Water as a Basic Human Right:

An exploration into the maintenance of Tanzanian water wells

By Micah Shackelford

Business Problem

Question: How can we help solve the water shortages in Tanzania?

Goals:

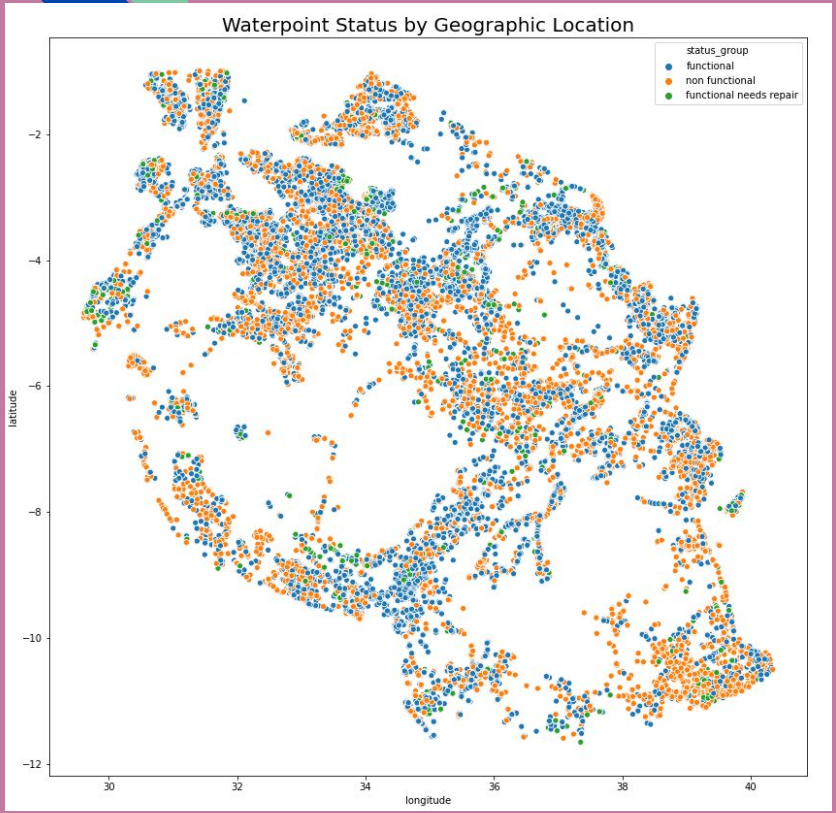
a. Figure out a way to improve are methods in identifying non functioning water wells
b. Detect key features to focus on in identifying these wells

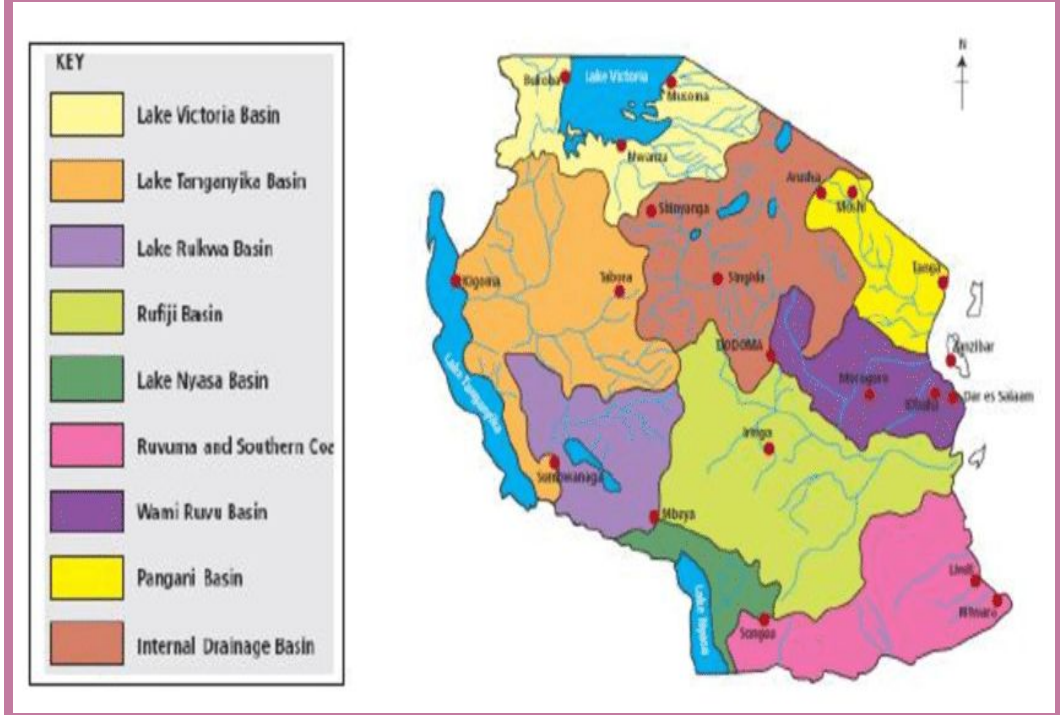
The Data

- Data comes from waterpoints all across Tanzania

- 60,000 different waterpoints are included and are classified into “functioning”, “non functioning”, and “needs maintenance”
- Data includes many different features of the wells including geographic location, water source, water quality, pump type, construction year, etc.

