Tanzanian Water Well Data

For The Ministry of Agriculture, Food Security, and Cooperatives

Produced by : Rachel Edwards and Svitlana Glibova

Outline

- Key Issues
- Target & Justification
- Data
- Significant Features
- Modeling & Results
- Goals

Key Issues

- Over 50% of Tanzania's GDP is reliant on livestock and agriculture
- 80% of Tanzanian employment is directly related to agriculture and livestock
- Management of water resources is crucial
  - fragmented planning
  - climate change
  - food self-sufficiency and malnutrition

Sources

https://www.ajol.info/index.php/tjags/article/view/153336/142927

https://na.unep.net/atlas/africaLakes/downloads/posters/Lake_Victoria.pdf

https://www.maji.go.tz/

Target

- Build a statistical model that predicts pump functionality
  - Predict pumps that need repair (positive class)

- Minimize false negative - maximize recall score

Justification

- Predicting that a well is functioning when it actually isn't could put peoples' lives and livelihoods at risk
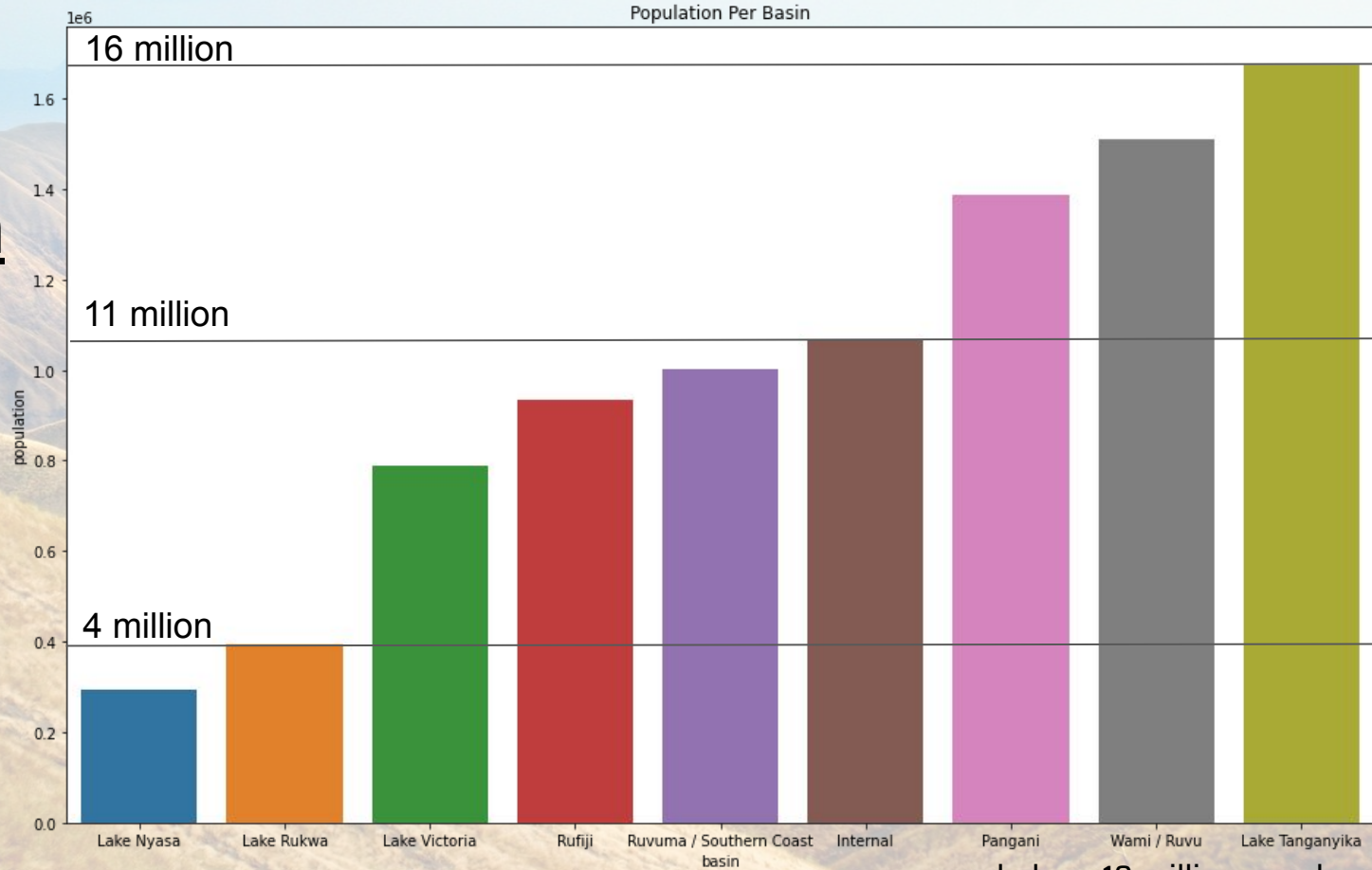
Map of Basins

