Tanzanian Water Well Prediction

Ryan Reilly & Kevin McDonough

Business Problem

Tanzanian Ministry of Water

Identify wells that are not functional
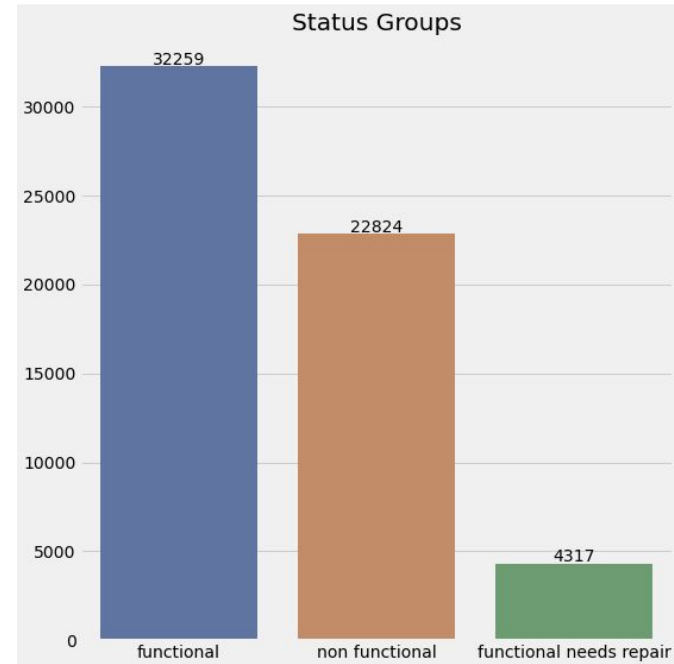
Identity features that predict non-functioning wells

Data

- 59,400 rows representing individual water wells in Tanzania
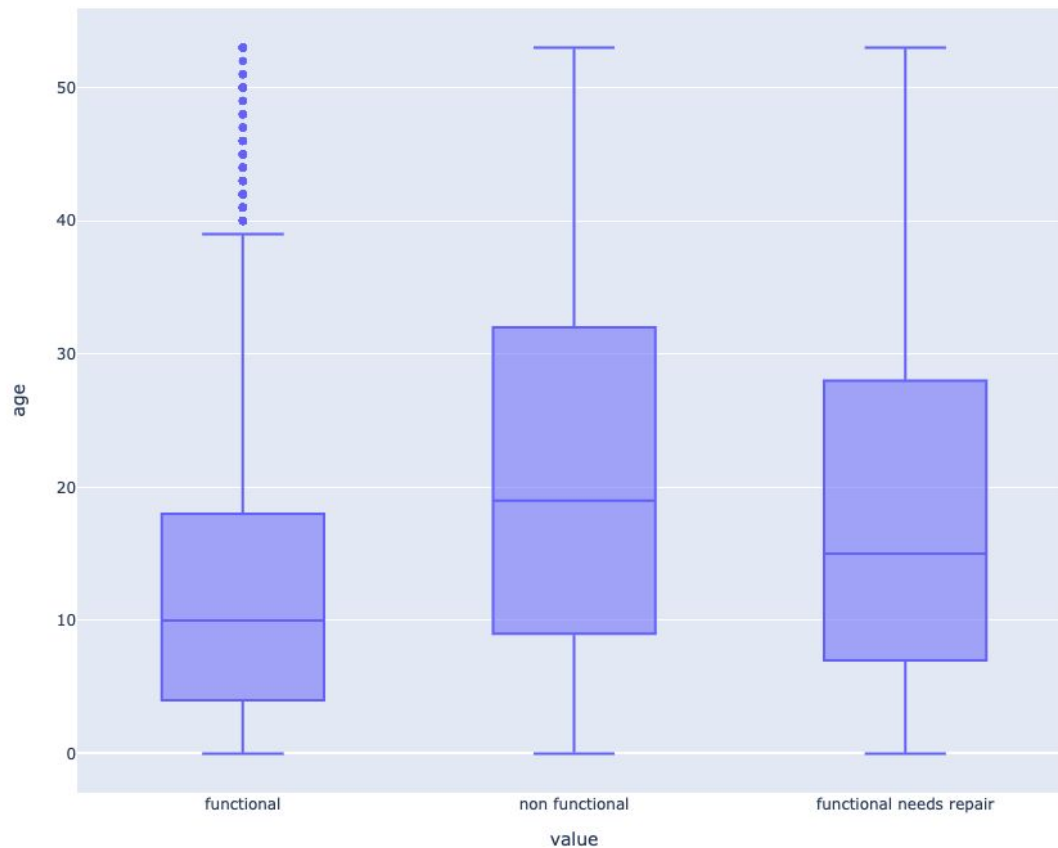- 41 columns

