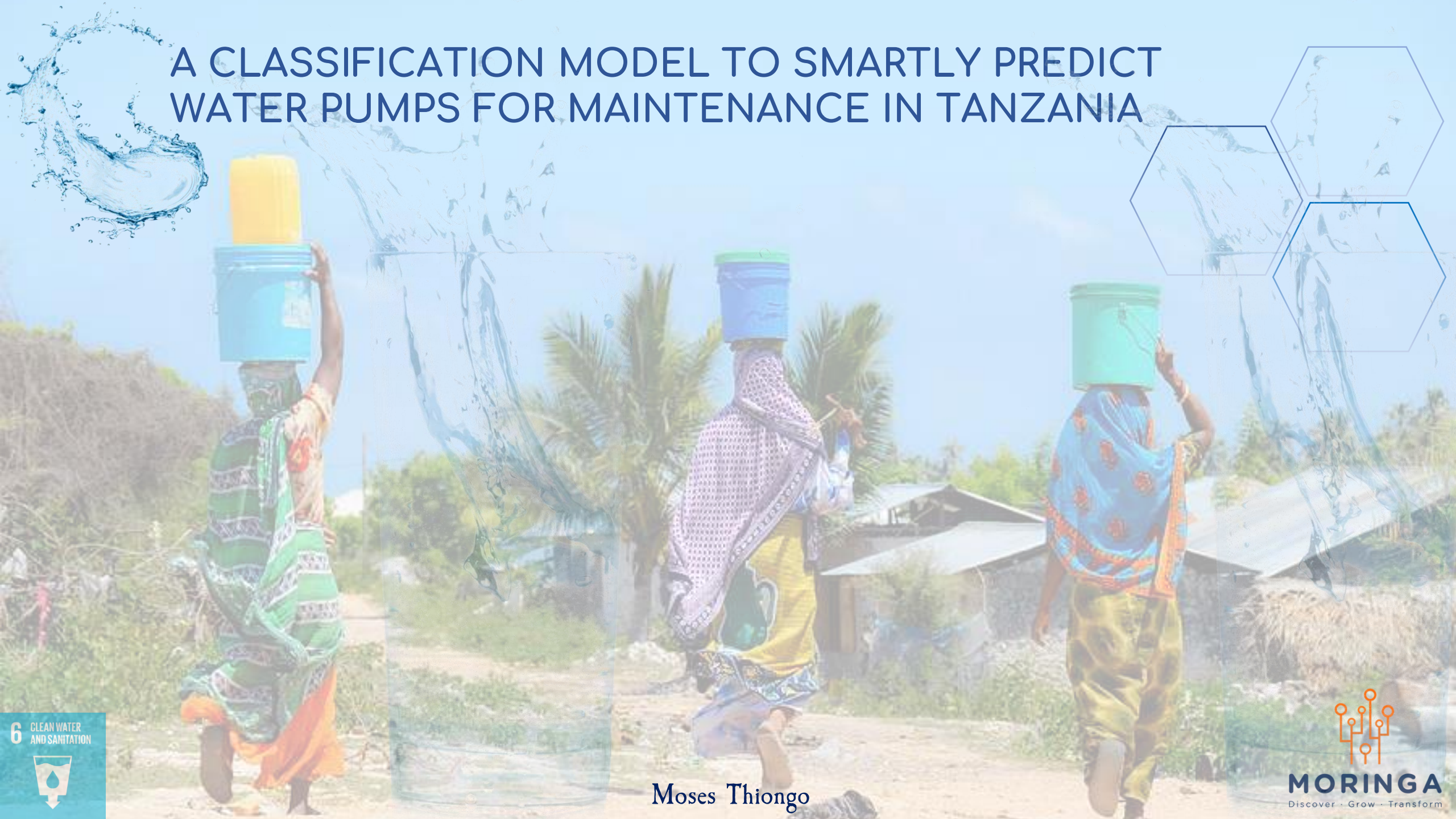
A CLASSIFICATION MODEL TO SMARTLY PREDICT WATER PUMPS FOR MAINTENANCE IN TANZANIA



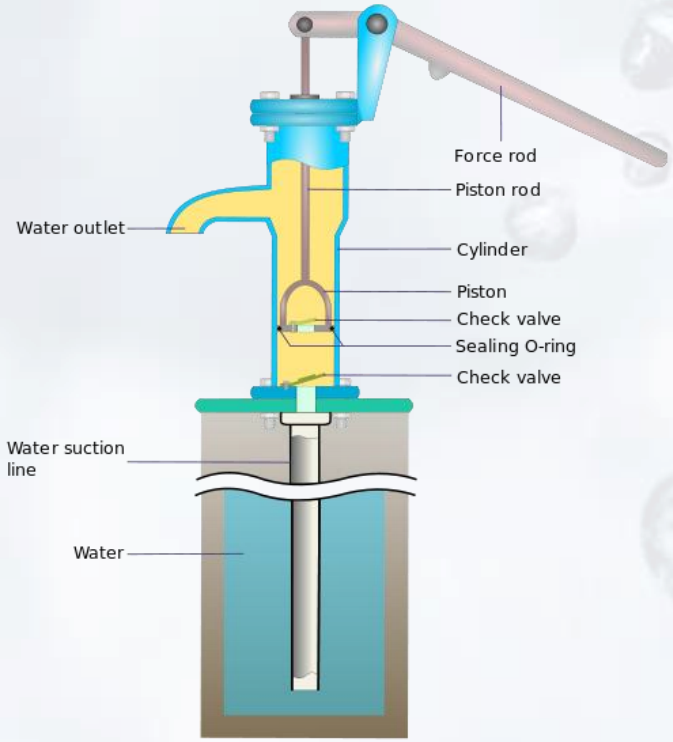
6 CLEAN WATER AND SANITATION

Moses Thiongo

MORINGA
Discover · Grow · Transform

INTRODUCTION

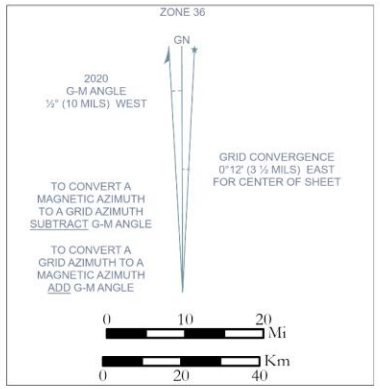
A Model to predict which water pumps will need maintenance to ensure that clean, and potable water is available to communities across the United Republic of Tanzania.