Data

Total Population
58.01 million

Population Per Basin

16 million

11 million

4 million

scaled per 10 million people

Data

Lake Victoria
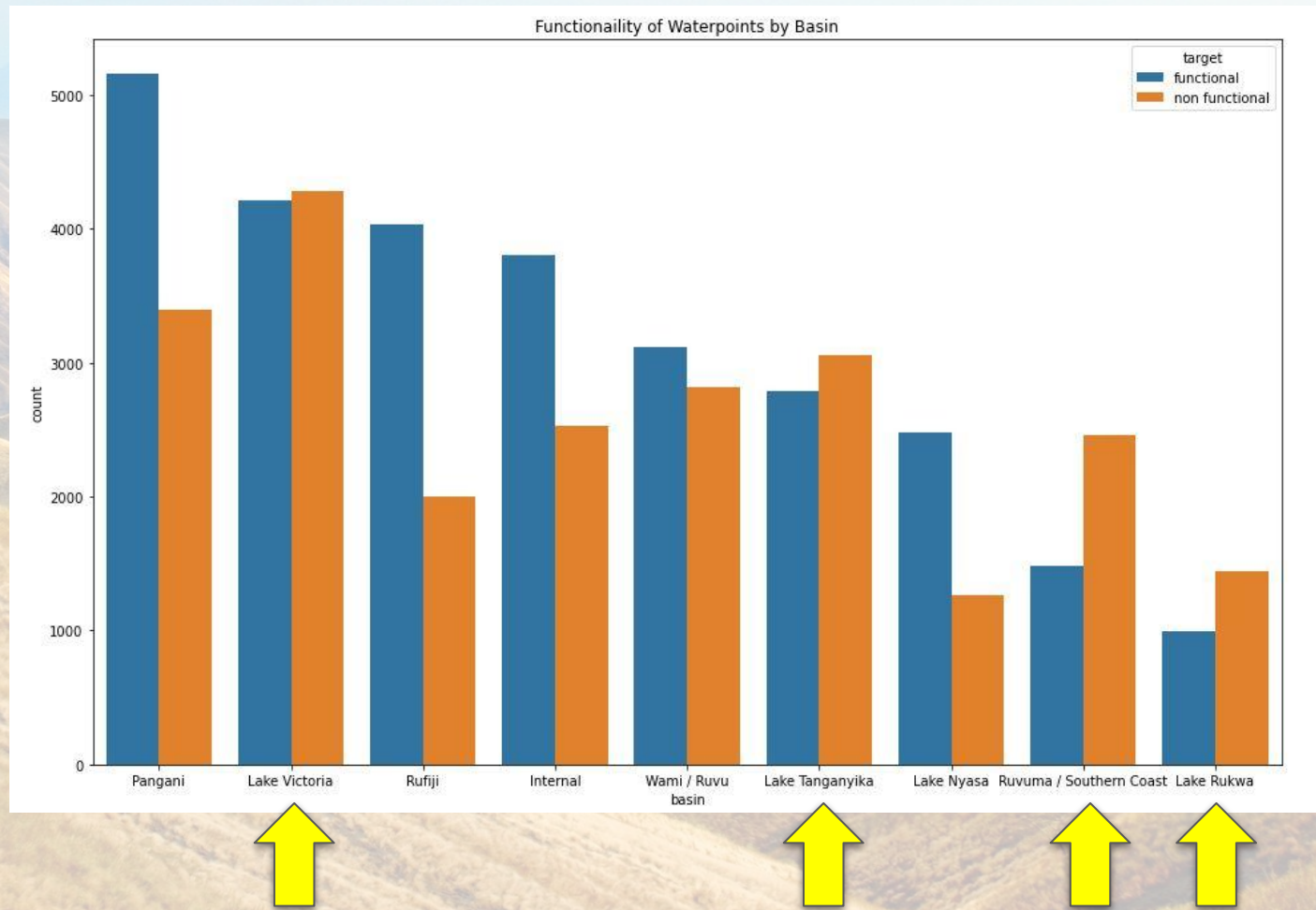- Pop. 40 million*
  (incl. bordering
  countries)