Target Categories

~54% **Functional**
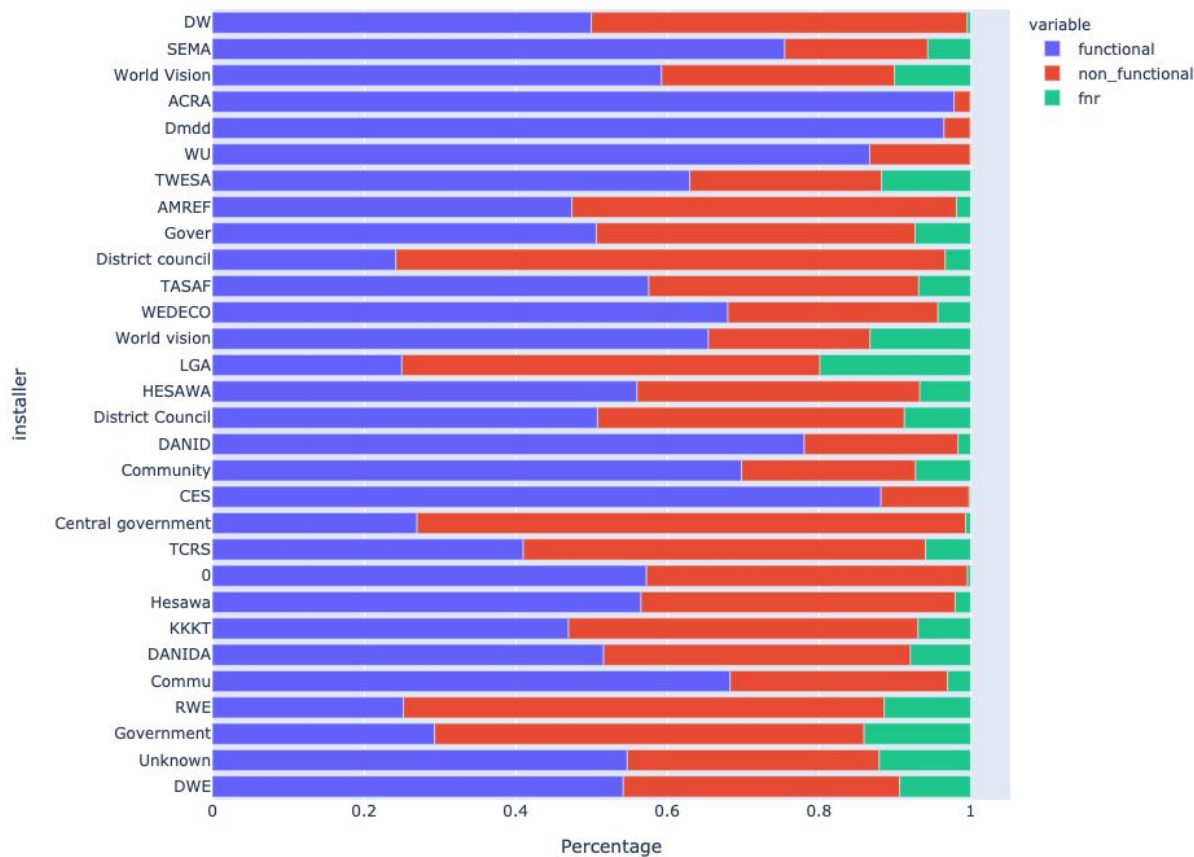
~38% **Not Functional**
