TANZANIAN WATER WELL PROJECT

MAJI NGO

MAJI NGO

Issues to Tackle

Maji NGO is a non-profit, Non-Governmental Organization dedicated to enhancing the welfare of people in African countries

Maji NGO wants to set up water pump station across the country to aid in more people having access to clean and safe water and furthermore help in repairing any pumps that have already been setup but have an issue that may or may not be affecting their functionality.

OBJECTIVES

- **MAINTENACE COST**

  We want to setup long lasting pumps so as to ensure more uptime and that the maintenance fees are not that high. This is due to the fact that if the pumps have low uptime, families benefit for a short period before going back to square one

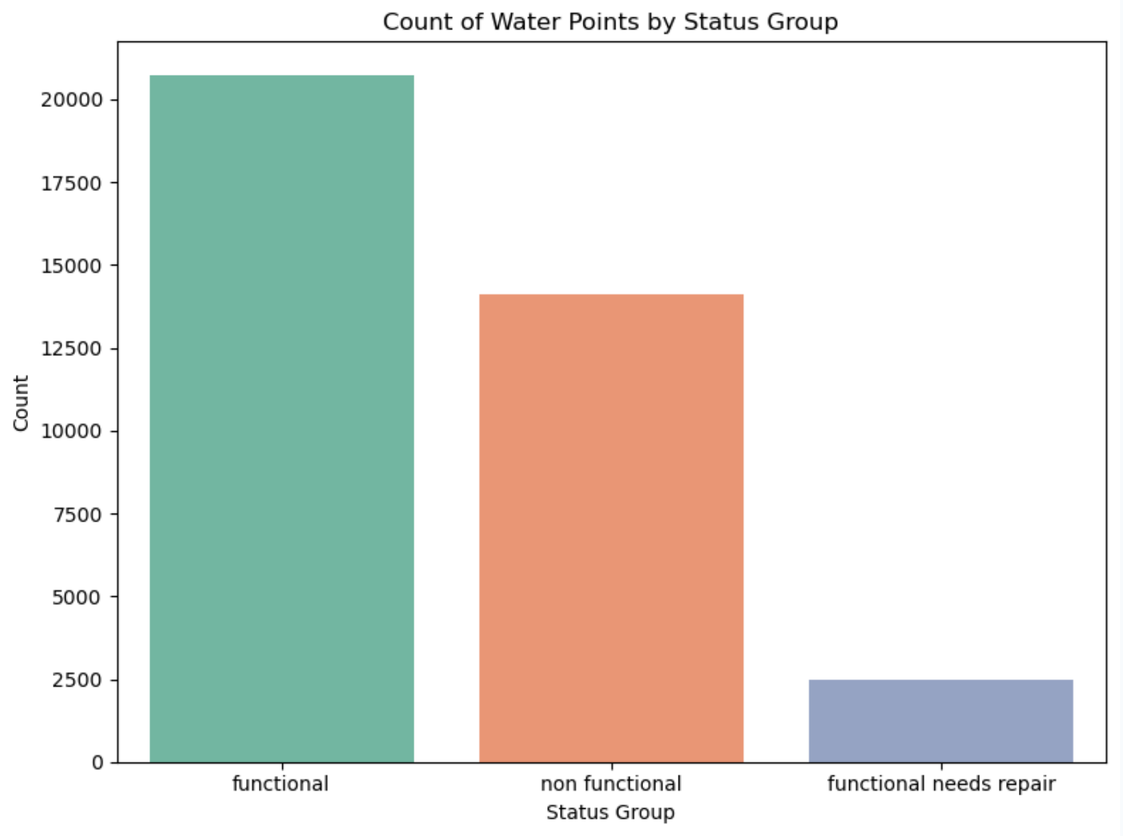
- **MORE UPTIME**

  Which pumps if setup will need a bit more maintenance than the rest, or rather what factors may increase the chances of a pump breaking down.

NUMBER OF WATER POINTS

We start off by visualizing the number of pumps that work, those that are working but need repairs, and those that are not working.