Lake Tanganyika
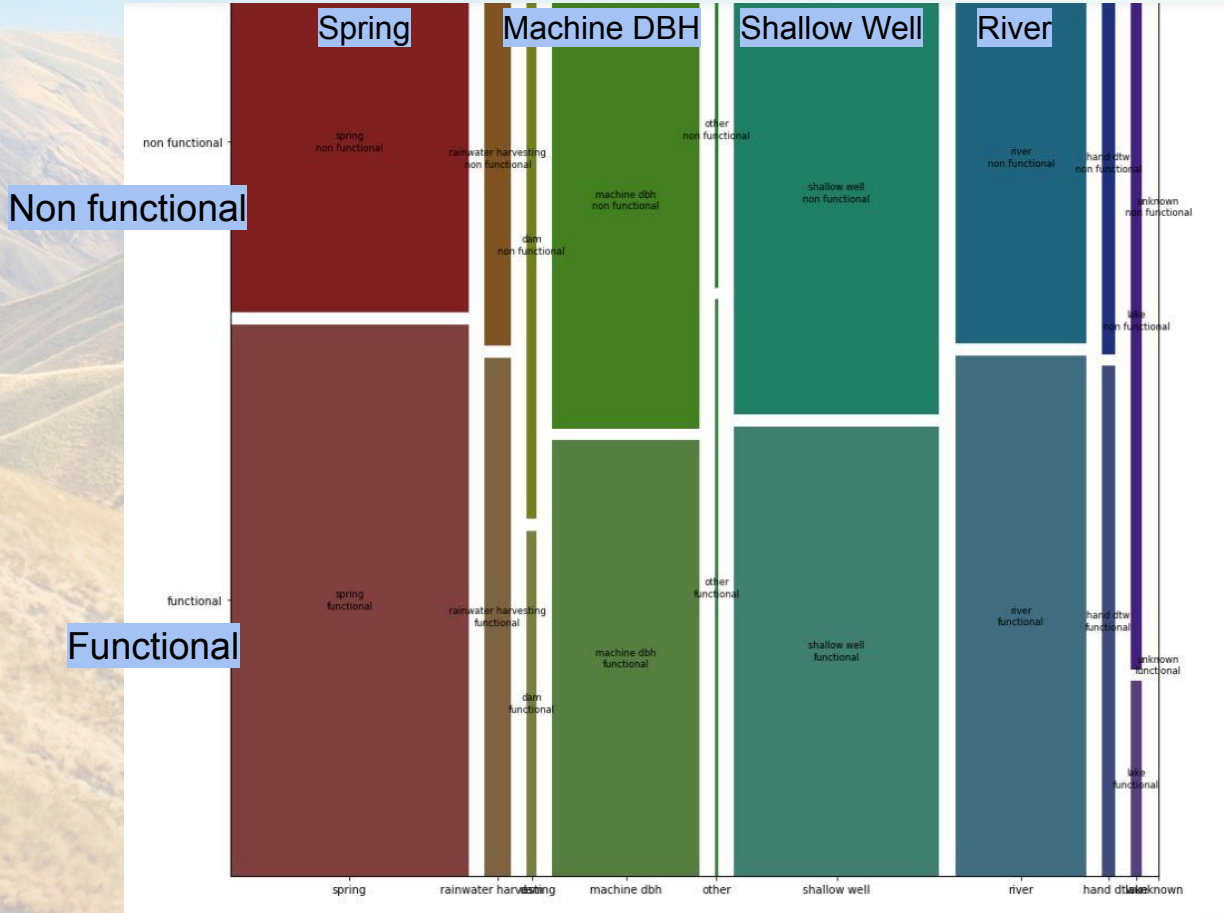- Pop. 10 million

Ruvuma Basin
- Pop. 1.37 million

Lake Rukwa
- Pop. 1.5 million

Functionaility of Waterpoints by Basin