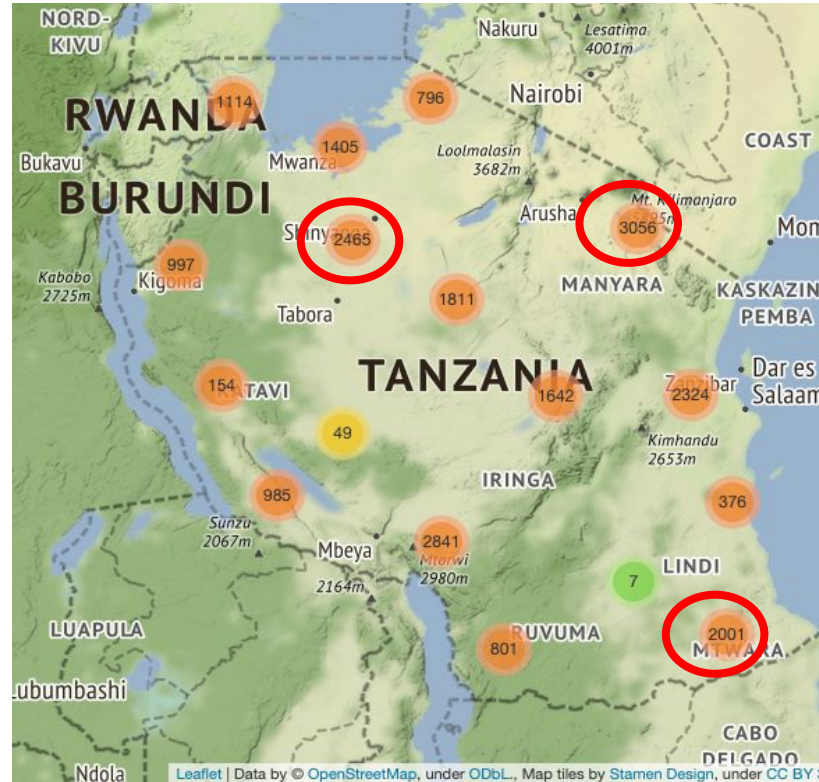
~7% **Functional Needs Repair**
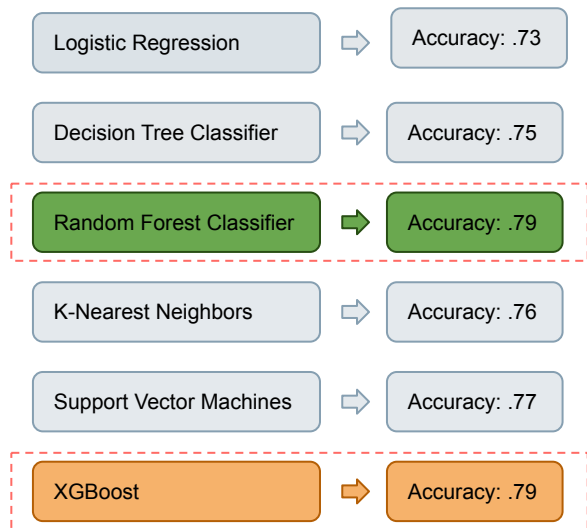
Status Group by Age