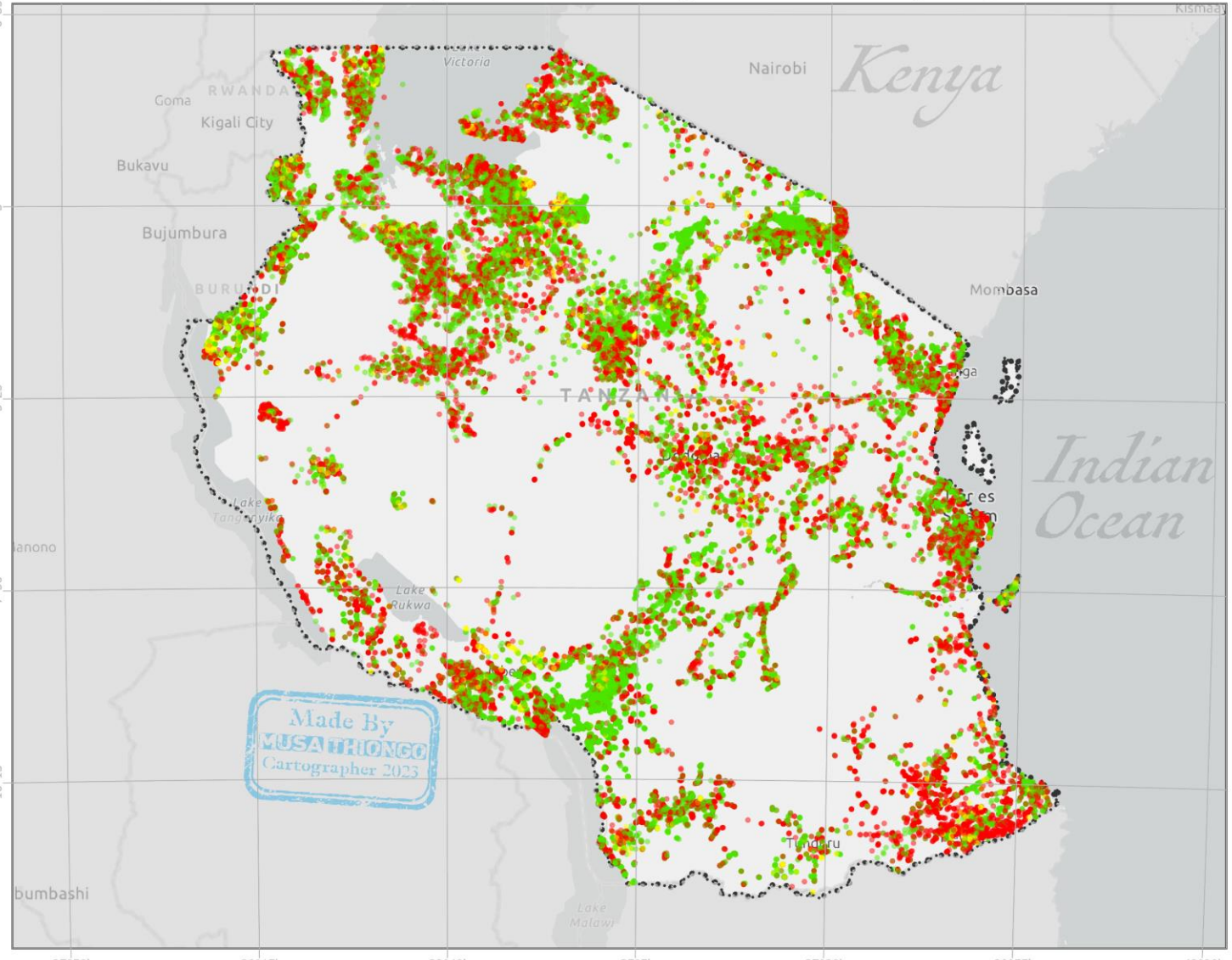
Tanzania Water Pumps

STATUS

- Functional
- Functional Needs Repair
- Non Functional



DISTRIBUTION OF WATER PUMPS IN TANZANIA



The map shows 57,588 water pumps out of 59,400 which are supplied in the original dataset. About 1,812 water pumps are mapped in the null island. This means that the water pumps have no valid geometry and cannot be mapped within Tanzania's defined boundaries. They appear somewhere in the Atlantic oceans, south of Nigeria. This point is called the Null Island (0,0). From the map, we can see that the water pumps are clustered in certain regions of the country and not others; there is a higher proportion of red dots, the non-functional water pumps, in the southeast and green dots, the functional water pumps, in the central and southwest parts of the country.