Model-Specific Features

Water Source Data by Functionality

- - Construction Year
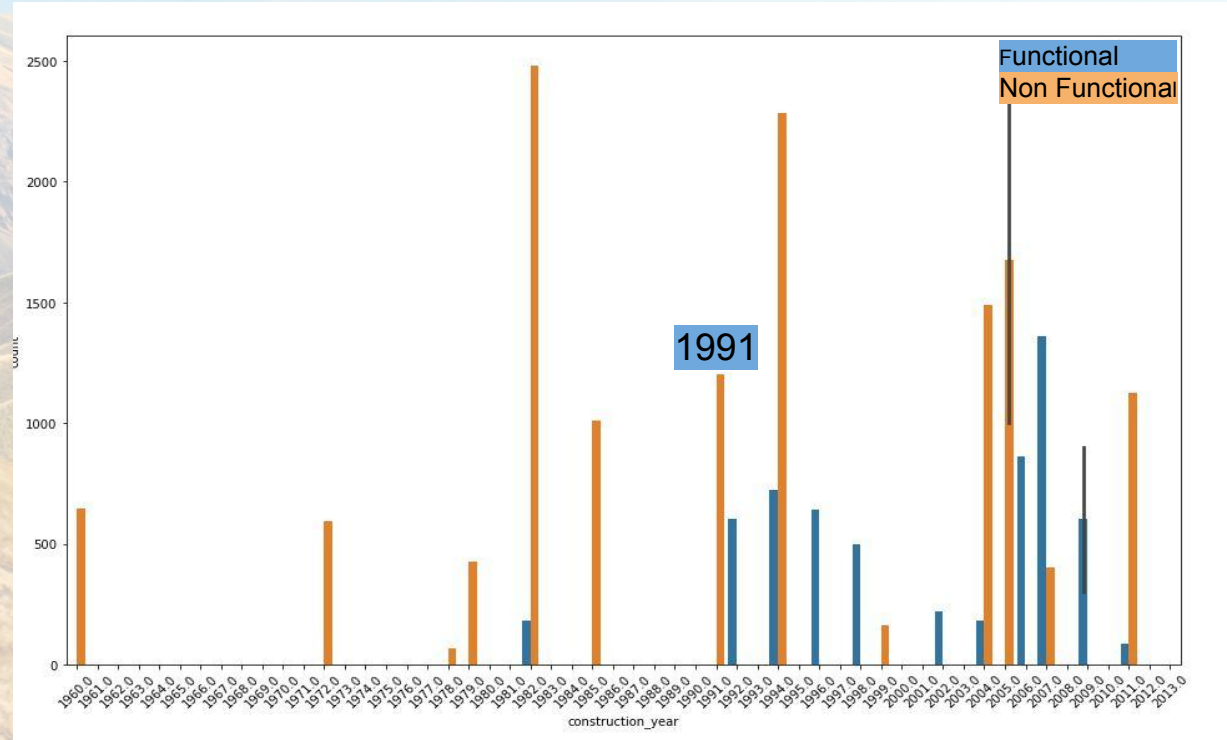- - Permit
- - Source