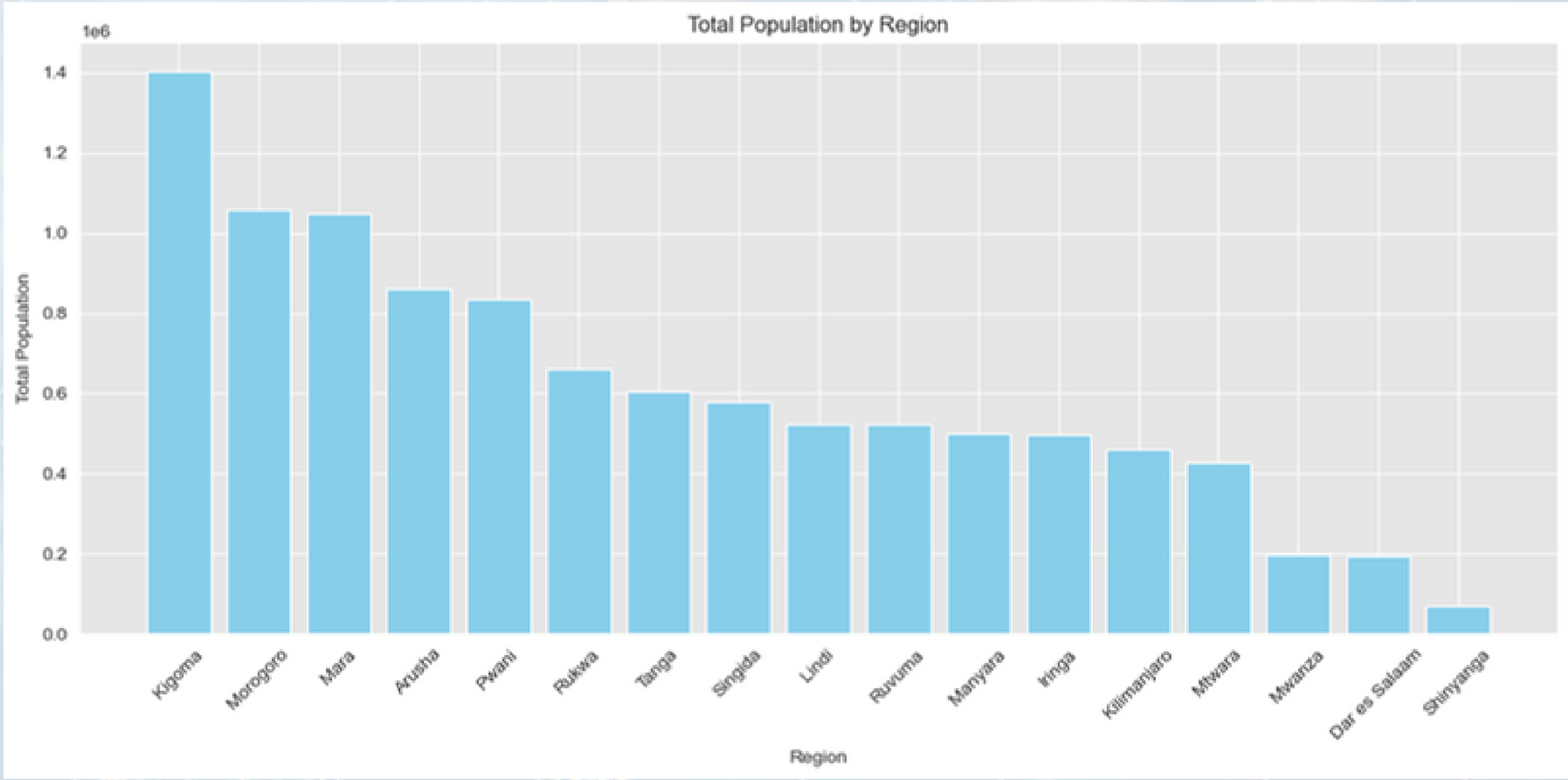
We can see most of the pumps are still functional even though some of them need repairs.