2023

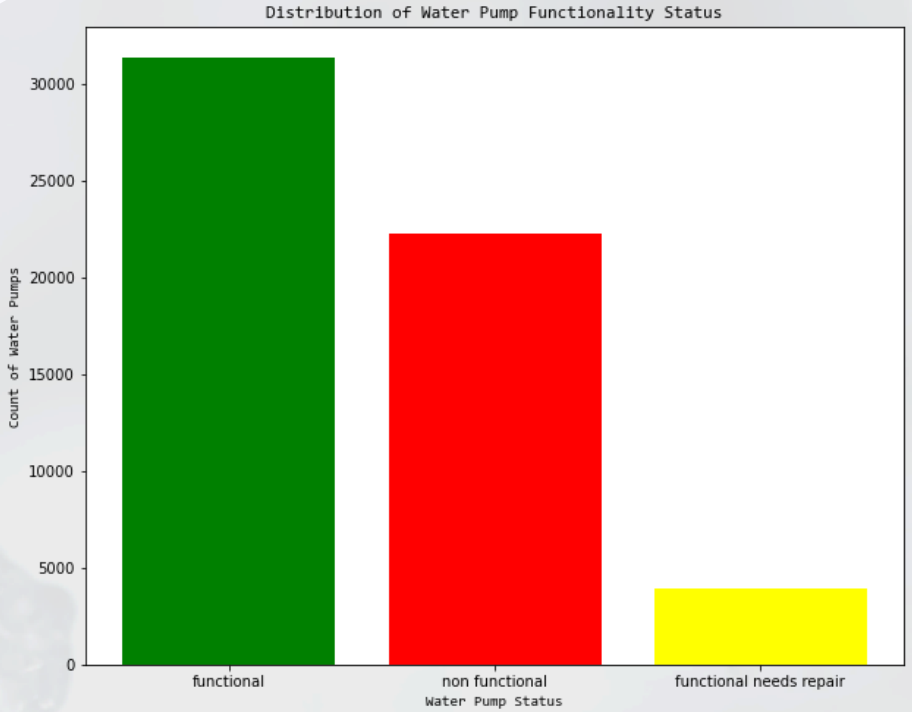
Sources: Esri, HERE, Garmin, FAO, NOAA, USGS, NaturalEarth.com, Taarifa-http://taarifa.org/whichaggregatesthe datafromtheTanzanianMinistryofWater

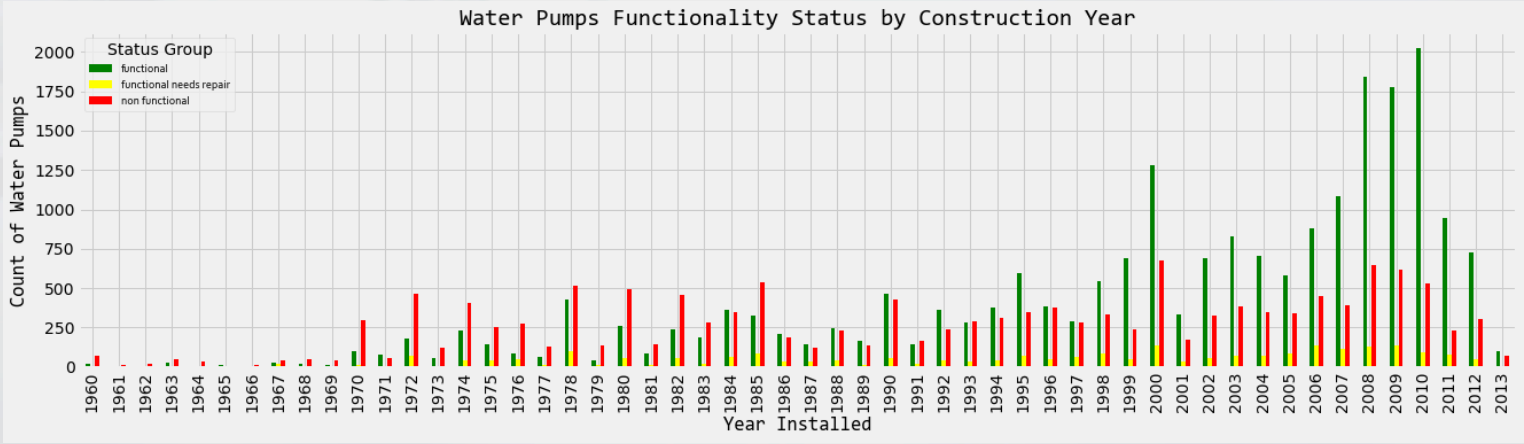
Coordinate System: Arc 1960 UTM Zone 36S

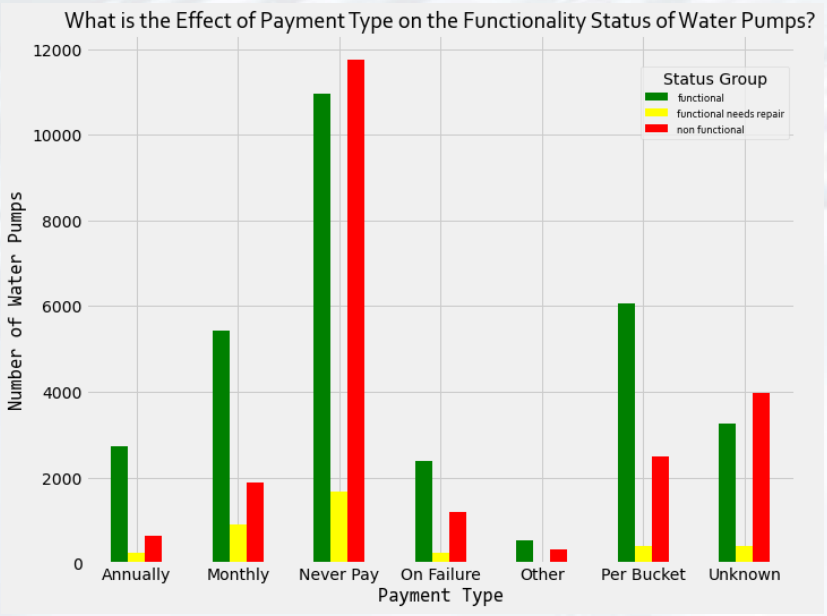
DATA

The Data is download from the Driven Data site and has been attributed to the Taarifa which aggregates the data from the Tanzanian Ministry of Water.

