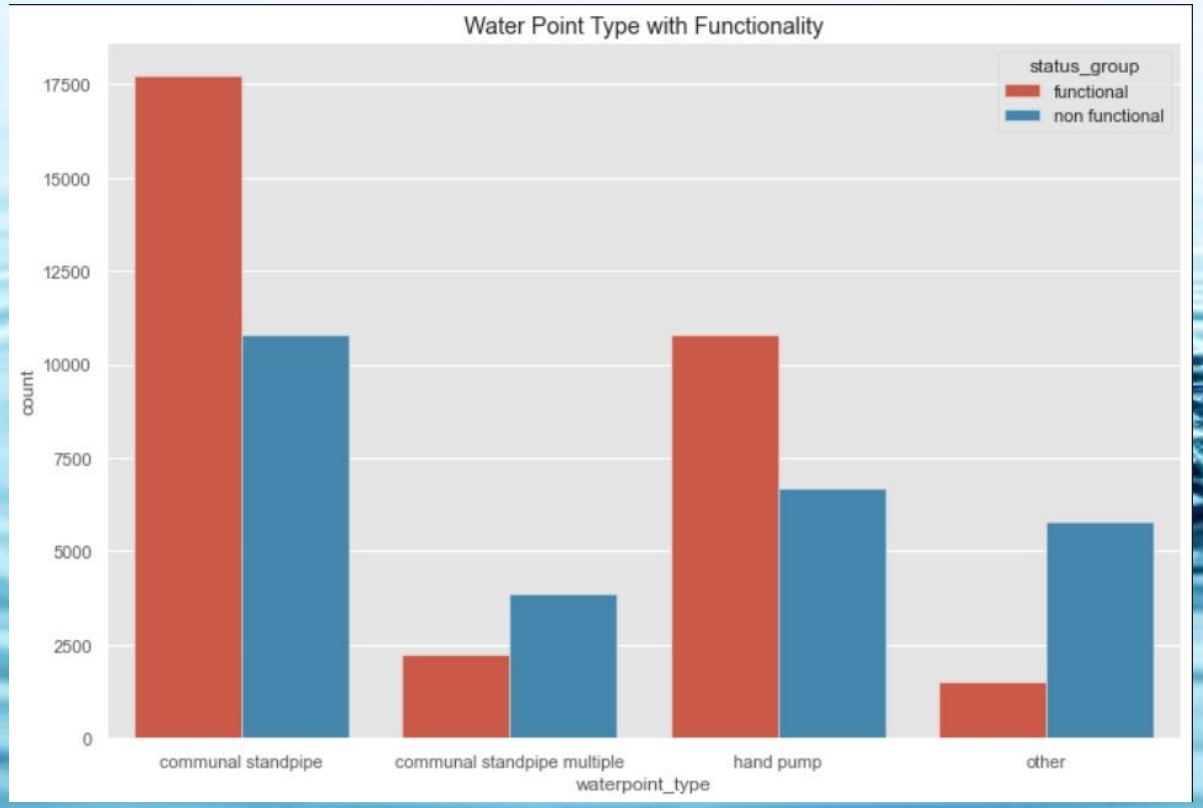


Water Quantity Functionality

It can be seen obviously that there is enough water quantity in some wells, they are non-functional. When looking at this graph, dry quantity water points have a high correlation with non-functionality. On the other hand, if the quantity is enough, there is a higher chance to find functional water points.

