Pump it Up: Data Mining the Water Table

Classifying waterpoints in Tanzania

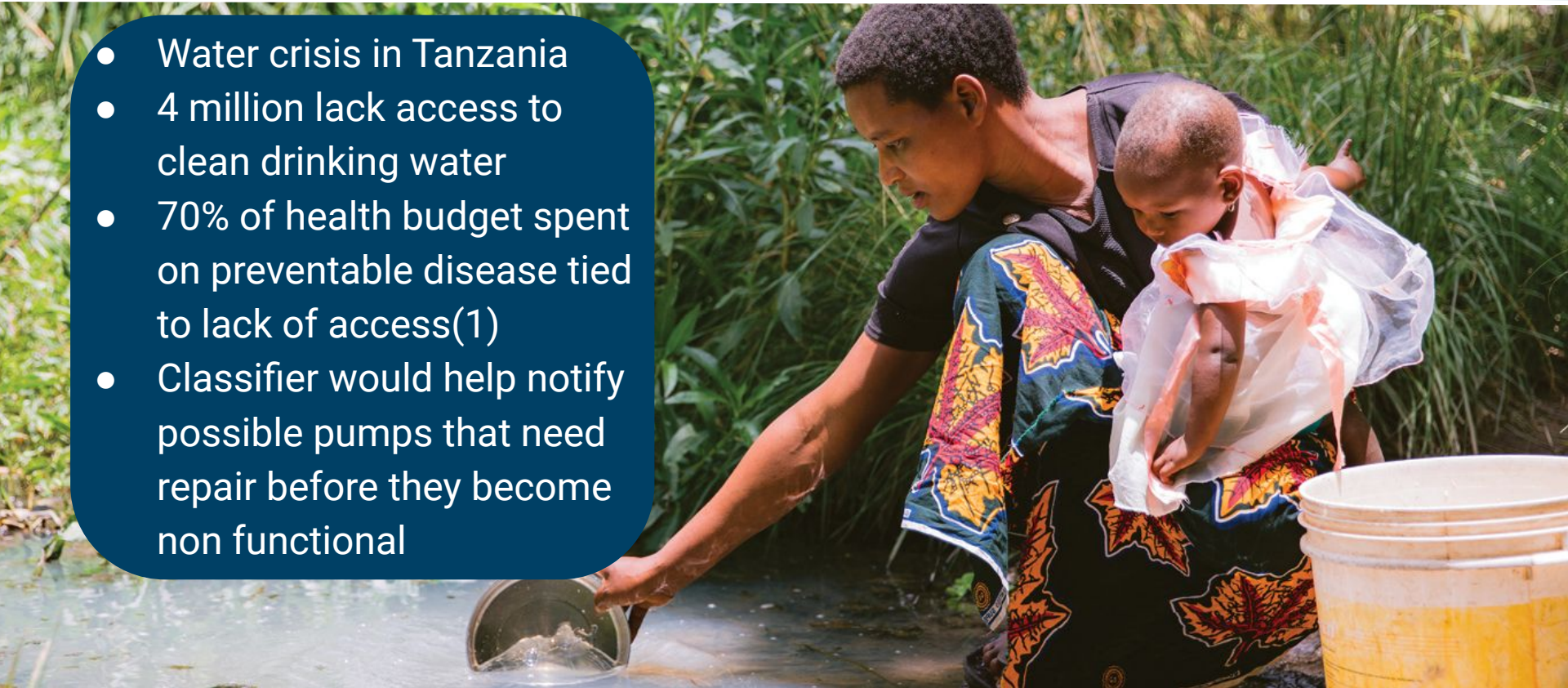
Gustavo Chavez

The Challenge

- Competition hosted by DataDriven.org
- Data gathered by Taarifa and Tanzanian Ministry of Water
- Ternary classification problem

