

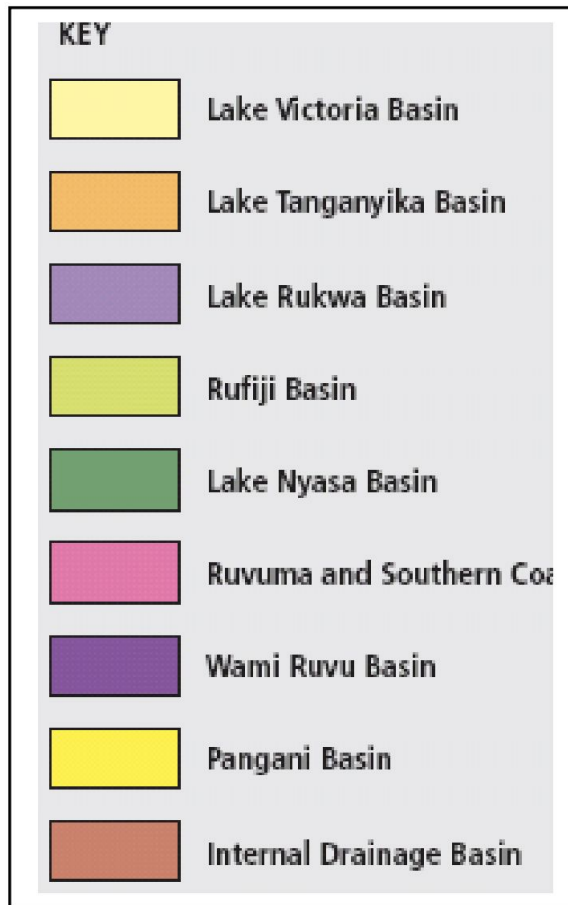
Pump It Up

Data Mining the Water Table
Bobby and Dan

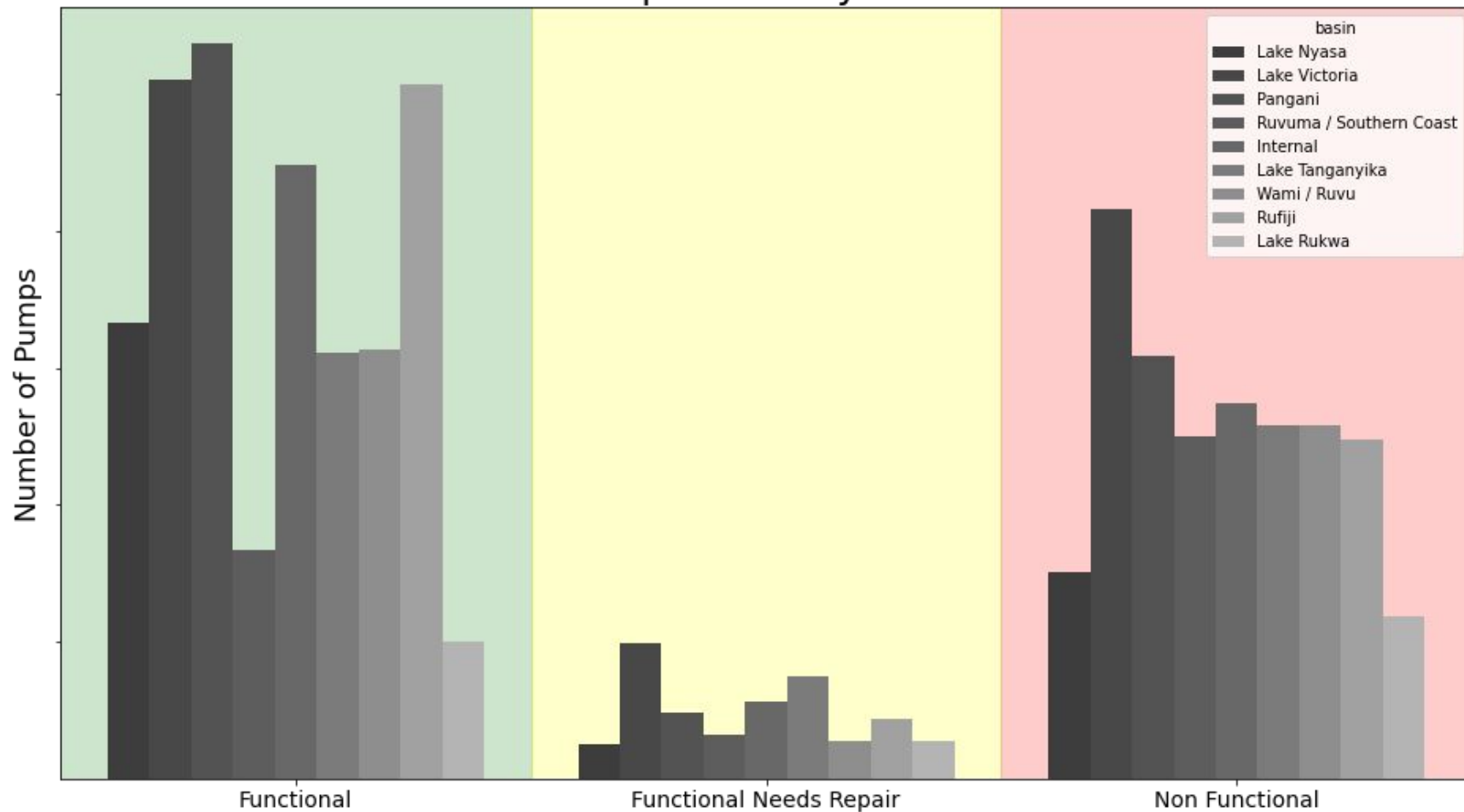
“Water is a shared common resource fundamental to life and in sustaining the environment and plays a central role in the social and economic development of Tanzania.”