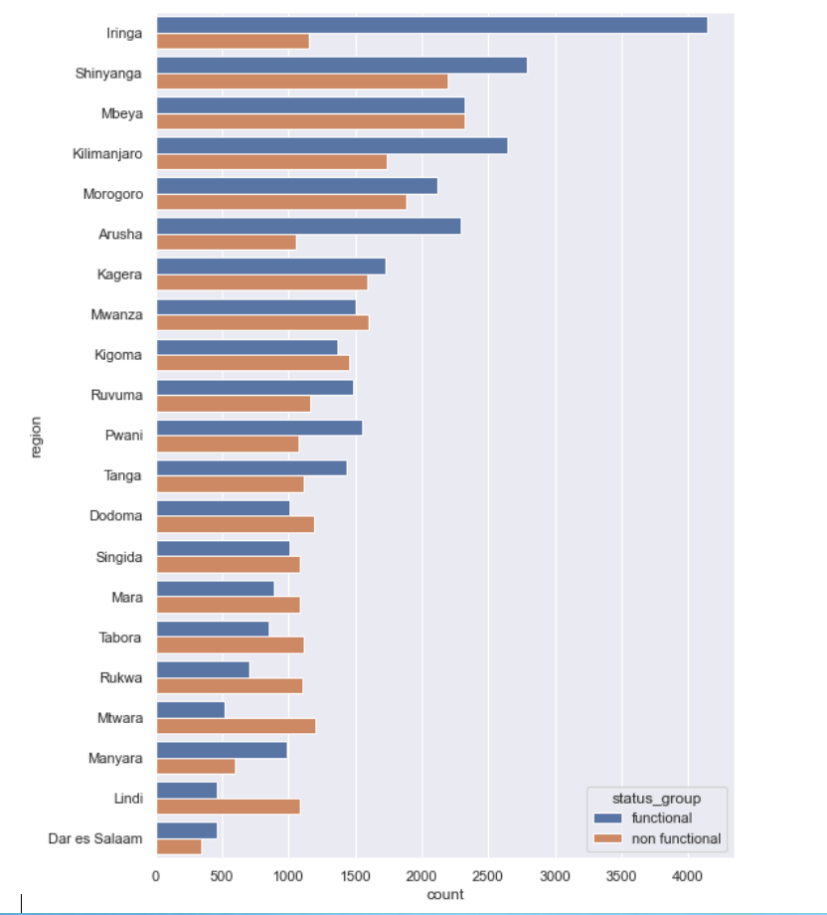


Water Point Type



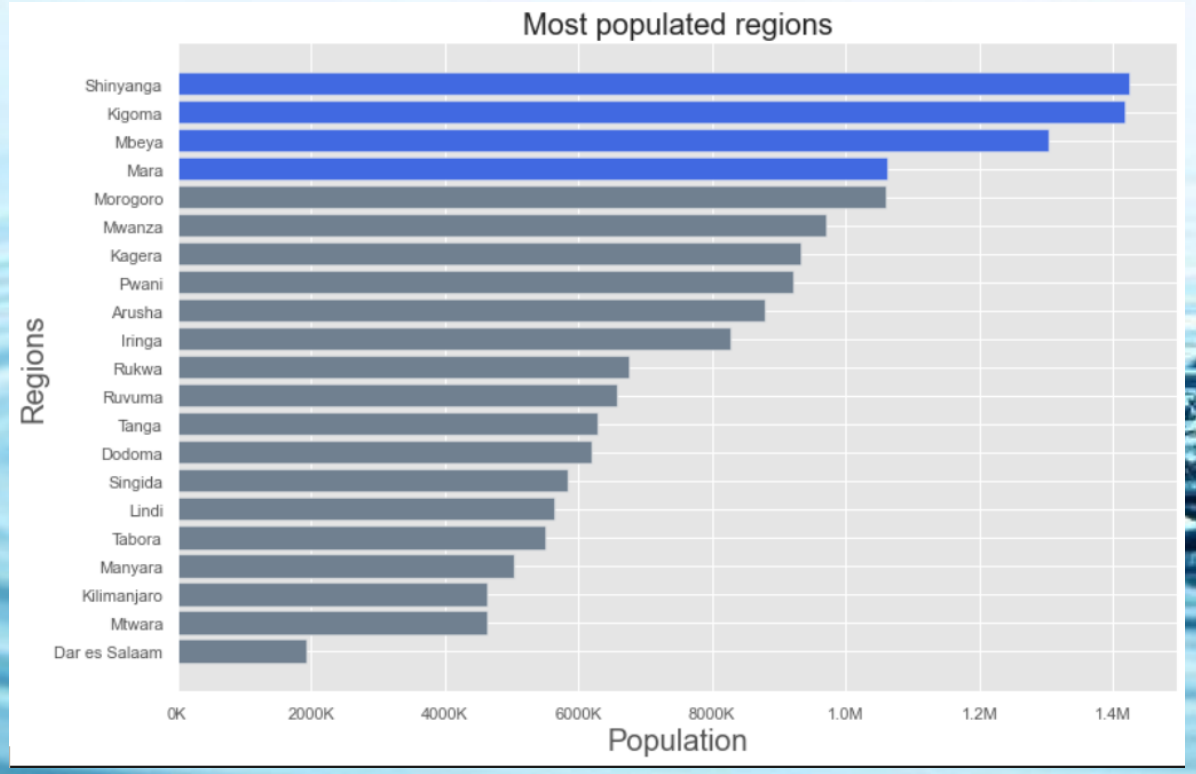
As we can see communal standpipe and hand pump mostly have functional. Communal standpipes with multiple types and others(dam,cattle trough) have mostly non-functional water points.