Top 25 Installers by Count
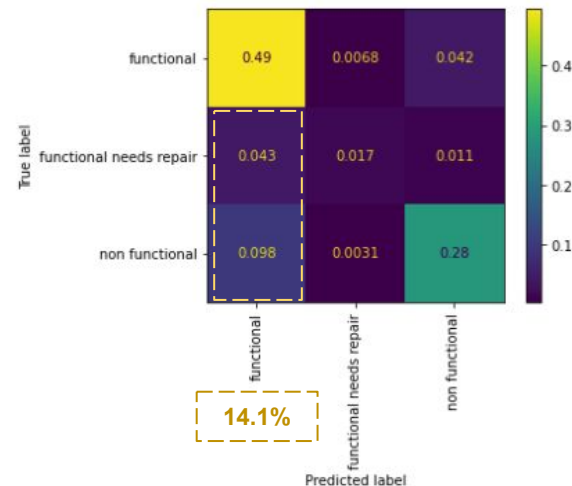
Non Functional Wells

Model Selection

Goal: Maximize Accuracy and Maximize Micro Precision

Logistic Regression	⇨	Accuracy: .73
Decision Tree Classifier	⇨	Accuracy: .75
Random Forest Classifier	⇨	Accuracy: .79
K-Nearest Neighbors	⇨	Accuracy: .76
Support Vector Machines	⇨	Accuracy: .77
XGBoost	⇨	Accuracy: .79

Random Forest Confusion Matrix

11.4%
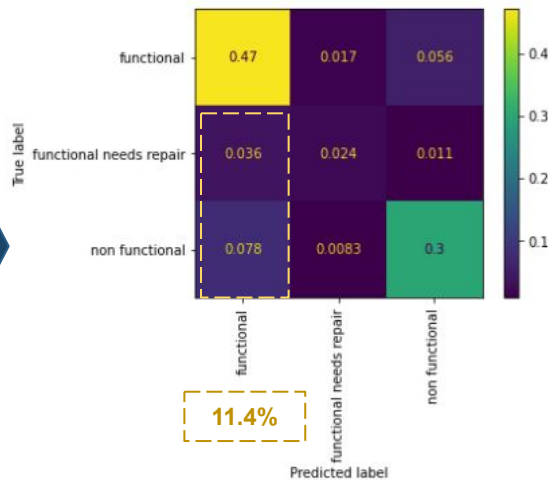
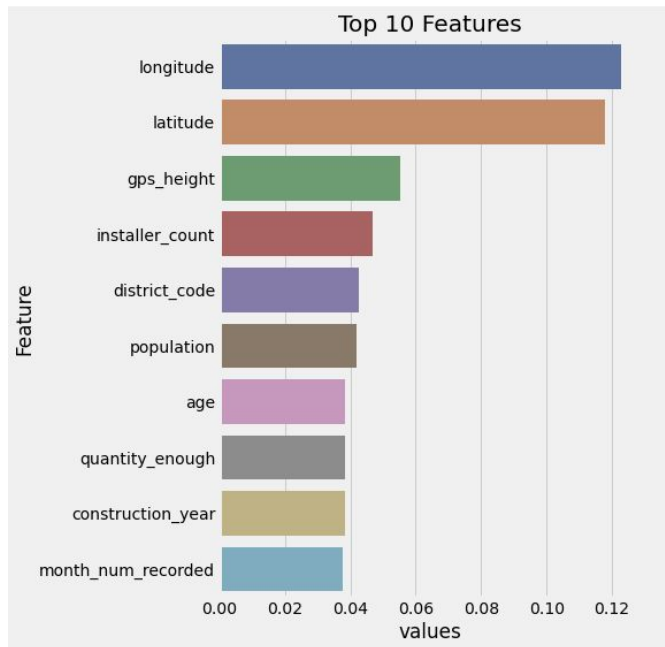**XGBoost Confusion Matrix**