Generally, there are many water pumps that are functional compared to the others. However, there is a high % of water pumps that are nonfunctional.

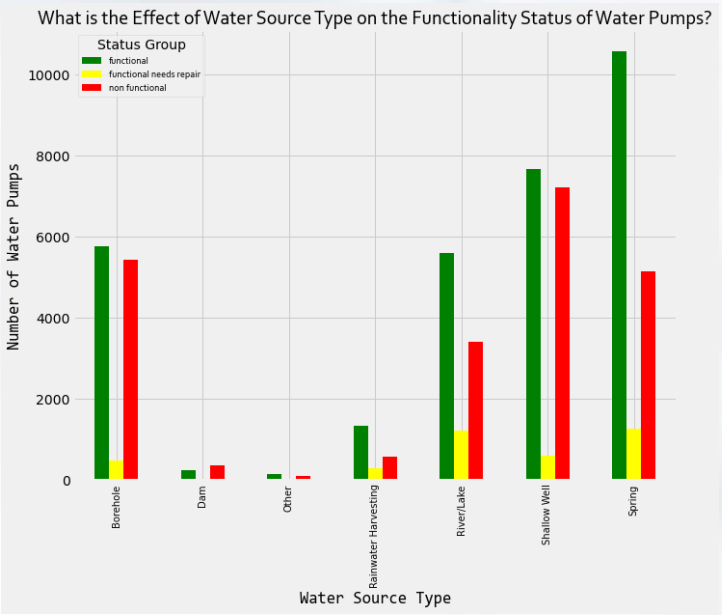


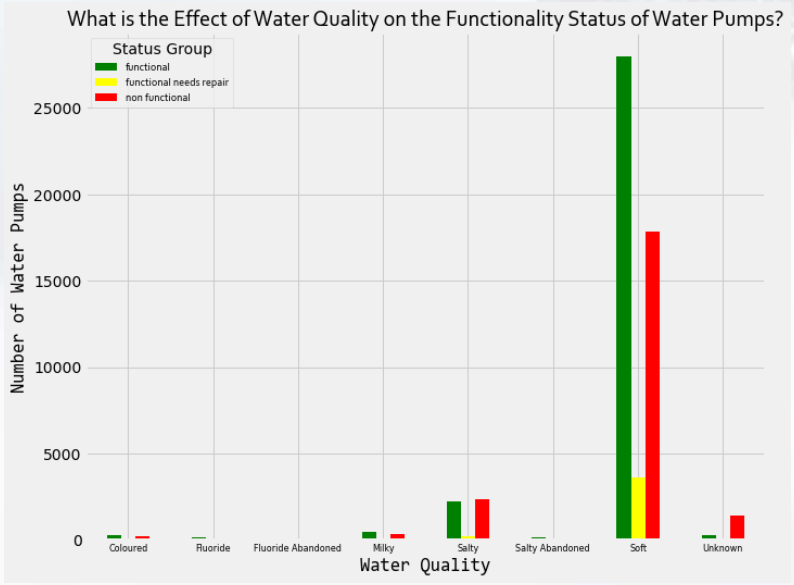




The Payment Type with ***Never Pay*** has the highest of the three status groups. c

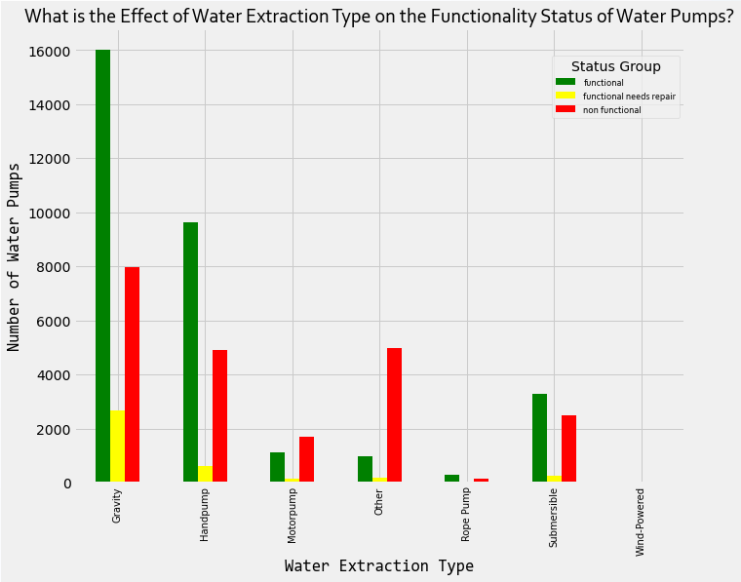
Water Source Type that comes from the spring has the highest number of functional water pumps. Shallow wells has the highest of nonfunctional water pumps.

