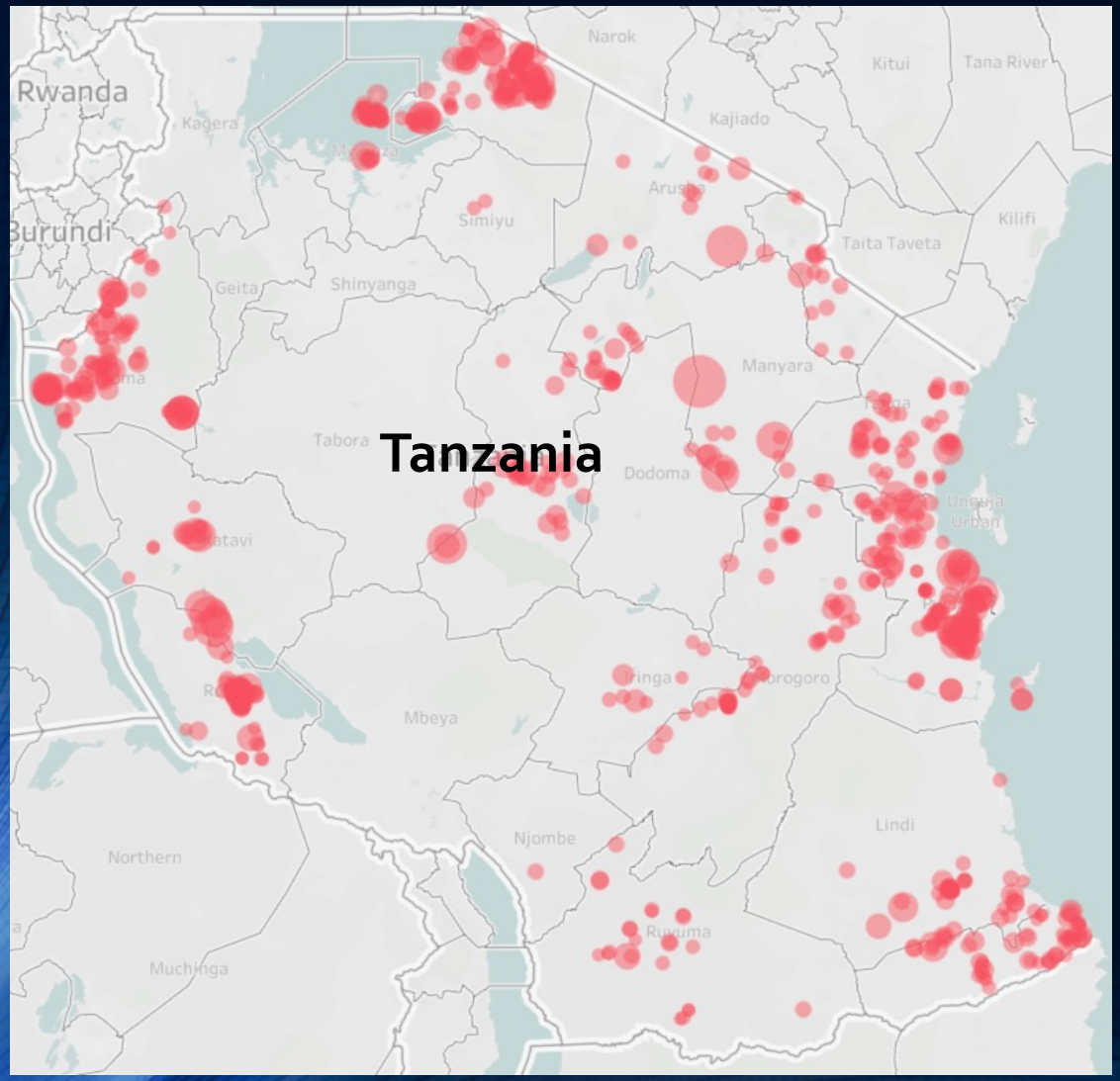
# Predicting Water Well Failures Using Machine Learning

