

# Tanzania Water Well

BY MAUREEN ANDUURU



# BUSINESS OVERVIEW

- ▶ This project aims to create a classifier that predicts the condition of water wells in Tanzania. Tanzania, as a developing country, faces difficulties in providing its population with clean, a water, the population of over 57 million people. Many of the existing water points require repair, have failed completely or are fully working,



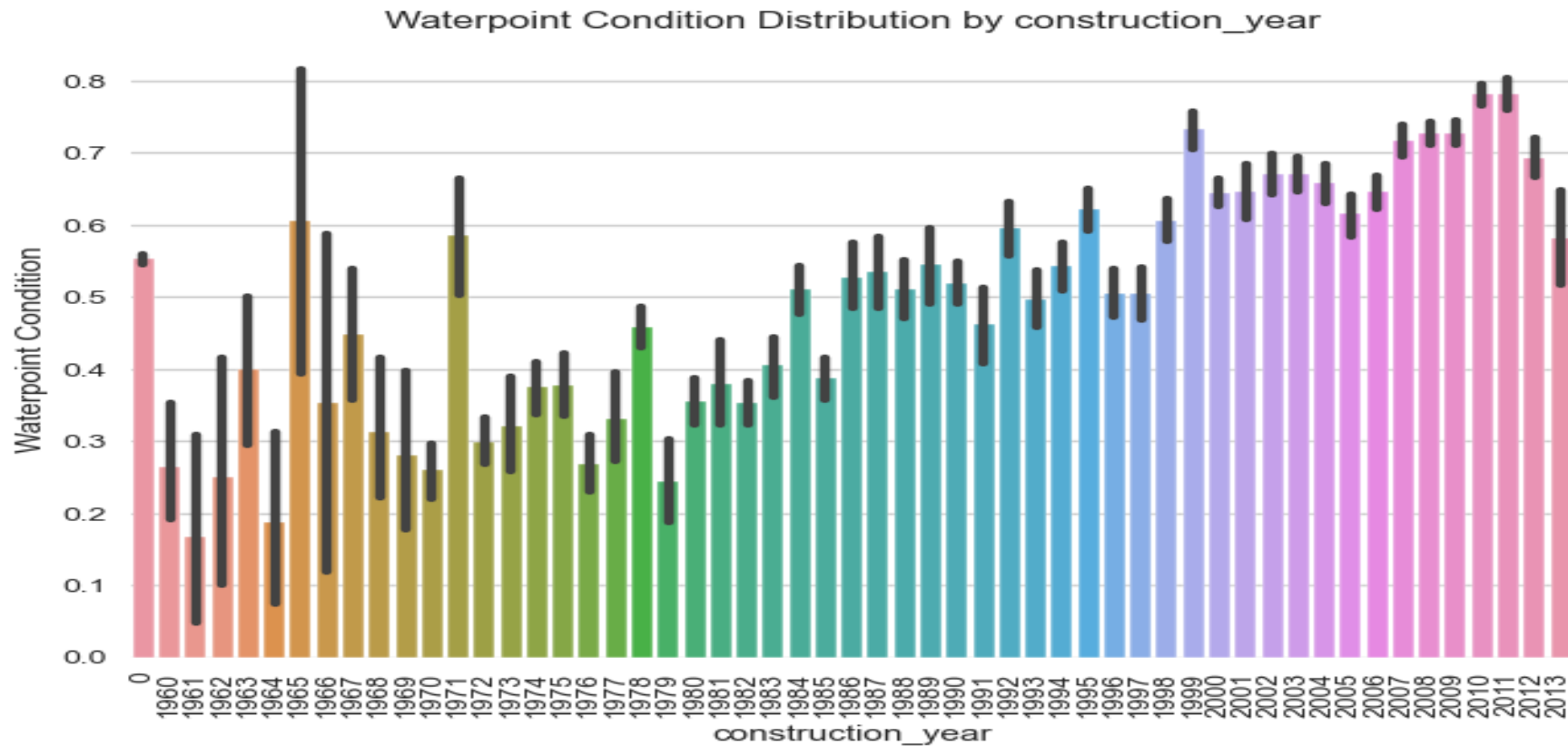
## BUSINESS PROBLEM

► Tanzania, as a developing country, struggles with providing clean water to its population of over 57,000,000. There are many water points already established in the country, but some are in need of repair while others have failed altogether.