Regions



From this visualization we can easily see the region which has mostly functional water points has the highest number of water point. On the other hand less water point regions have mostly non-functinoal water points.

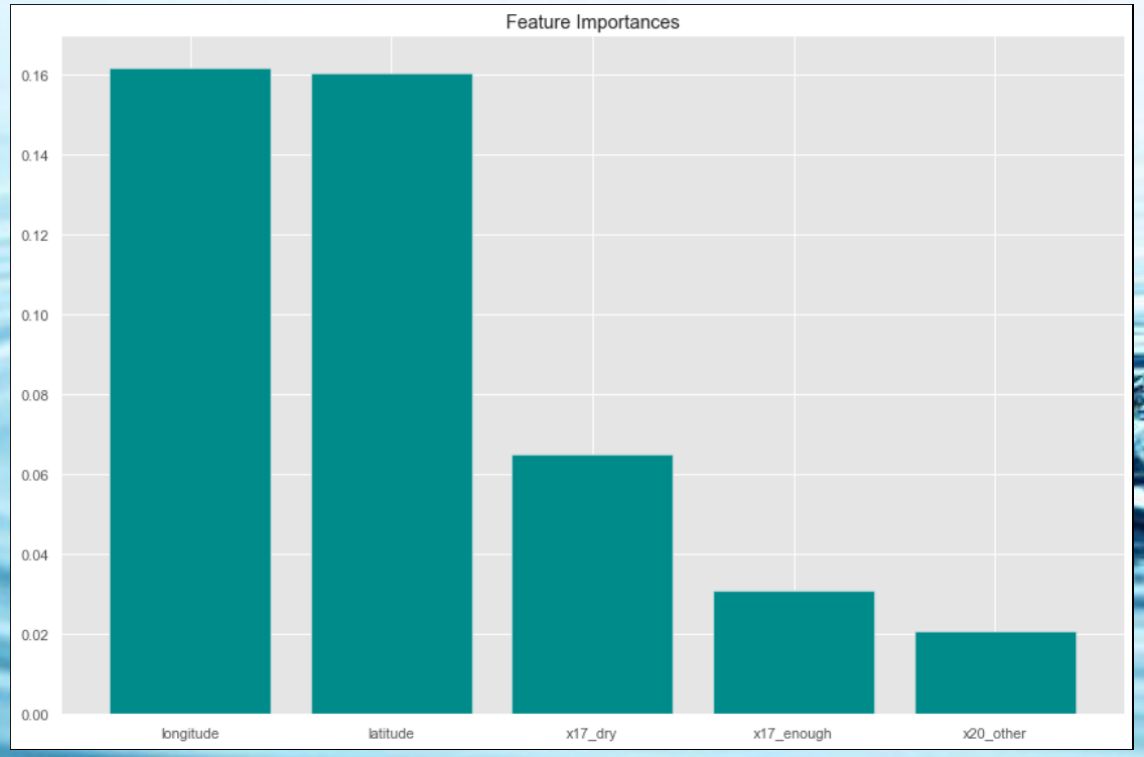
Regions Population



With this graph and previous one;

We can say there is not directly positive relation between population and water points at the regions. For instance; Kigoma is the second most populated region but has around 1500 non-functional water points. On the other hand, Shinyanga has around 2700 functional water points according to our data.

Findings



This is our most important features; as our model predicted longitude,latitude most important features. After those, in order dry water quantity and enough water quantity comes.

Conclusion

I created multiple models for the best result and my last model predicted(focused on) non-functional water points as **79** percent accurately. It shows **21** percent wrong prediction as; water point was functional but my model predicted non-functional in 21 percent.It is not the best but not bad also.

Future Step

1. Gather better quality data for prediction model.
2. Bring together old and new data for preparing for modeling.
3. Work on models to predict better.


Thank You