ANALYSIS BY BRIAN BENTSON

# Predicting Well Failure Can Save Lives

Non-Functional wells that support over 1,000 citizens
(sized by population)



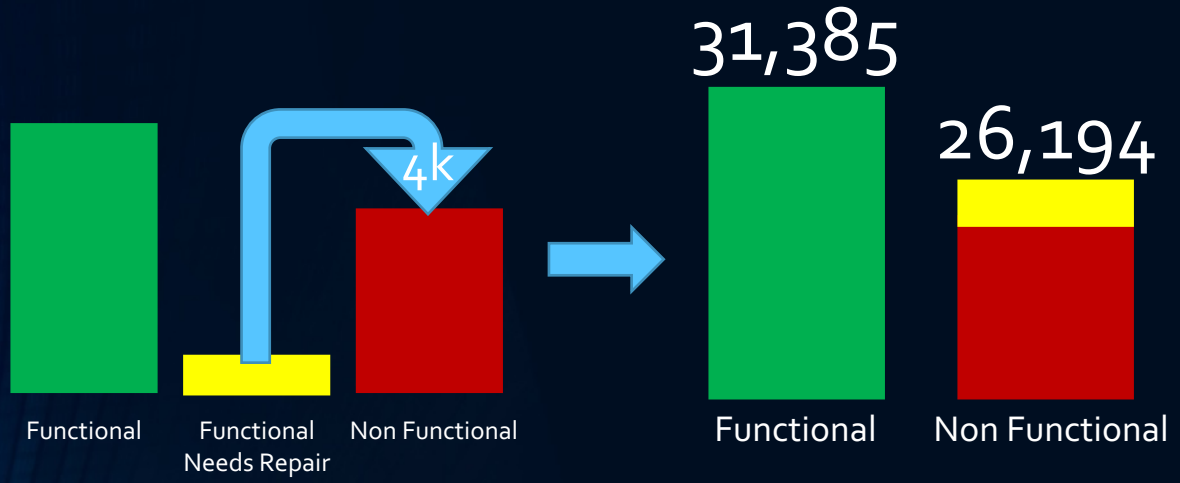

- Tanzania has almost 60,000 water wells in dataset with 45% not functional, leaving 4.6M people without a reliable water source
- Humans can only live up to 3 days without water
- Ability to predict water well failures and respond quickly can be the difference between life and death