# OBJECTIVES

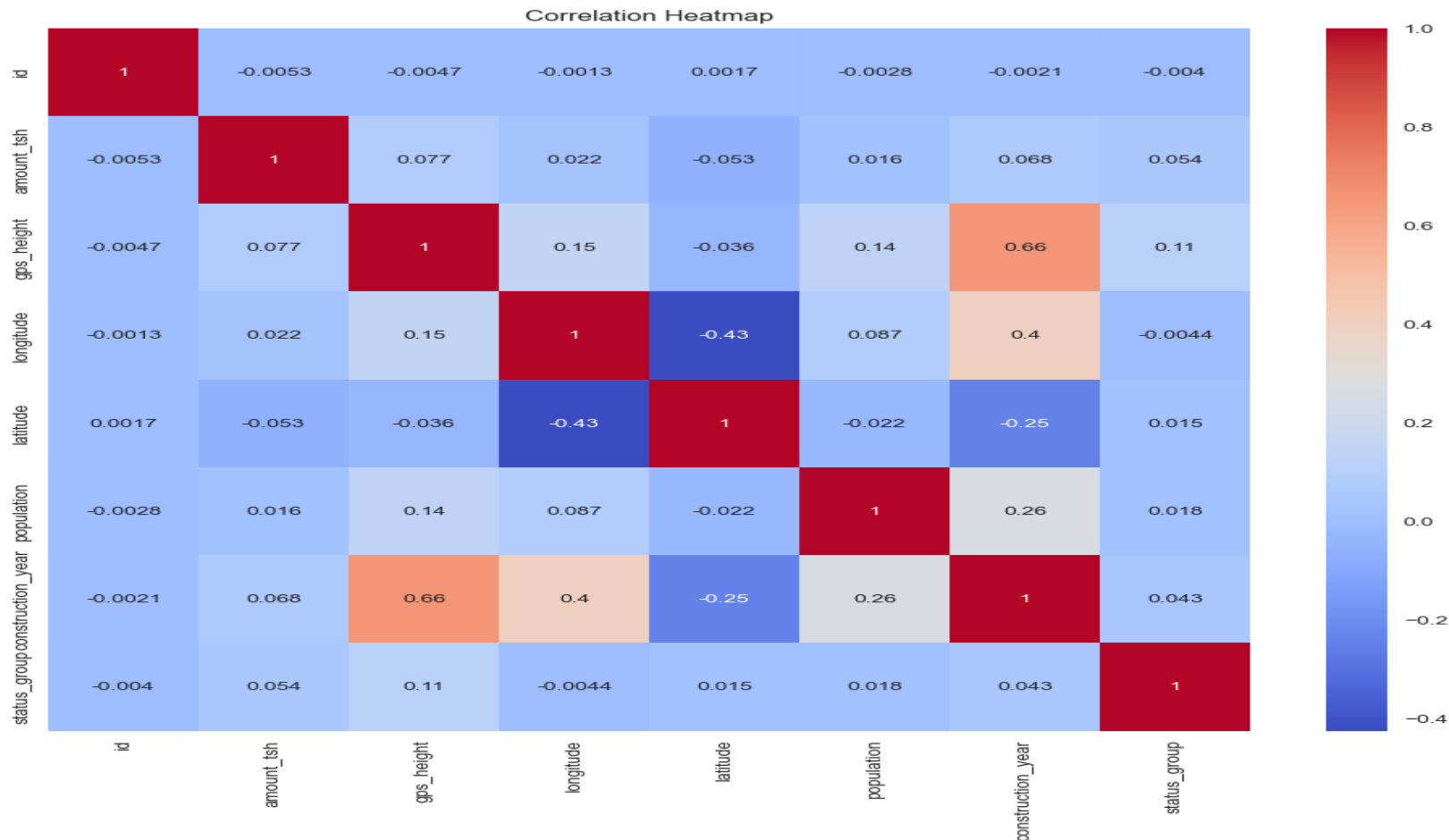
- ▶ 1.To build a classifier to predict the condition of water wells in Tanzania.
- ▶ 2.For the NGOs to focused on locating wells needing repair
- ▶ 3.The Government of Tanzania, which aims to identify patterns in non-functional wells to improve future well construction.
- ▶ 4.We will develop a classifier to predict the condition of water wells.

