# PUMP IT UP: Data Mining the Water Table

Amanda Rowe

# Tanzania Water Access

- Fifth most populous country in Africa
- Large population scatter in rural areas
- Impact of water and food scarcity

## The Task

# Data Driven Model

The aim of the project is to help the Tanzanian government (and local/international support organizations) detect when a water well will be in need of repair.
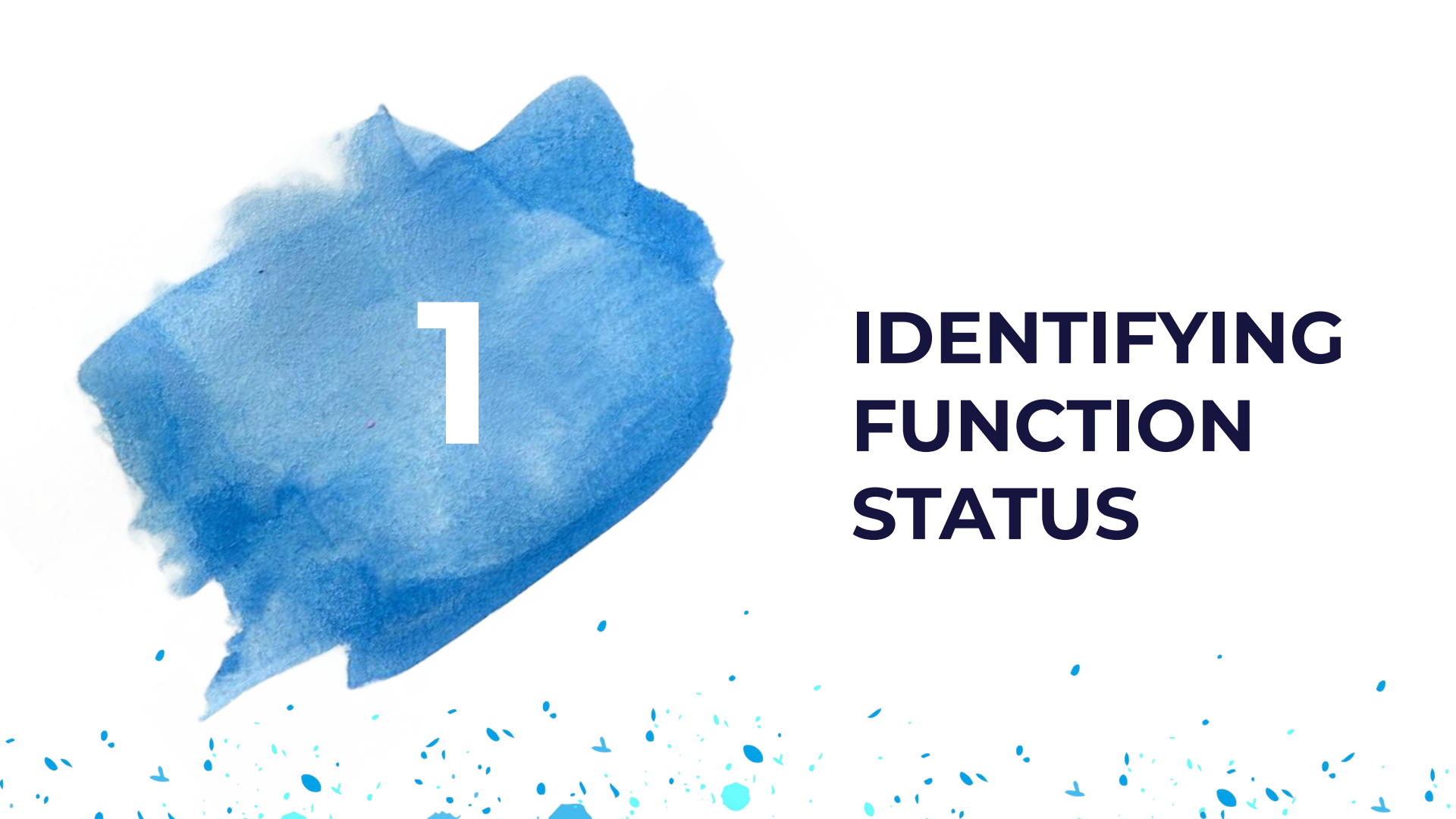
# The process