Total Population by Region

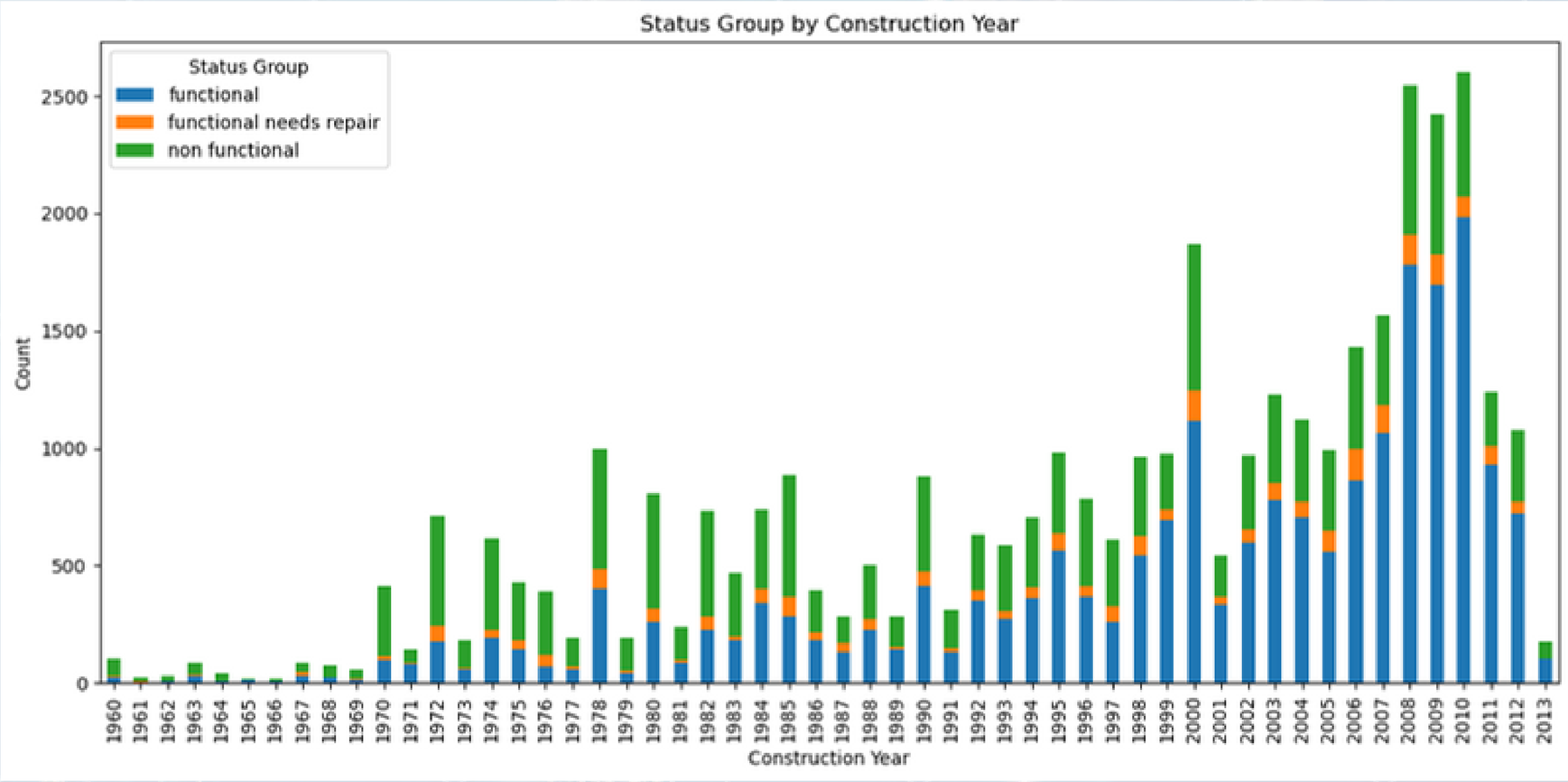
We can see the region with the highest population is Kigoma and the one with the least is Shinyanga.

This can help guide which areas are in more need of functional pumps due to more people living there.



