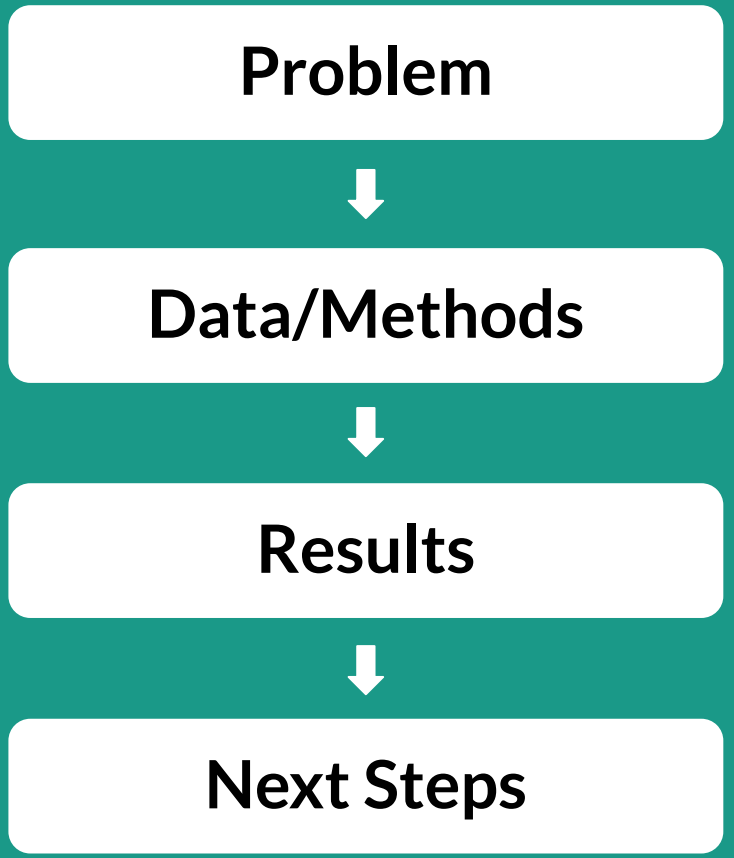
Predicting Waterpoint Functionality in Tanzania

Luke Dowker // Data Scientist
Flatiron School

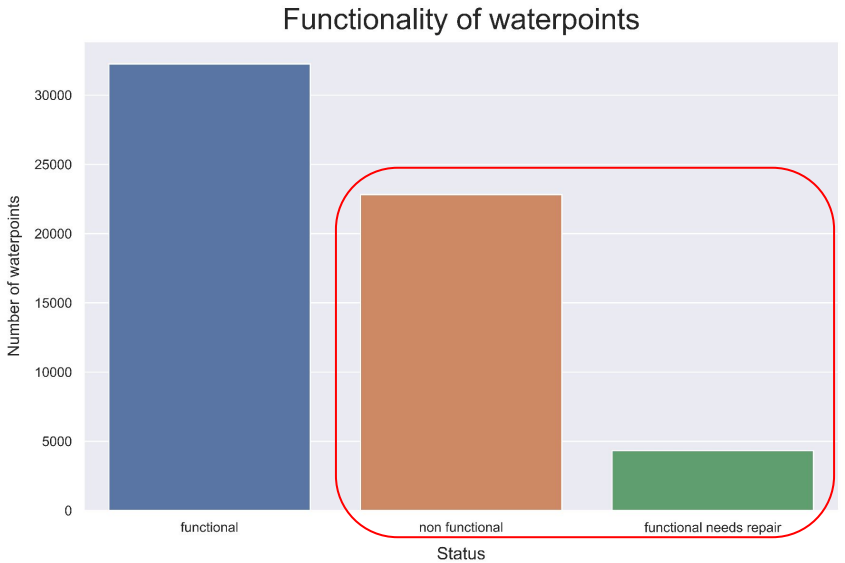
Today's agenda

79%
accurate



The Problem:

- **Which** waterpoints to investigate?
- **How** to predict functionality?



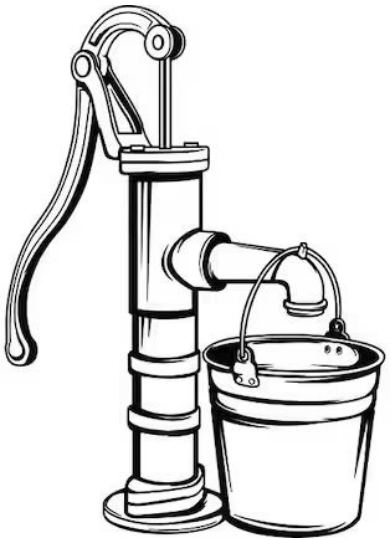
Predictors Used

- Source (lake, river, etc.)
- Water quality
- Year constructed
- Area population
- …and more

Pangani River

Results ✔

- **79% accurate**[1] on unseen data
  - Correct on **4 out of 5** predictions
- Identifies **84%** of non-functional waterpoints[2]



1. accuracy score
2. precision score

Limitations

- **Missing values -** year, amount of water, etc.
- Class **imbalance**
- …but these can be improved!

Next Steps...

- Improve **collection practices**
- Dive deeper into **individual predictors...**
- ...to **build better waterpoints!**

Thank you!

Full analysis, including Jupyter notebook, available in the GitHub repository for this project.