



# Predictive Modeling for Tanzanian Wells

Alex Marshall, Jennifer Cobb, Nick Indorf

Water, water everywhere?



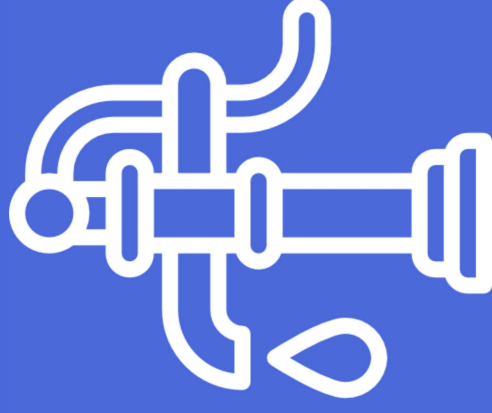
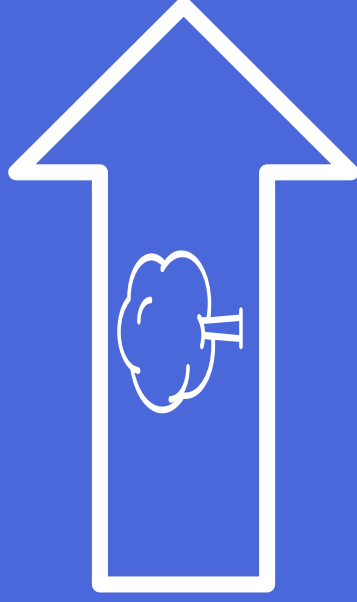