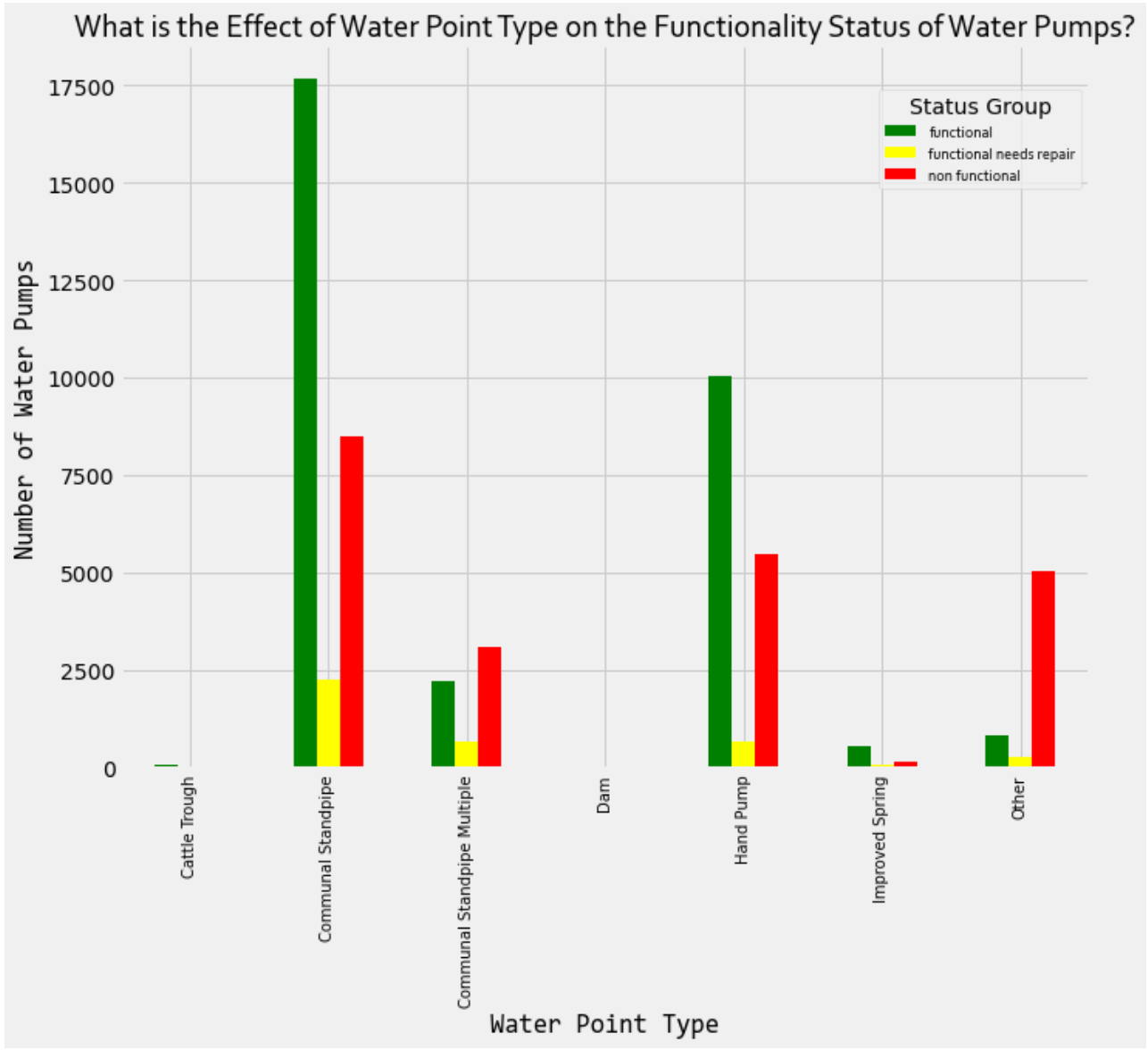


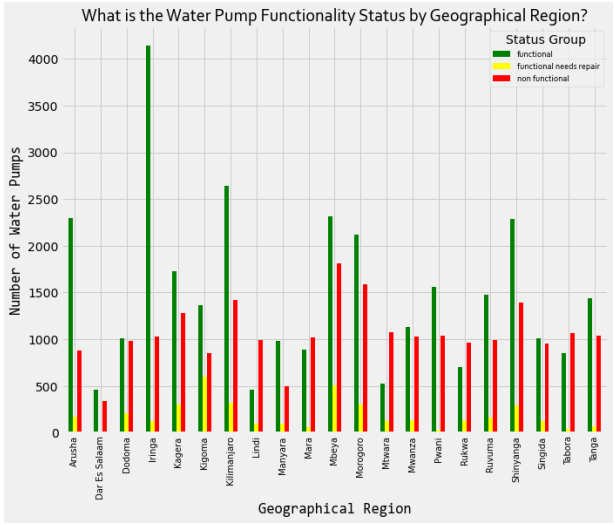


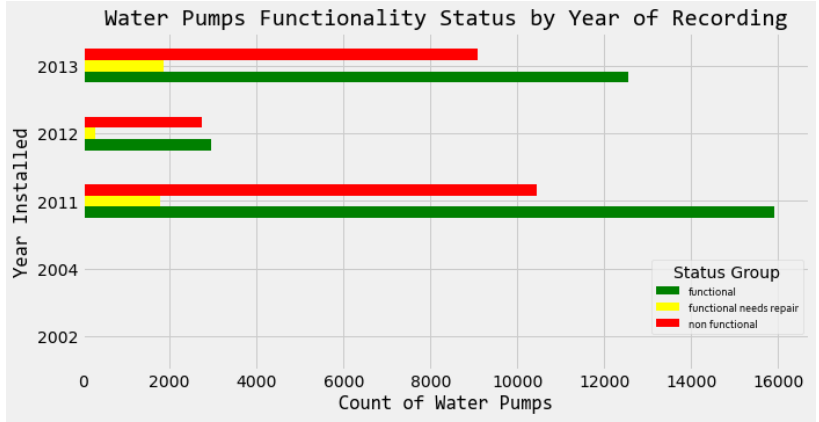
Soft water has the highest number of water pumps that are functional, functional needs repair and nonfunctional

Water Extraction type with the highest functional and nonfunctional water pumps is Gravity followed by handpump.