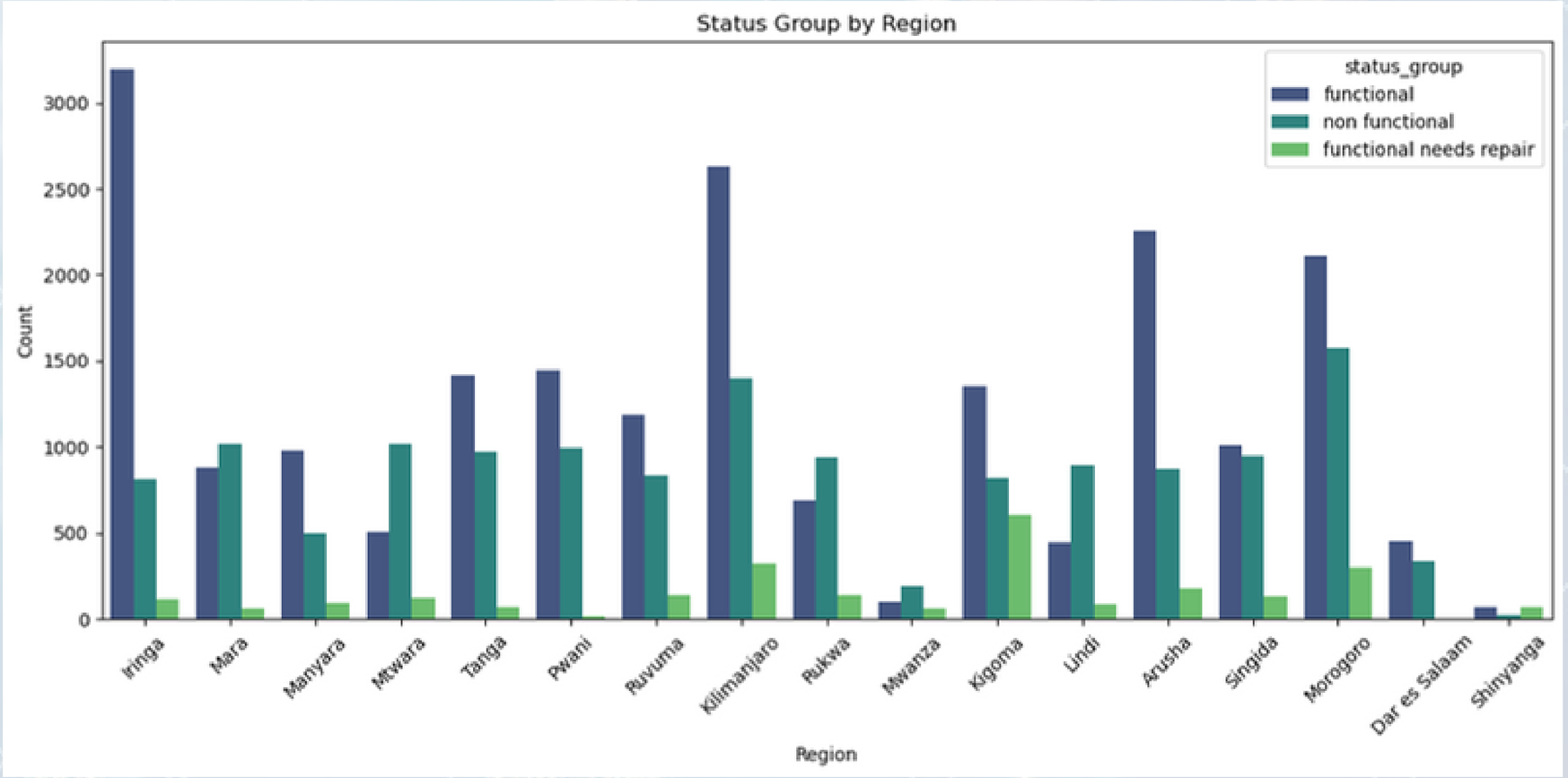
Functionality by Construction Year

We can see that most of the pumps set up earlier are non-functional and most of the more recent ones are still working. This shows what we suspected that the pump will wear down with age. The older it is the more likely it is to have broken down.



Status Group by Region

This graph shows us regions with high functionality. These are the regions where we would want to set up future pumps. These regions are like Iringa and Kilimanjaro.



MODEL PERFORMANCE VALIDATION

LOGISTIC REGRESSION MODEL	DECISION TREE MODEL	TUNED DECISION TREE MODEL	FOREST MODEL
Precision: 78.17% Recall: 65.78% Accuracy: 76.95% F1-Score: 71.44%	Precision: 75.39% Recall: 76.74% Accuracy: 78.83% F1-Score: 76.06% *Overfitting model	Precision: 79.66% Recall: 60.92% Accuracy: 76.06% F1-Score: 69.04%	Precision: 85.81% Recall: 64.10% Accuracy: 79.62% F1-Score: 73.38%

The Random Forest Model performed best on our training data.

Conclusion

I believe the forest model among all our models had the best results on the training data and can thus be used with 79.62%accuracy to determine which factors affect the water pump functionality

Next Steps

- Some of the columns had a large amount of missing data, If it is possible to collect this data from the field that would greatly improve model performance in terms of making inferences
- Run more models that have not been run in this notebook and counter-check with the ones in this notebook.
- If satisfied with the model results, then

TEAM

ONESPHORO M.
KIBUNJA

Data Science Student

DSPT-04

onesphoro.kibunja@student.moringaschool.com