Location

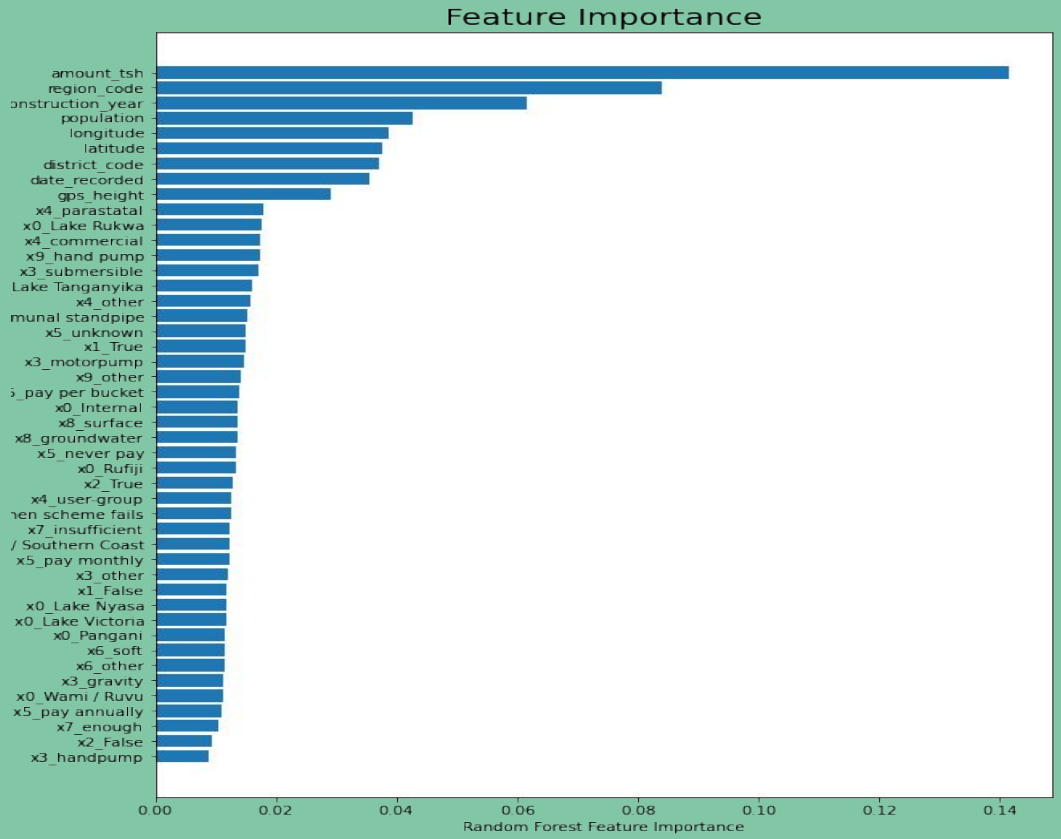

Waterpoint Status by Geographic Location
status_group
functional
non functional
functional needs repair
latitude
longitude
−2
−4
−6
−8
−10
−12
30
32
34
36
38
40


KEY
Lake Victoria Basin
Lake Tanganyika Basin
Lake Rukwa Basin
Rufiji Basin
Lake Nyasa Basin
Ruvuma and Southern Coa
Wami Ruvu Basin
Pangani Basin
Internal Drainage Basin
N
Lake Victoria
Mwanza
Shinyanga
Arusha
Singida
DODOMA
Zanzibar
Dar es Salaam
Iringa
Mbeya
Songea

The Model

Type	Random Forest Classifier
Number of Features	38(20)
Accuracy Score	Training: 81.52% Test: 74.65%
Important Features	Geographic Location, Altitude, Year Constructed, Population

Most Important Features



Summary

1. In this model we can accurately predict the current state of the well 75% of the time
2. The most important feature to look at are geographic location, altitude, year constructed, and population

Recommendations

Look into these water points with low amounts of access to water. Why are these not getting the water they need? If we can give these wells better access to water, we should be able to solve a large portion of the problems.

There is definitely a pattern with location and the status of the wells. We saw this in the visualizations and in our models. Try to find why this is. Is it a problem with the local regulations or something larger?

The population surrounding the wells also seems to be important. A lower population probably means that there are less regulations and maintenance. These wells are still needed though. The government needs to focus on getting these wells back online.

Future Analysis

- Explore different types of models to see if we can improve our accuracy in predicting the state of wells. Especially interested in XGBoost
- We were much more accurate in predicting “functioning” and “non functioning” wells. Maybe we can remove the “needs maintenance” classification, so we can focus on the more pressing issues
- This data only go up to 2013. Update the data with more recent well information

Thank You!

Micah Shackelford

Github: https://github.com/shackemn