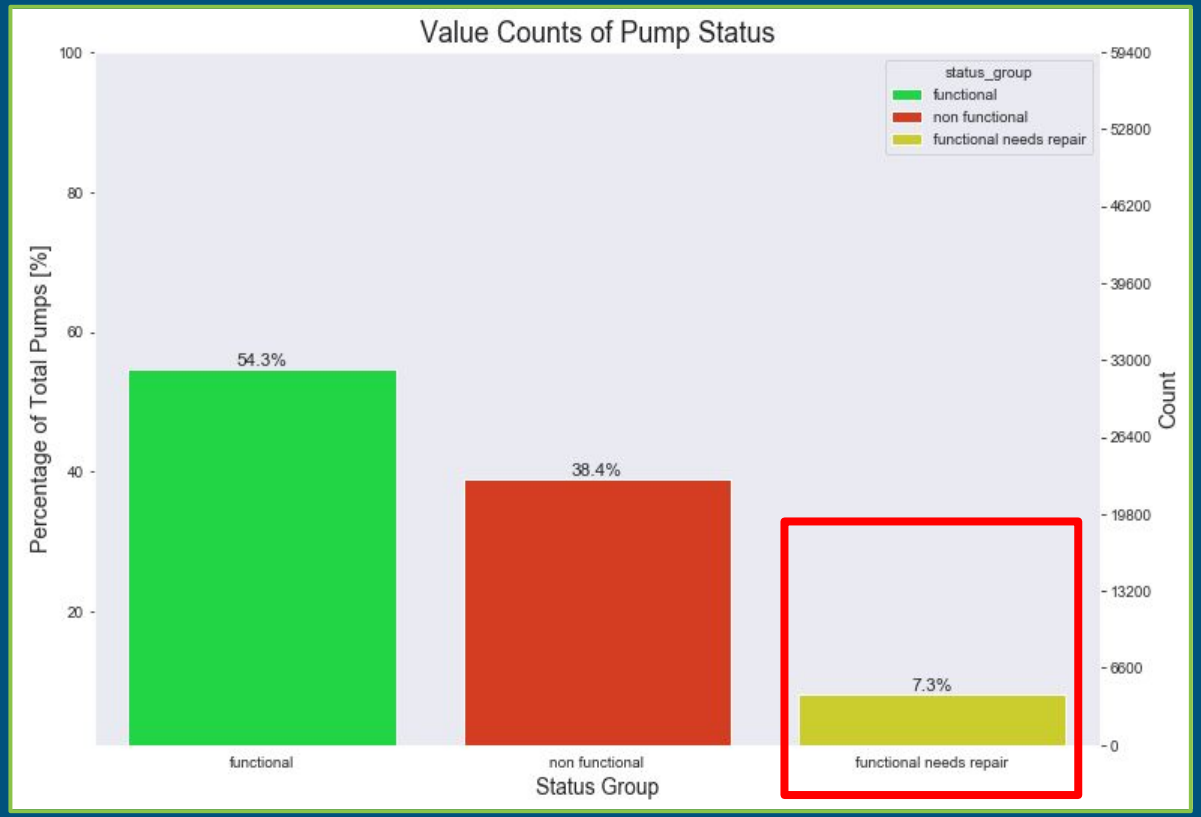
Purpose of Creating Predictive Model

- Water crisis in Tanzania
- 4 million lack access to clean drinking water
- 70% of health budget spent on preventable disease tied to lack of access(1)
- Classifier would help notify possible pumps that need repair before they become non functional