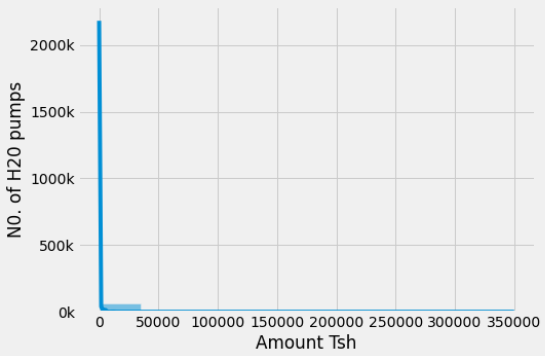

What is the Effect of Water Point Type on the Functionality Status of Water Pumps?
Number of Water Pumps
17500
15000
12500
10000
7500
5000
2500
0
Status Group
functional
functional needs repair
non functional
Cattle Trough
Communal Standpipe
Communal Standpipe Multiple
Dam
Hand Pump
Improved Spring
Other
Water Point Type

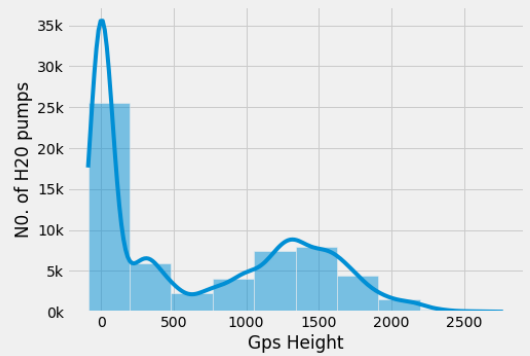

What is the Water Pump Functionality Status by Geographical Region?
Status Group
functional
functional needs repair
non functional
Number of Water Pumps
4000
3500
3000
2500
2000
1500
1000
500
0
Arusha
Dar Es Salaam
Dodoma
Iringa
Kagera
Kigoma
Kilimanjaro
Lindi
Manyara
Mara
Mbeya
Morogoro
Mtwara
Mwanza
Pwani
Rukwa
Ruvuma
Shinyanga
Singida
Tabora
Tanga
Geographical Region

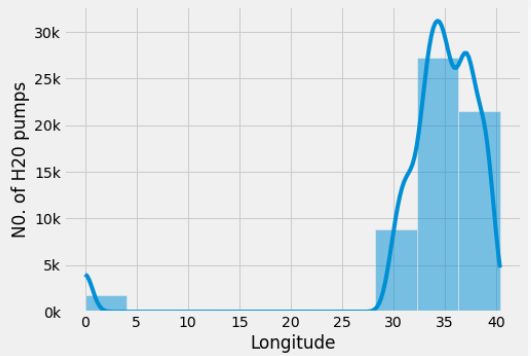

Water Pumps Functionality Status by Year of Recording
Year Installed
2013
2012
2011
2004
2002
Status Group
functional
functional needs repair
non functional
0
2000
4000
6000
8000
10000
12000
14000
16000
Count of Water Pumps

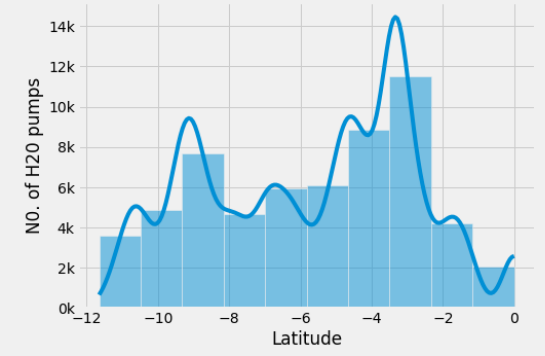
OUTLIERS

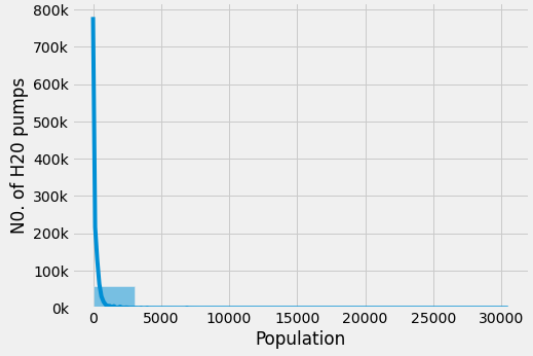
Outliers are examined and dealt with in our classification model. The column placeholders were also carefully taken care of.