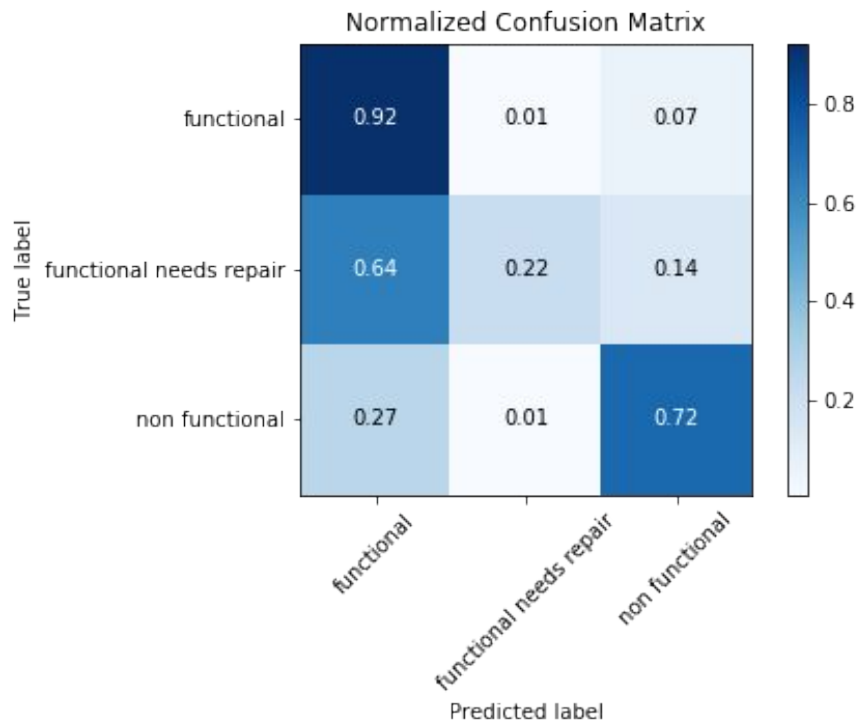Analyze the features

Limit features

Train, Test, and Tuning

# 1

# IDENTIFYING FUNCTION STATUS

# Ternary Model Performance

| | Accuracy |
|---|---|
| K Nearest Neighbor | .77 |
| Decision Tree/Bagged Tree | .70/ .71 |
| Random Forest | .79 |
| XG Boost | .79 |
| AdaBoost/ Gradient Boost | .72/ .75 |
| Logistic Regression | .55 |

# Finding Wells in Need of Repair



Normalized Confusion Matrix

# Identified Well Labels



Identified Wells in Training Data

Well Status: functional, non functional, functional needs repair
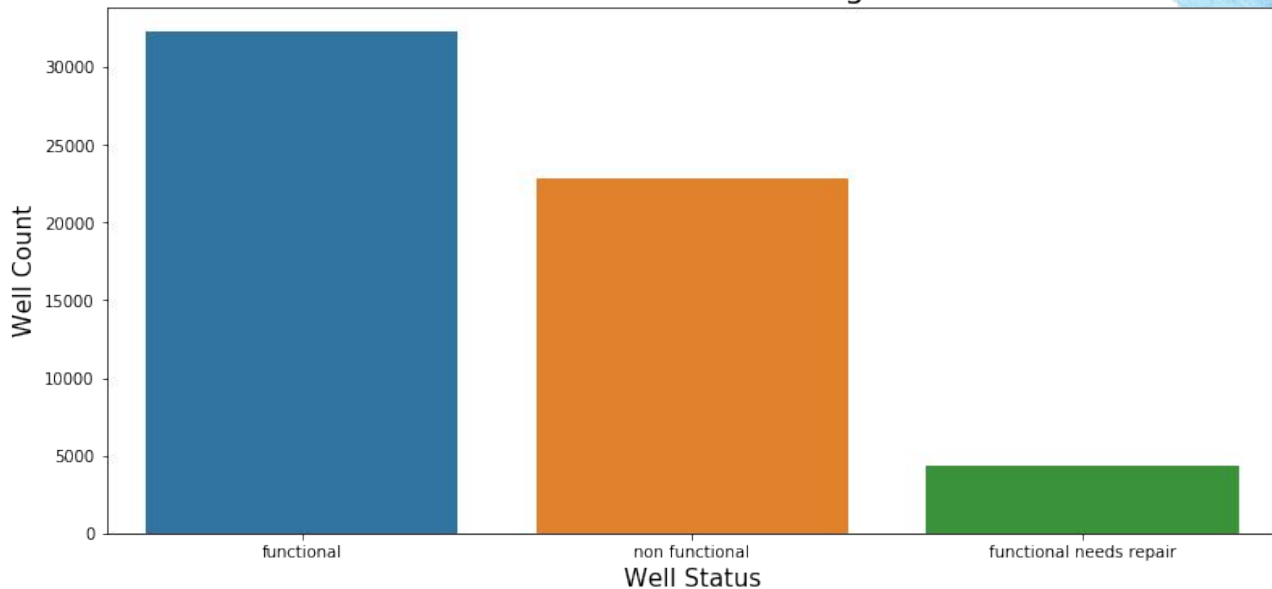
**2**