# DATA ANALYSIS



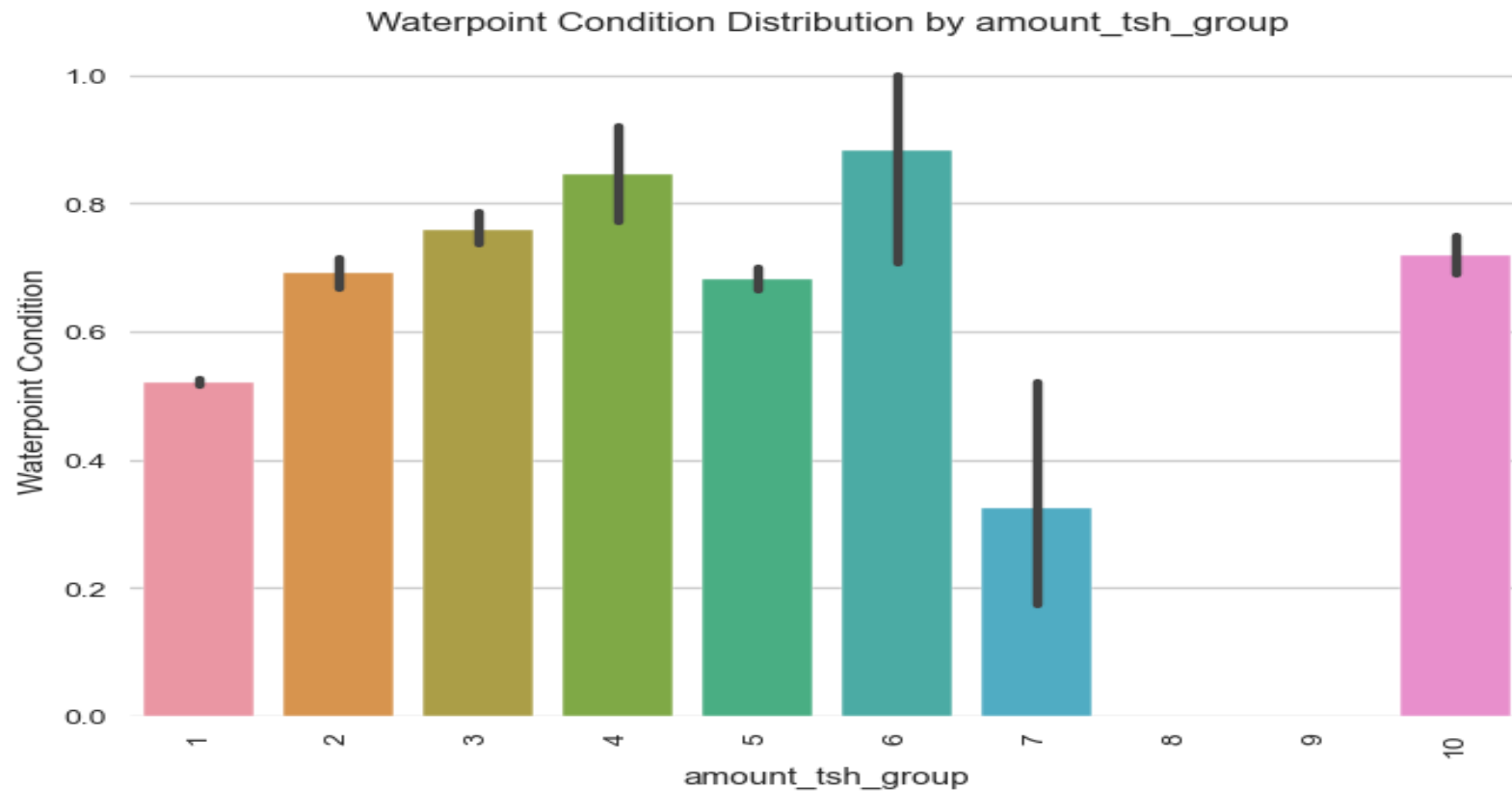
I plotted a graph waterpoint condition distribution and it showed that water wells constructed from 2007 to 2013 are functioning well as compared to wells in 1960.

# DATA ANALYSIS



A correlation matrix can be used to identify variables that are strongly correlated with each other, and may therefore be important predictors of a target variable. This can help in feature selection for predictive modeling tasks.

# DATA ANALYSIS



The graph shows the amount of money spent on the well depends the how well the water wells are taken care of and how they function the more the amount the more they functionality

# Modelling

## Logistic regression

Confusion Matrix

Actual	functional	functional needs repair	non functional
	5095	1	1361
	614	0	237
non functional	2481	8	2083
		Predicted	
		functional	non functional

## Random Forest

Confusion Matrix

Actual	functional	functional needs repair	non functional
	5572	213	672
	411	300	140
non functional	909	98	3565
		Predicted	
		functional	non functional

## K-Nearest Neighbors (KNN),

Confusion Matrix

Actual	functional	functional needs repair	non functional
	5273	174	1010
	487	200	164
non functional	1665	105	2802
		Predicted	
		functional	non functional



# CONCLUSION

- ▶ In conclusion, the Random Forest Classifier shows relatively good performance in predicting the condition of water wells, with higher accuracy, precision, recall, and F1-score for the "functional" and "nonfunctional" classes compared to the "functional needs repair" class. there is room for improvement in identifying wells that need repair.

# RECOMMENDATIONS

- ▶ I recommend
- ▶ 1. There is a need to do further exploration into other features in order to better understand the determinants of house prices.
- ▶ 2. The Government should consider partnering with the Locals and non-governmental organizations in maintaining the water wells.
- ▶ 3. The model in Project Phase 3 can be improved by using a binary classification approach.
- ▶ 4. More feature engineering to be carried out.
- ▶ 5. The government should offer incentives to the NGO to enhance more water wells to be dug for its people for example give tax waivers for NGOs that dig water wells and maintain them.

# NEXT STEPS

- ▶ Next Steps:
- ▶ 1. The road to the water points should be easily accessible for use and maintenance.
- ▶ 2. Revise the models so that it reflects the water pumps, installations so that this will allow for better accuracy in predicting the functional, nonfunctional water wells
- ▶ 3. The analyst can query how to improve in identifying wells that need repair.



Thank  
you.