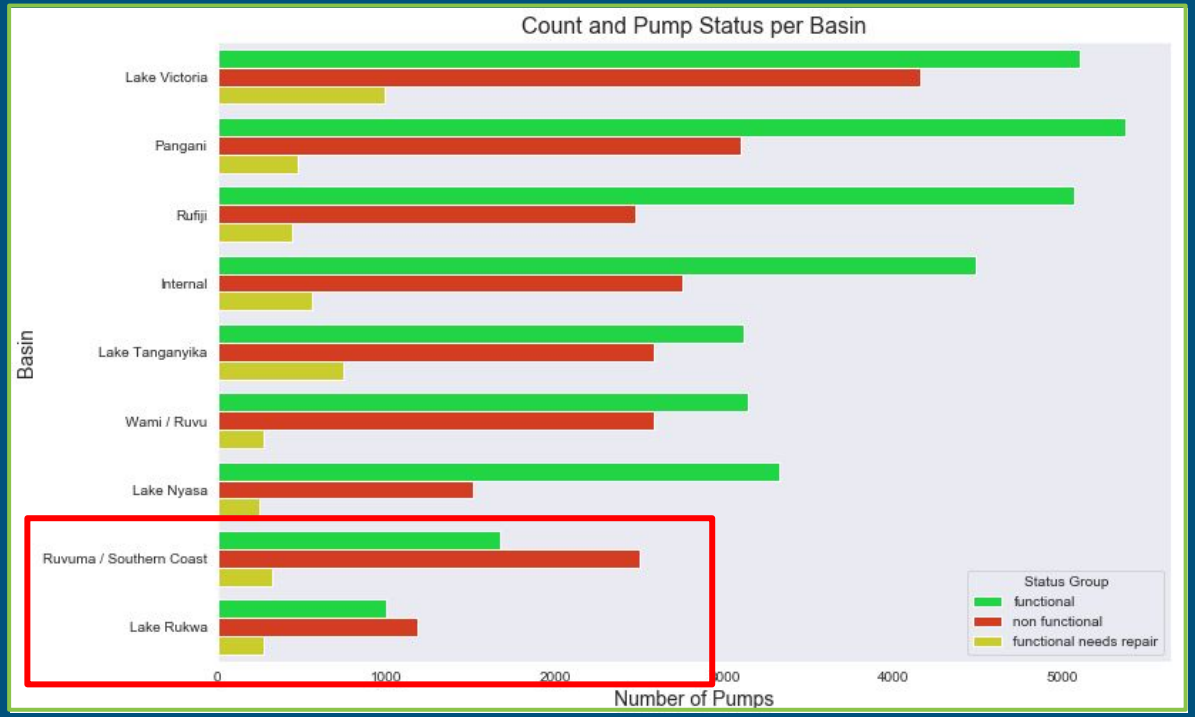


The Water Pumps

- Over half of all pumps are functional
- One-third of all pumps do not work
- Knowing which pumps need repair before they stop functioning helps reduce water scarcity.



What is the Spread of Pumps in each Basin?

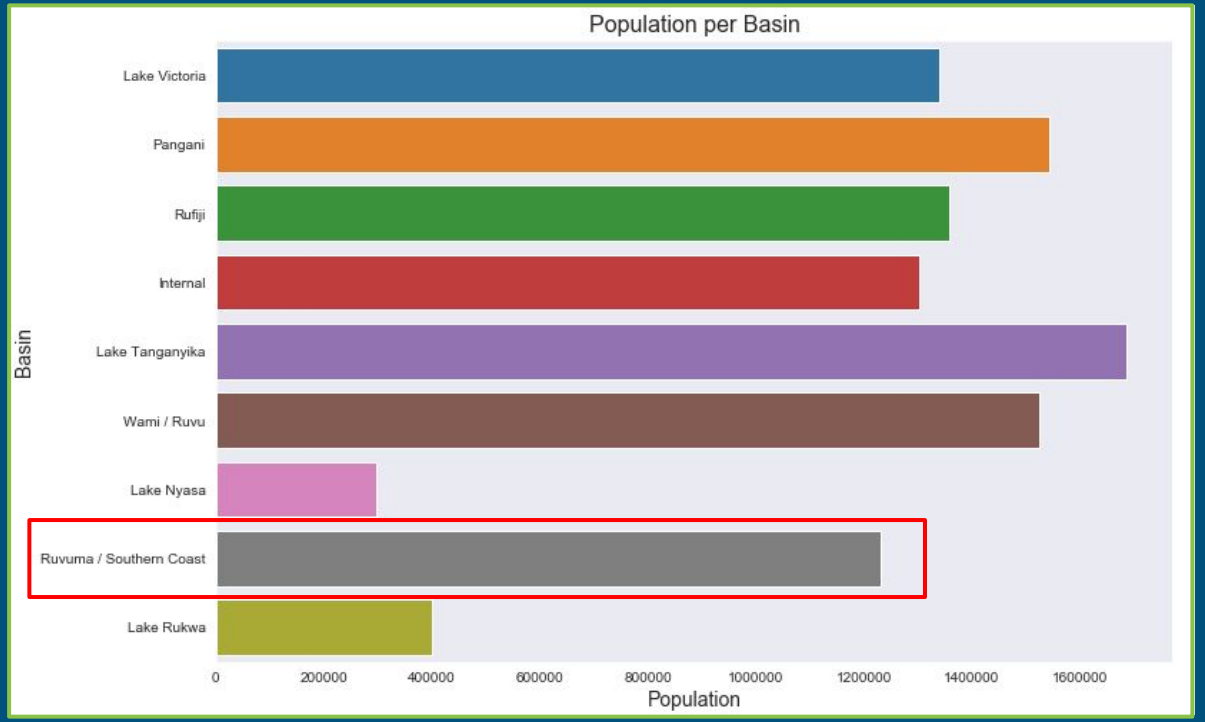


- Most regions have more functioning pumps than non-functioning ones
- Must improve pumps that need repair to improve ratios
- Certain Basins appear underserved

Basins in Need

Ruvuma / Southern Coast contains more non-functional pumps than functional pumps as well as supporting a high population.