Model-Specific Features

Take away:
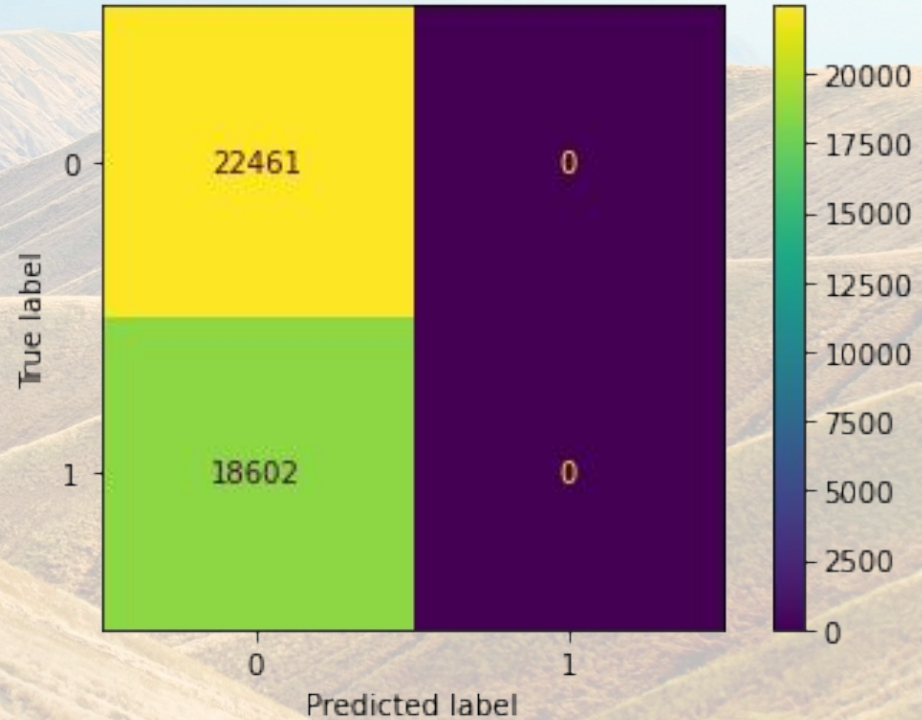-   Most pumps built before 1991 are non functional

Year Constructed by Functionality

Initial Model

Frequency

Recall: 0%

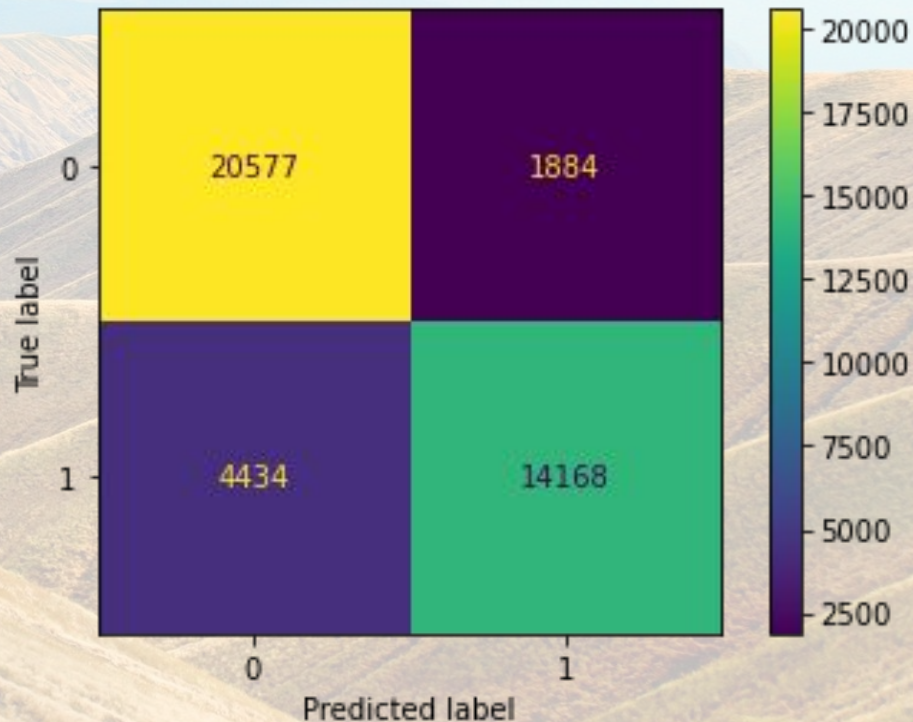
Accuracy: 54.6%
Precision: undefined

Modeling & Results

XGBoost Classifier

Recall: 76.4% train
69.8% test
(87% for negative class)

Important parameters
- Number of estimators
- Learning Rate
- Regularization

Precision: 81.4%
Accuracy: 79.1%

Goals

- Update pump infrastructure

- Cross-regional communication

- Creating stability in planning and policymaking

- Technological approach to water resource management

THANK YOU!

Rachel Edwards [LinkedIn](#) [GitHub](#)

Svitlana Glibova [LinkedIn](#) [GitHub](#)

Appendix

Lake Victoria
- This population does not represent the population of Tanzania but the population of those in the region of access of Lake Victoria

Internal Basin
- "a drainage basin that normally retains water and allows no outflow to other external bodies of water, such as rivers or oceans, but drainage converges instead into lakes or swamps, permanent or seasonal, that equilibrate through evaporation"