14.1%

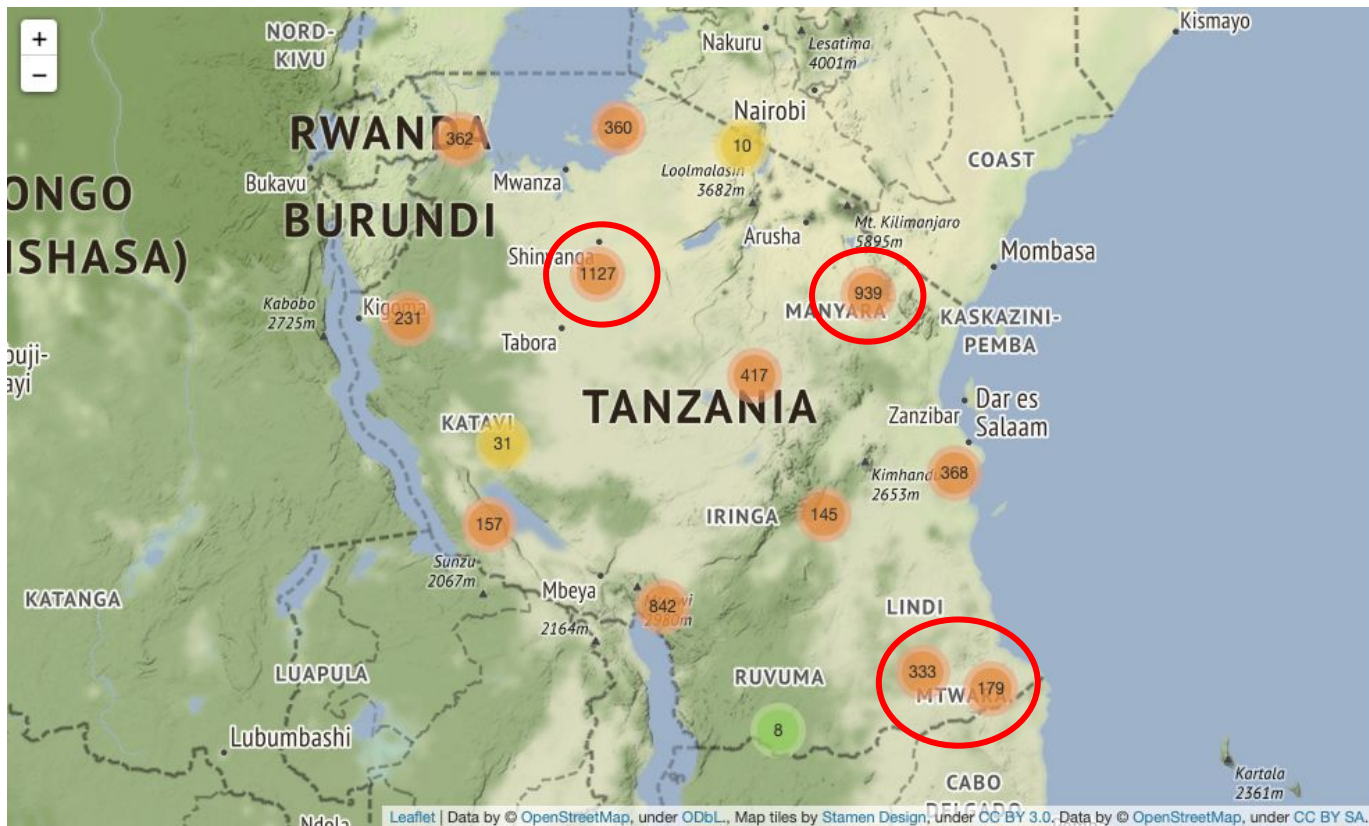# Final Model Evaluation

Accuracy: **.79**
Micro Precision: **.82**
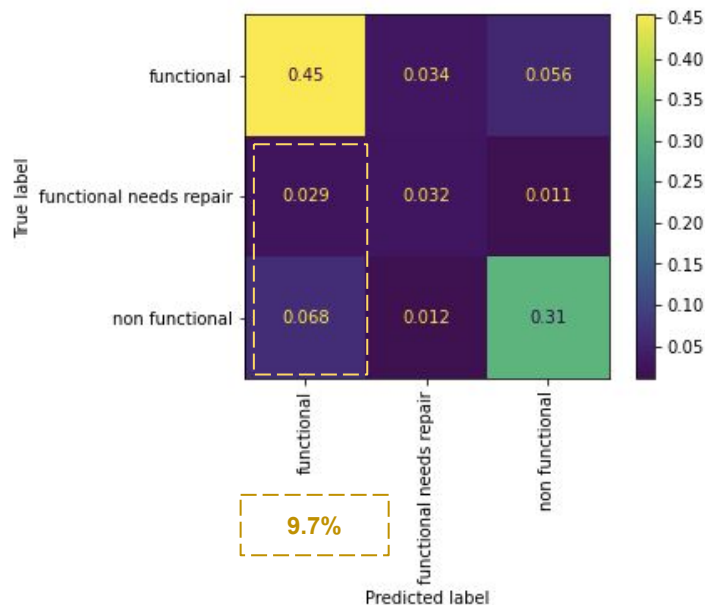
Feature Importance