- The United Republic of Tanzania Ministry of Water and Irrigation
National Water Sector Development Strategy

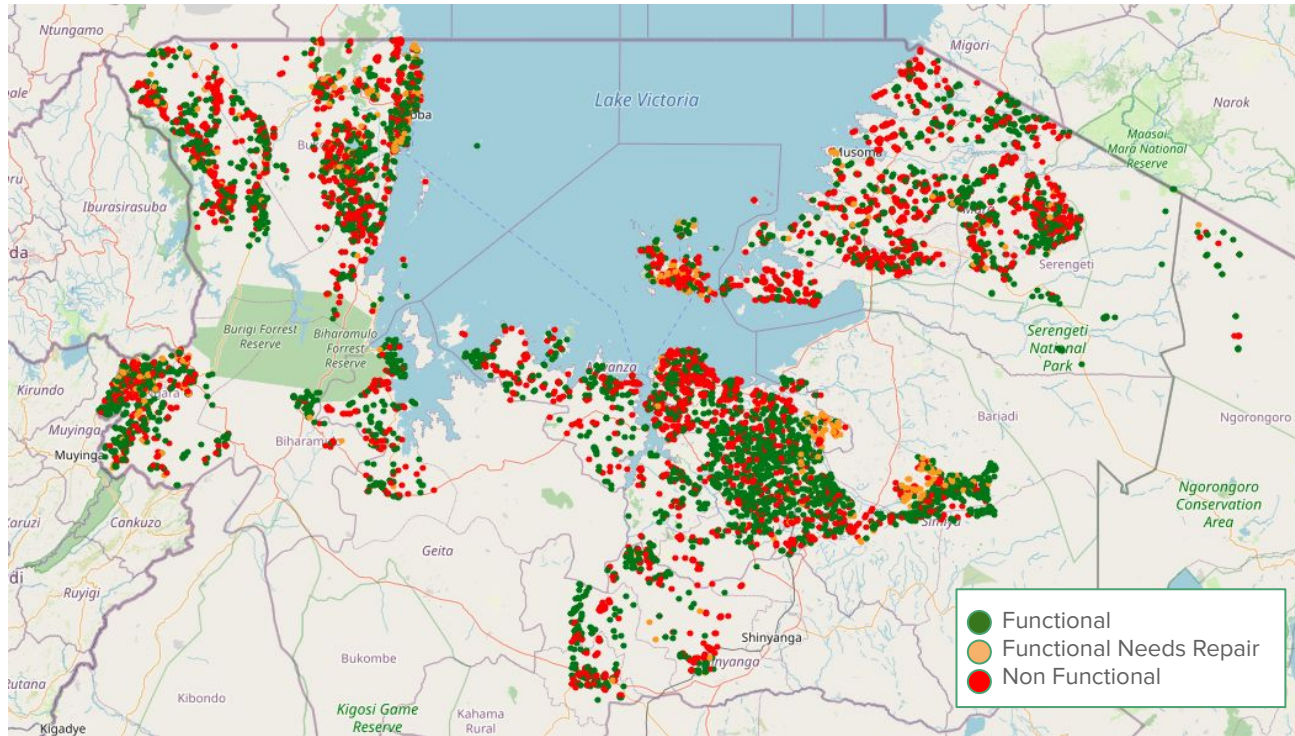




Pump Status By Basin



Water Points in the Lake Victoria Water Basin



Goal

Create a model to predict whether or not a pump needs repair.