# NARROWING THE SCOPE

# Binary Model Performance

|  | Accuracy |
|---|---|
| **K Nearest Neighbor** | .79 |
| **Decision Tree/Bagged Tree** | .72/ .73 |
| **Random Forest** | .81 |
| **XG Boost** | .76 |
| **AdaBoost/ Gradient Boost** | .74/ .76 |
| **Logistic Regression** | .54 |

# Finding Wells in Need of Repair



Normalized Confusion Matrix
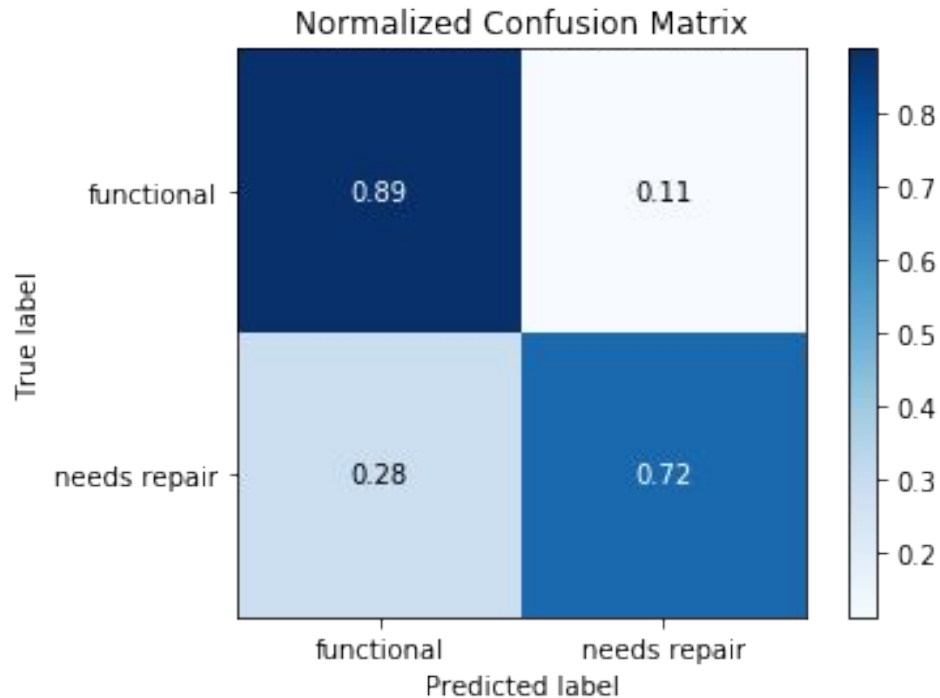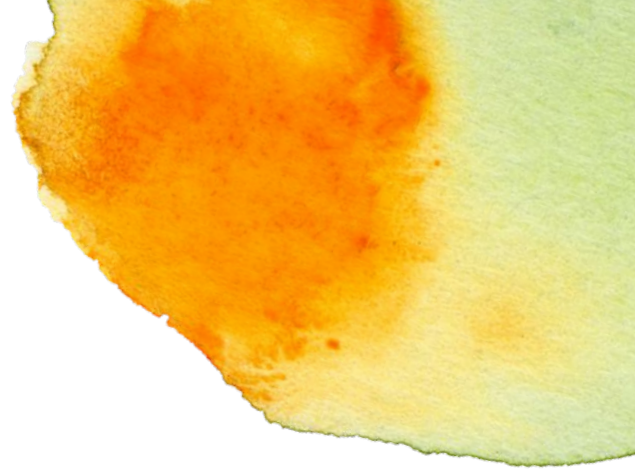
# Future Work

### Causes

Identify causes for the wells to need repair and add that to the model to be able to specify the support needed.

,

### Support

Provide information about operating wells to the communities affected by the non-functioning wells.

### Climate

With climate change, how to anticipate and prepare communities for the future.

# Thanks!

**Any questions?**

# Credits

Special thanks to all people who made and share these awesome resources for free:

- > Presentation template designed by Slidesmash

- > Photographs unsplash.com

- > Watercolors by graphicburguer.com

- > Data provided by drivendata.org

- > Additional facts about Tanzania provided by fundforpeace.org