Predicted Non-Functional Wells

Conclusion

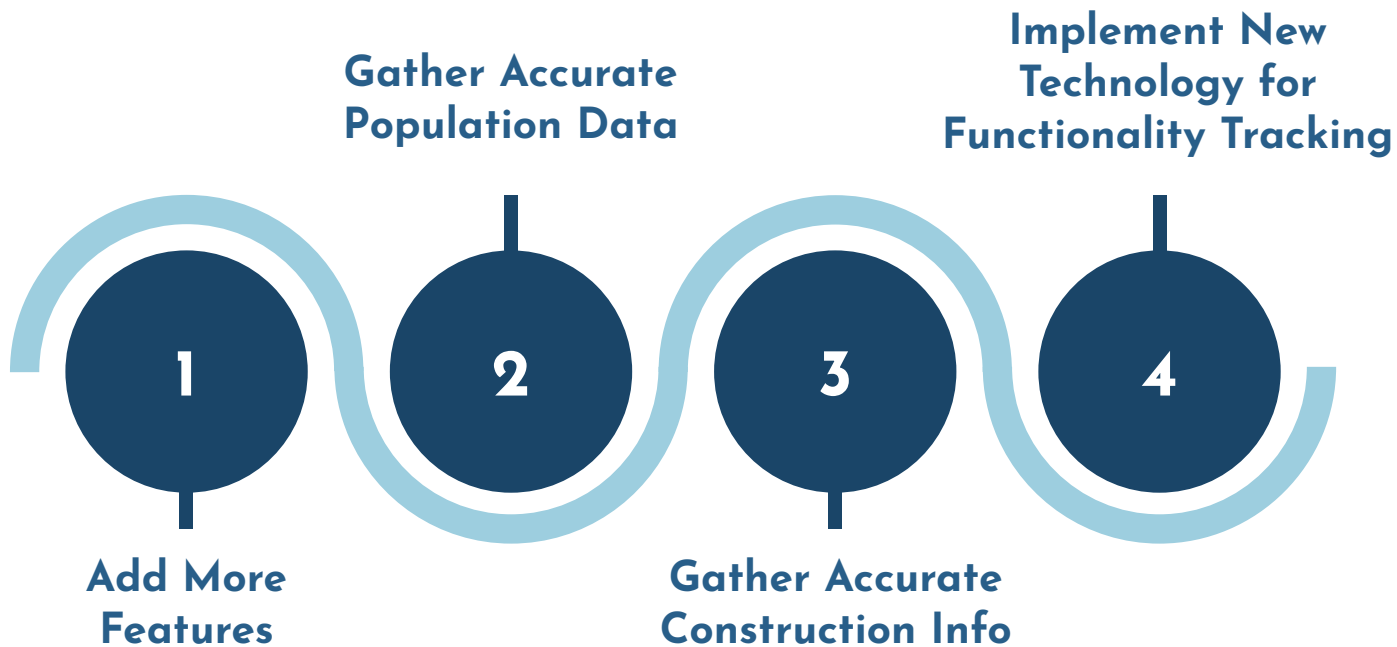Begin fixing non-functional wells in locations with highest predicted non-functional counts

Use top installers for well replacements and new installations

Focus on replacing older wells first

Next Steps

1 Add More Features

2 Gather Accurate Population Data

3 Gather Accurate Construction Info

4 Implement New Technology for Functionality Tracking

Thank You

Ryan Reilly
Email: ryan.m.reilly@gmail.com
Github: https://github.com/ryanreilly
Linkedin: https://www.linkedin.com/in/ryanreilly1/

Kevin McDonough
Email: kpmcdonough@gmail.com
Github: https://github.com/KPMcDonough49
Linkedin: https://www.linkedin.com/in/kevin-mcdonough-01466a178/