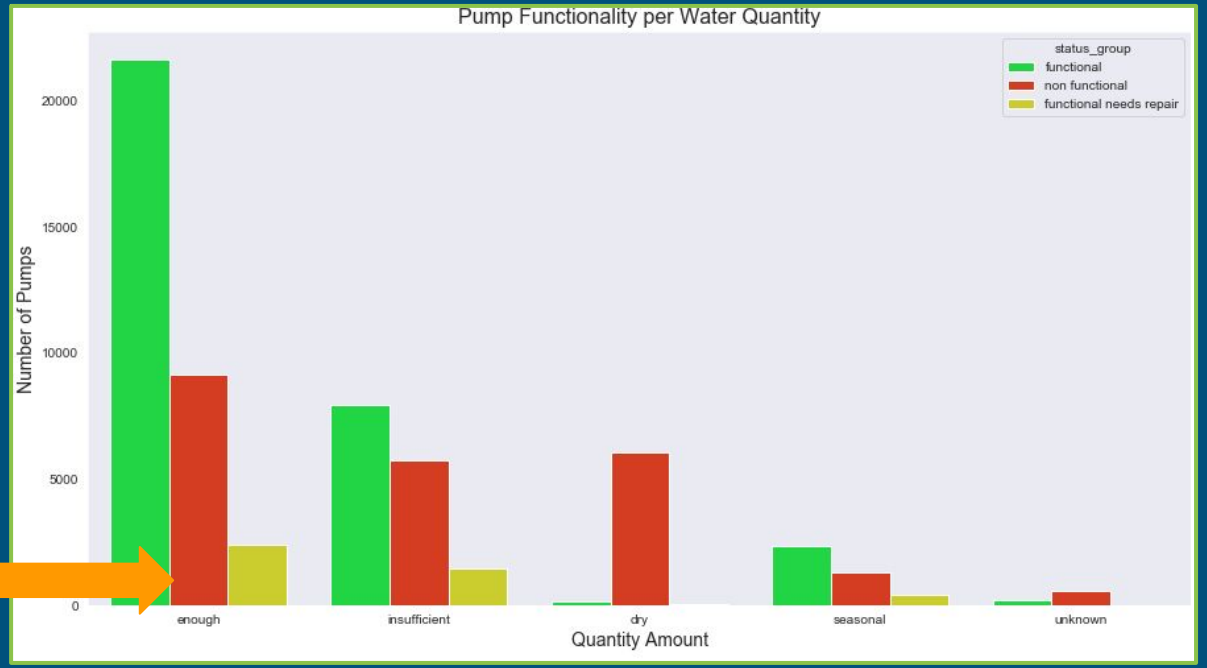
Recommendation: Have improved water access be a priority for regions within this basin