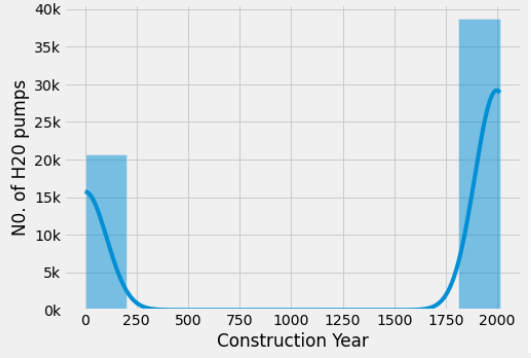




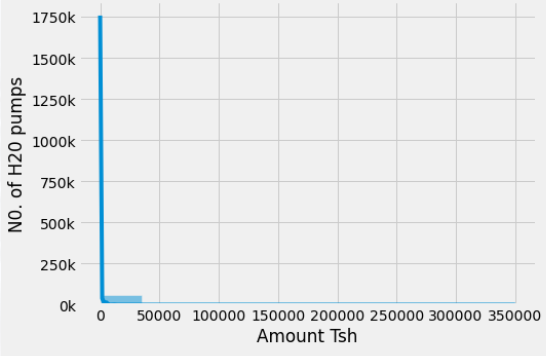


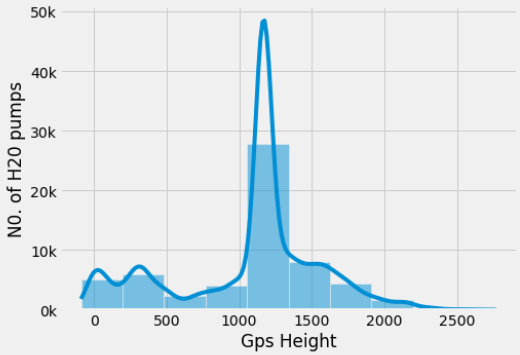


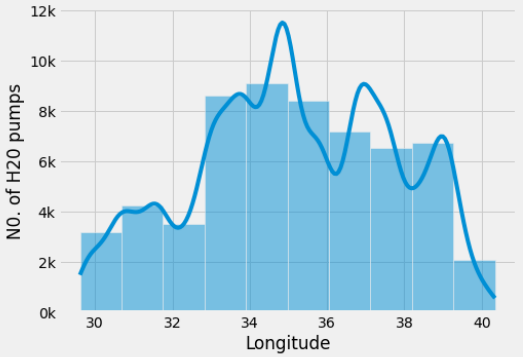


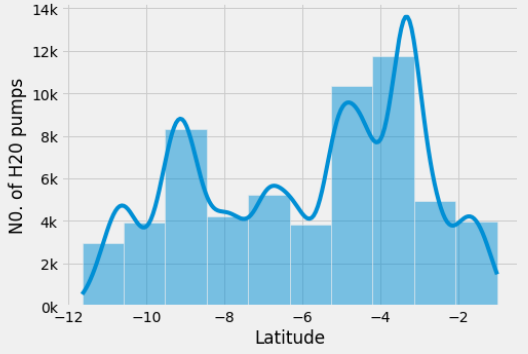


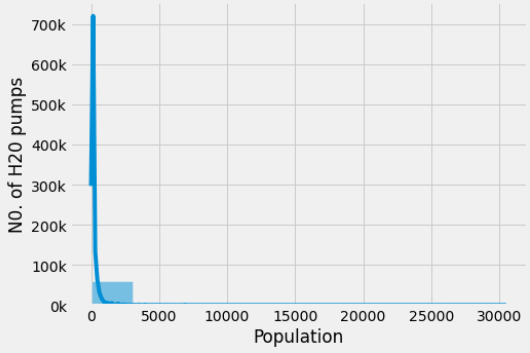
The Longitude and Latitude column that had 0,0 were removed to avoid those entry records which could not be properly mapped in Tanzania.