Get clean water to residents in rural regions.



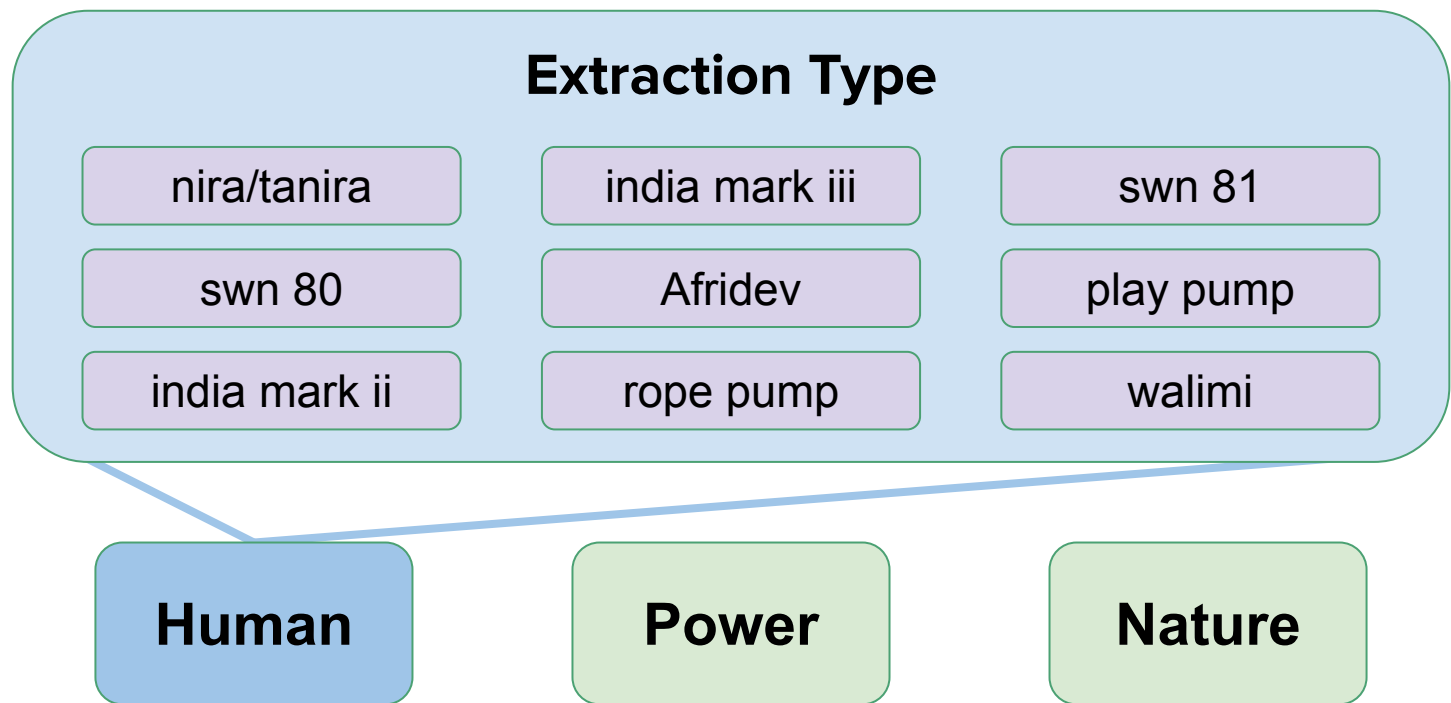
The Dataset

There are a number of inconsistencies in the data for a variety of reasons, including:

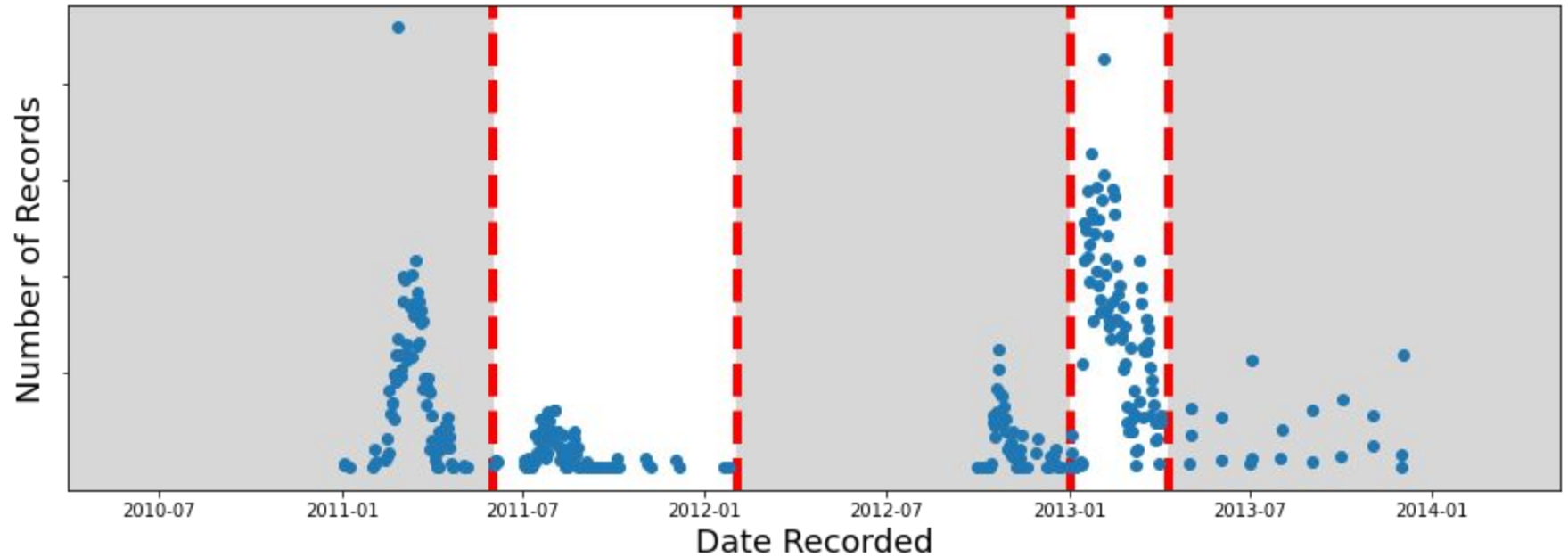
- Missing data
- Subjective fields
- Form modifications
- Redistricting
- Malfunctioning equipment
- Untrained operators



Predictors



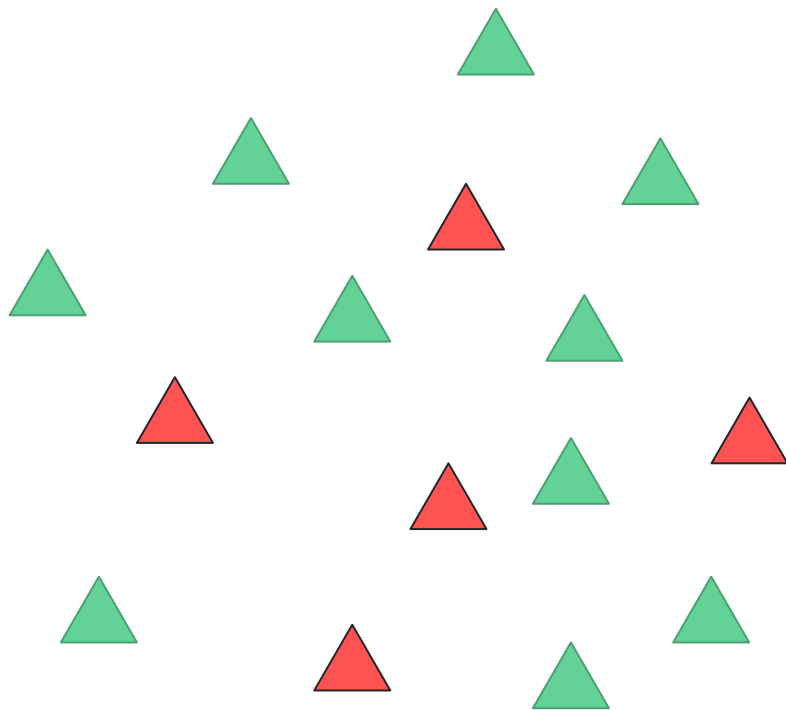
Phases of Data Collection



Best Model: Decision Tree Classifier

“Overlook rate” of 32%

Overall accuracy of 76% across all pump statuses.



Recommendations



- Repair water points in need
- Reduce maintenance costs

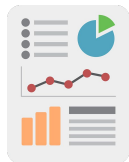
Next Steps



Tune the model



Collect more data



WPMS Dashboard

Contact Information

Bobby Williams

<https://github.com/bobbyiestofjos>

Dan Hales

<https://github.com/danhales>