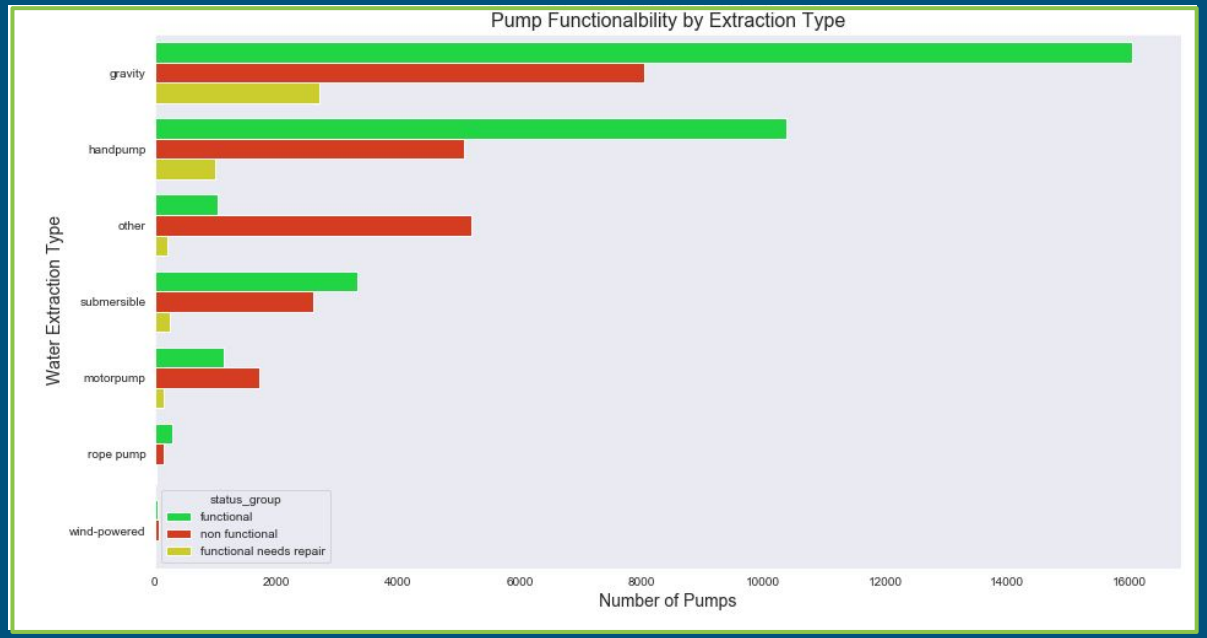
Which Pumps to Consider

- Pumps that already produce enough best candidates to be fixed if not already operational

- “Dry” non-functioning pumps not worth fixing

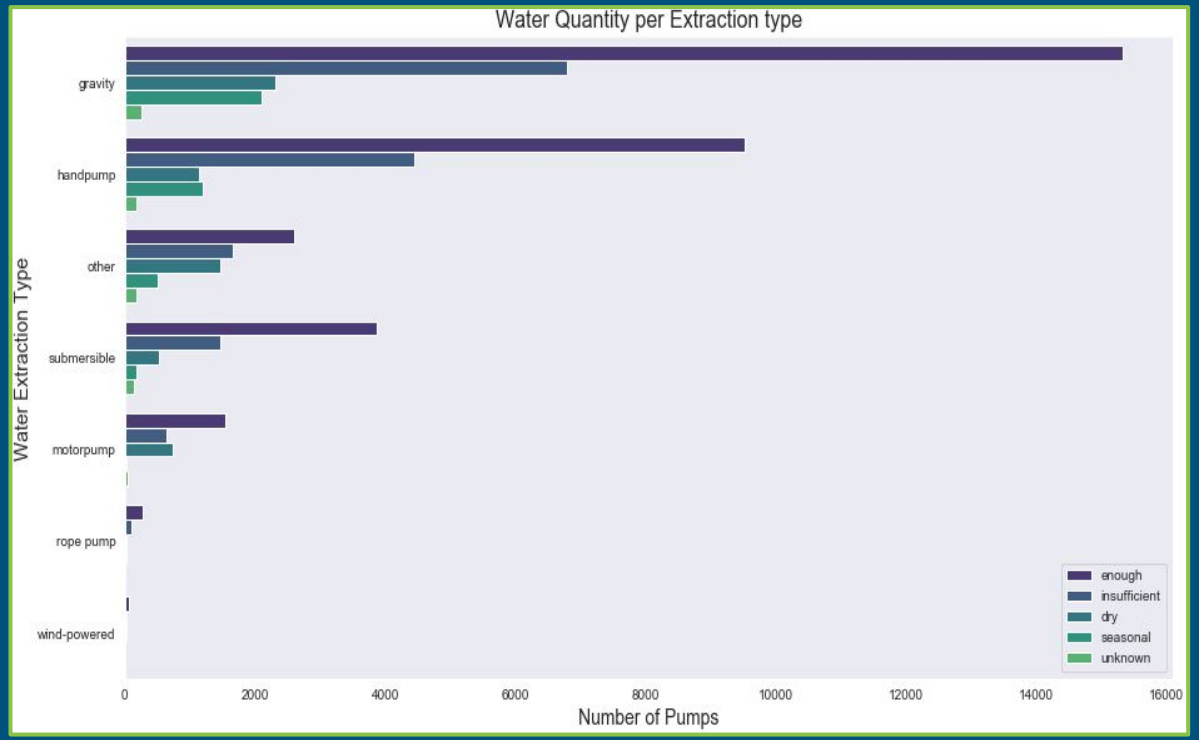


Best Extraction Methods



- Gravity pumps and hand-pump wells are the most common and most functional.
- Other types of pumps are not plentiful enough to comment on.
- Most pumps in ‘Other’ do not work.