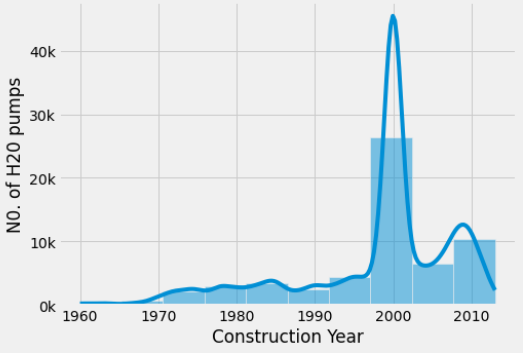


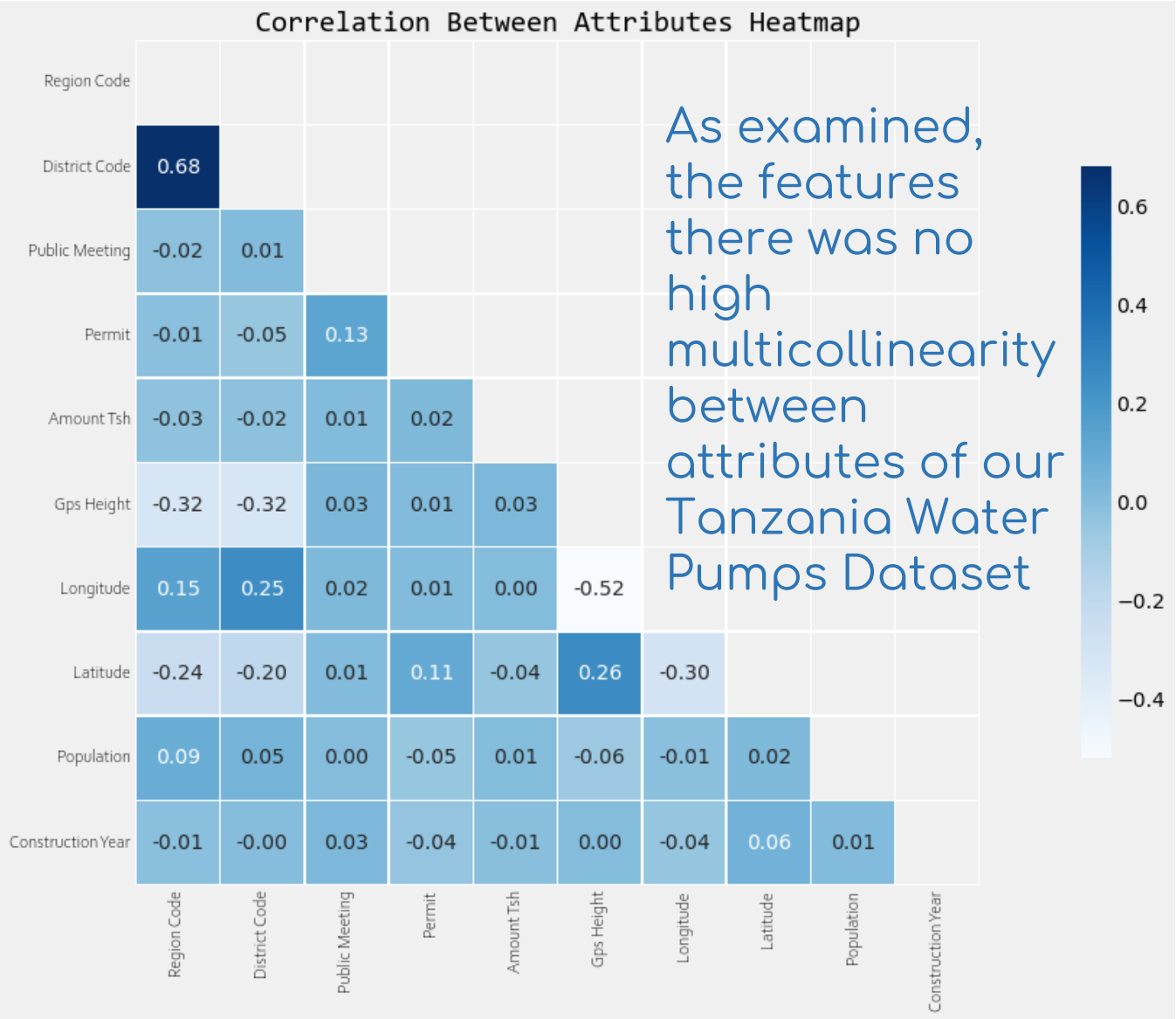






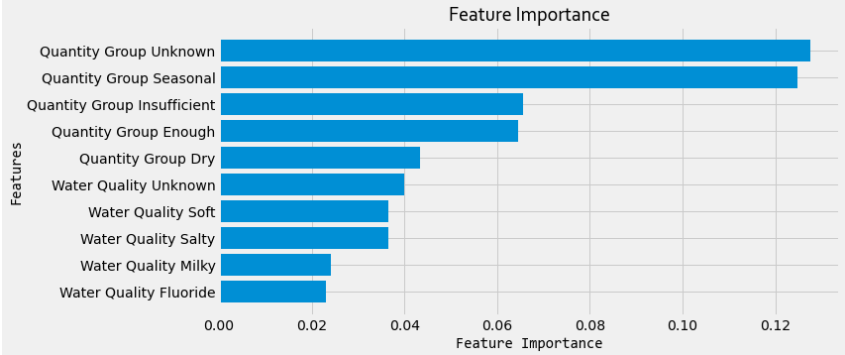


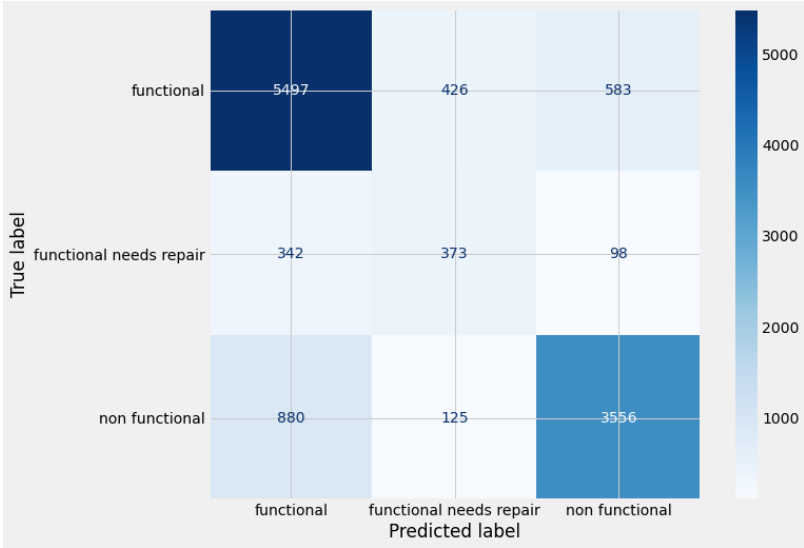




As examined, the features there was no high multicollinearity between attributes of our Tanzania Water Pumps Dataset

MODELING





	Model	Train Accuracy Score(%)	Test Accuracy Score(%)
0	Baseline Decision Tree	99.0	76.0
1	Second Decision Tree	99.0	75.0
2	Baseline Random Forest Classifier	99.0	79.0
3	Baseline Gradient Boost	99.0	71.0
4	XGBoost Classifier	86.0	77.0
5	Random Forest Classifier-Grid Search	98.0	79.0
6	XGBoost Classifier-Grid Search	95.0	79.0
7	Final Model-Random Forest Classifier	98.0	79.0

The Model Finally Achieved an Accuracy Prediction of *79.41%*

CHALLENGES

- The data had a lot of gaps that were either not filled or got filled but in a very general way.
- For such dataset, there were a high number (about 3%) of the water pumps that had not been mapped in Tanzania. This brought a challenge in determining where they were spatially supposed to be located. If this results are to be relied upon, such water pumps are supposed to accurately located.

RECOMMENDATIONS

- When gathering the data, it is important to use, among other reliable instruments, a high precision GPs. This would help the technical team visiting the water pump have an easy time getting their way around.
- Data cleaning and verification while in the field would be so crucial to ensure that valid results are taken back to the office for analysis.
- The government or the responsible agencies (e.g., the Ministry of Water) should put up a task force that ensures periodic data update to stay up-to-date with the requirements of all the water pumps and the type of care they need.
- The data collection should be focused more on the nonfunctional water pumps.
- A research to understand what mainly caused breakage.

END

GitHub Repository