**Model Optimization**

Populations  
without water

Misallocation  
of labor

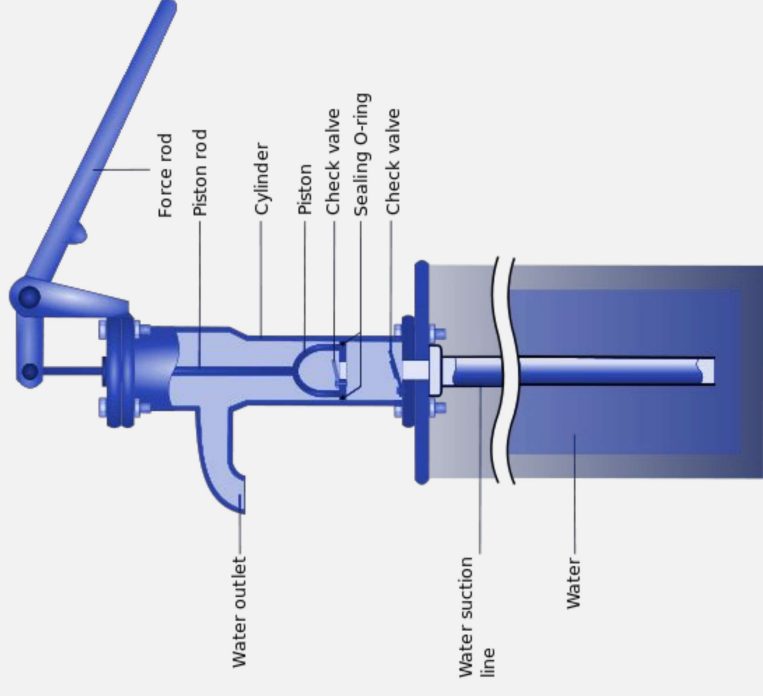
# Process





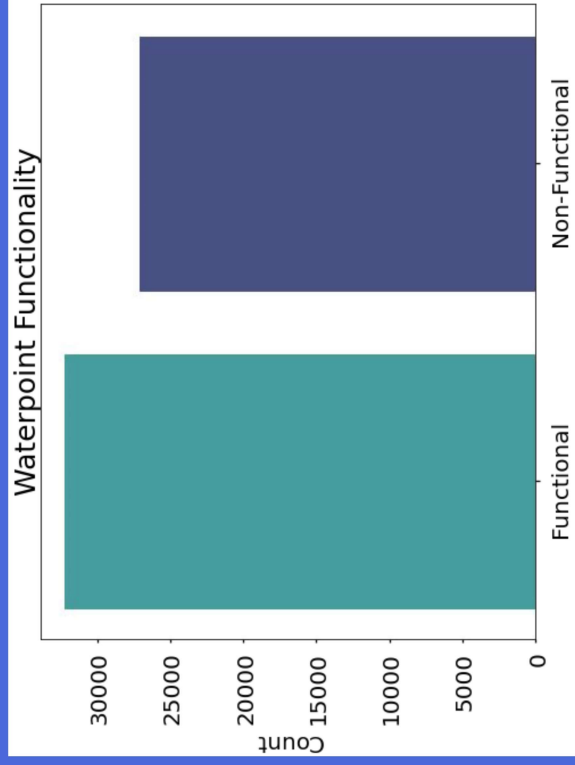
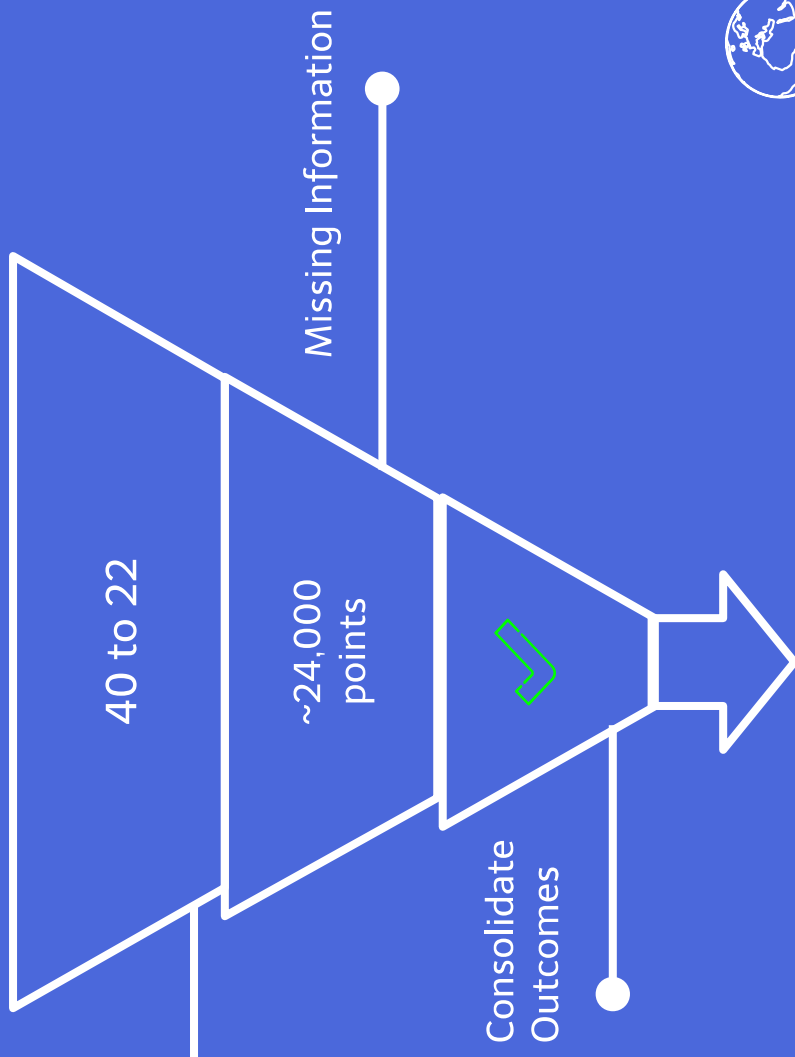
# Dataset Characteristics

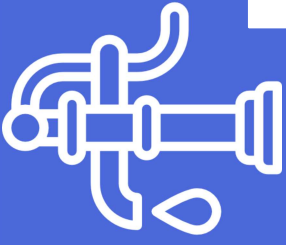
- ~59,000 water pumps
- Collected between 2002 - 2014
  - From Ministry of Water & Taarifa
- 31 categorical, 9 numeric features





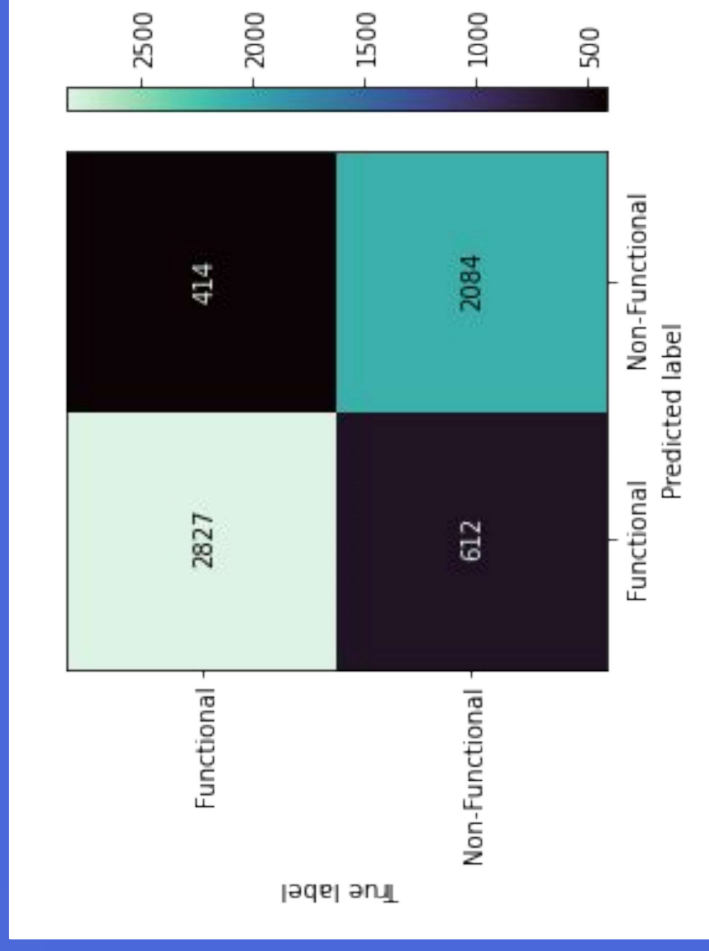
# Data Cleaning





# Modeling & Evaluation

- Based on Decision Tree
- 82.7% - accurate for all predictions
- 77.3% - accurate for truly non-functional
- 83.4% - accurate for predicted non-functional



# Next Steps

---

- Revise data collection processes
- Integrate external geodata
- Expand number of classifiers in model







Thank  
You