Best Extraction Methods



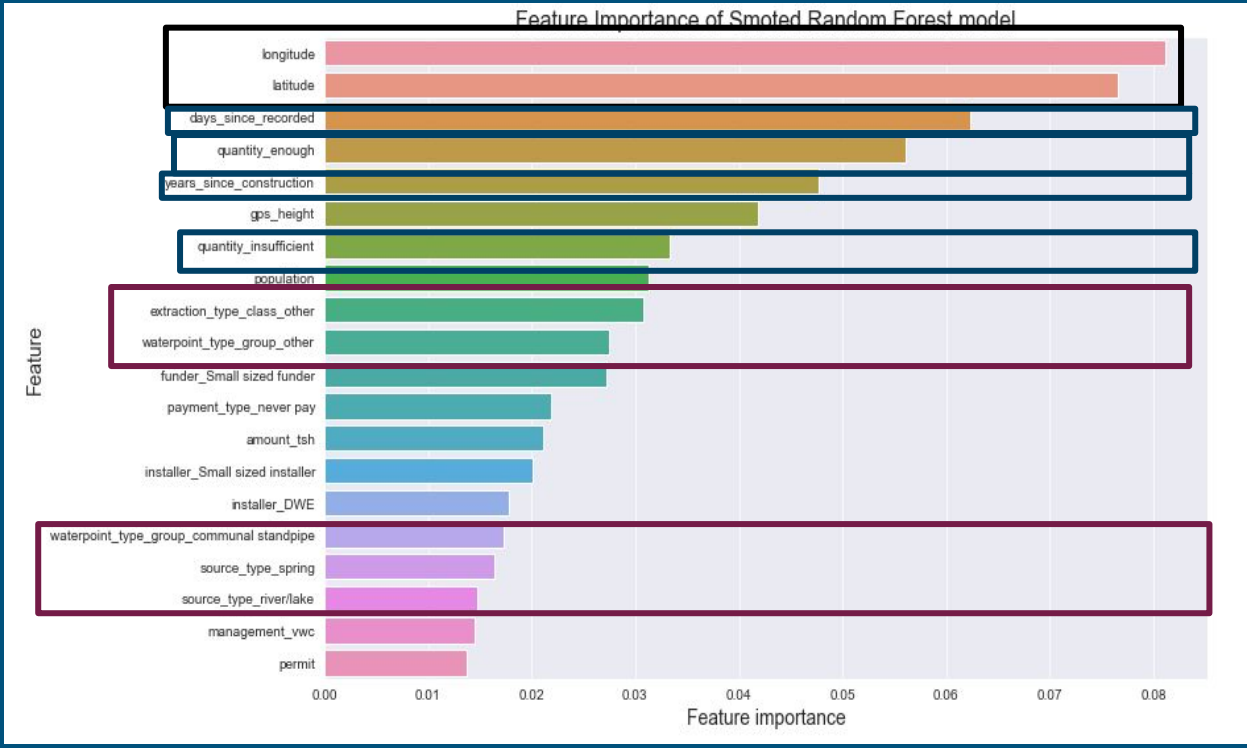
- Majority of gravity pumps provide plenty of water
- Correlation between functional and enough, as well as non functional and insufficient
- Gravity and Handpumps are the best methods both in terms of functionality and quantity

79.69%

The model created was able to accurately predict almost 80% of the functionality of the pumps.

Most Important Features

- Location mattered
- Days since recording and years since building
- Quantity of water
- Type of water and method of extraction also mattered a lot.



Recommendations

- Seek to build new water infrastructure in basins that are lacking water supply.
- Attempt to find sources where building a hand-pump or gravity pump are possible.
- For existing water infrastructure, seek to repair pumps that have enough water quantity so as to allow more water for that population.

Future Work

- Examine the effects of date recorded to check for seasonality
- Analyze how construction year affects pump functionality
- Stack multiple strong models together for stronger predictive power
- Find out what contributes to the high number of non functional pipes in Ruvuma / Southern Coast Basin.

Thank you for listening!

Special Thanks to:

Flatiron cohort lead Abhineet Kulkarni

Educational Coach Dara Paoletti

Fellow members of my Flatiron cohort.

References

(1) https://www.unicef.org/tanzania/what-we-do/wash

Image Sources

Slide 2 - https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/

Slide 3- https://lifewater.org/blog/tanzania-water-crisis-facts/