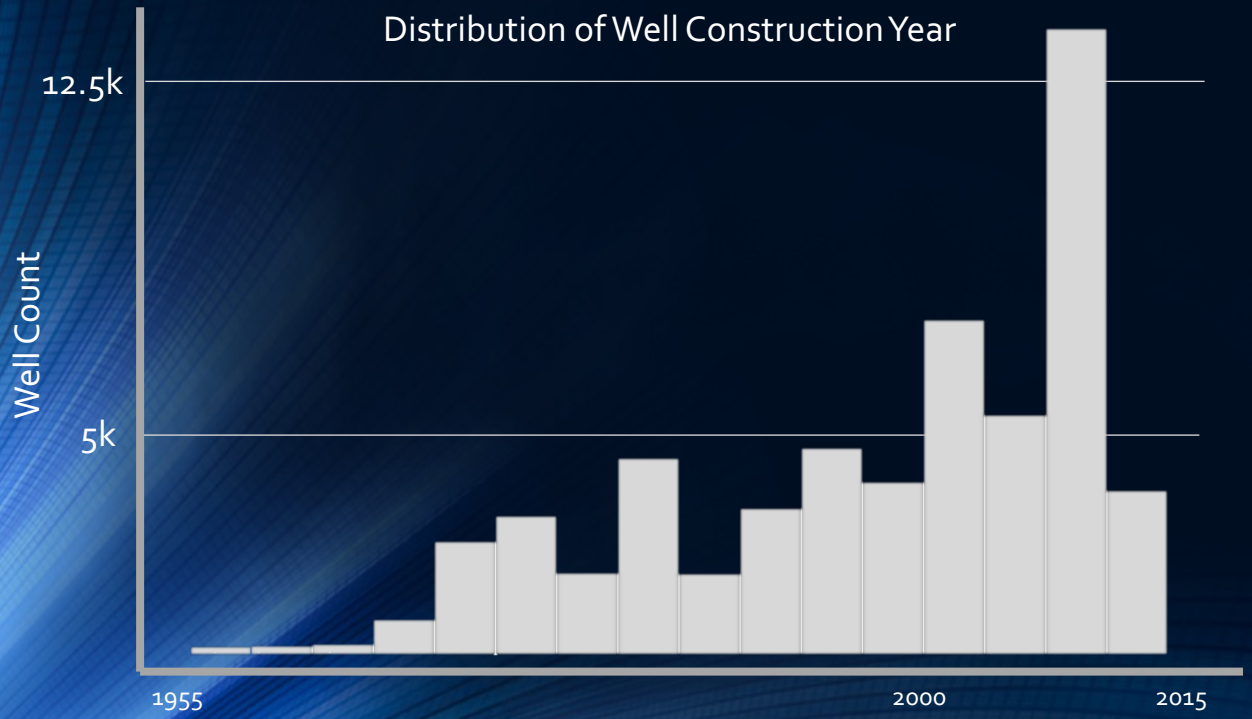
# Data Quality a Potential Issue

# 19K

- Almost 19,000 water wells with zeros for population, head, well elevation and construction year
- Outliers and improbable values  in head and population that can skew results
- Zeros in construction year were converted to years based on distribution of non-zero values

# Analysis Overview

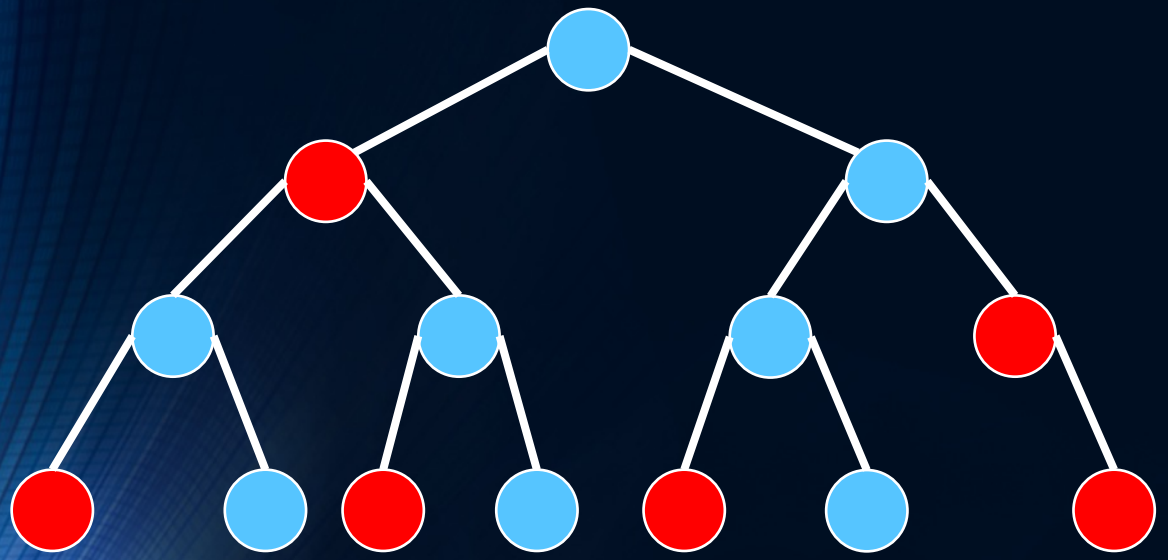



- Equated wells functioning at a reduced capacity as not functioning
- Created well_age feature
- Classification modeling to use well features to predict non-functional water wells can save lives by increasing reliability and maintenance response time

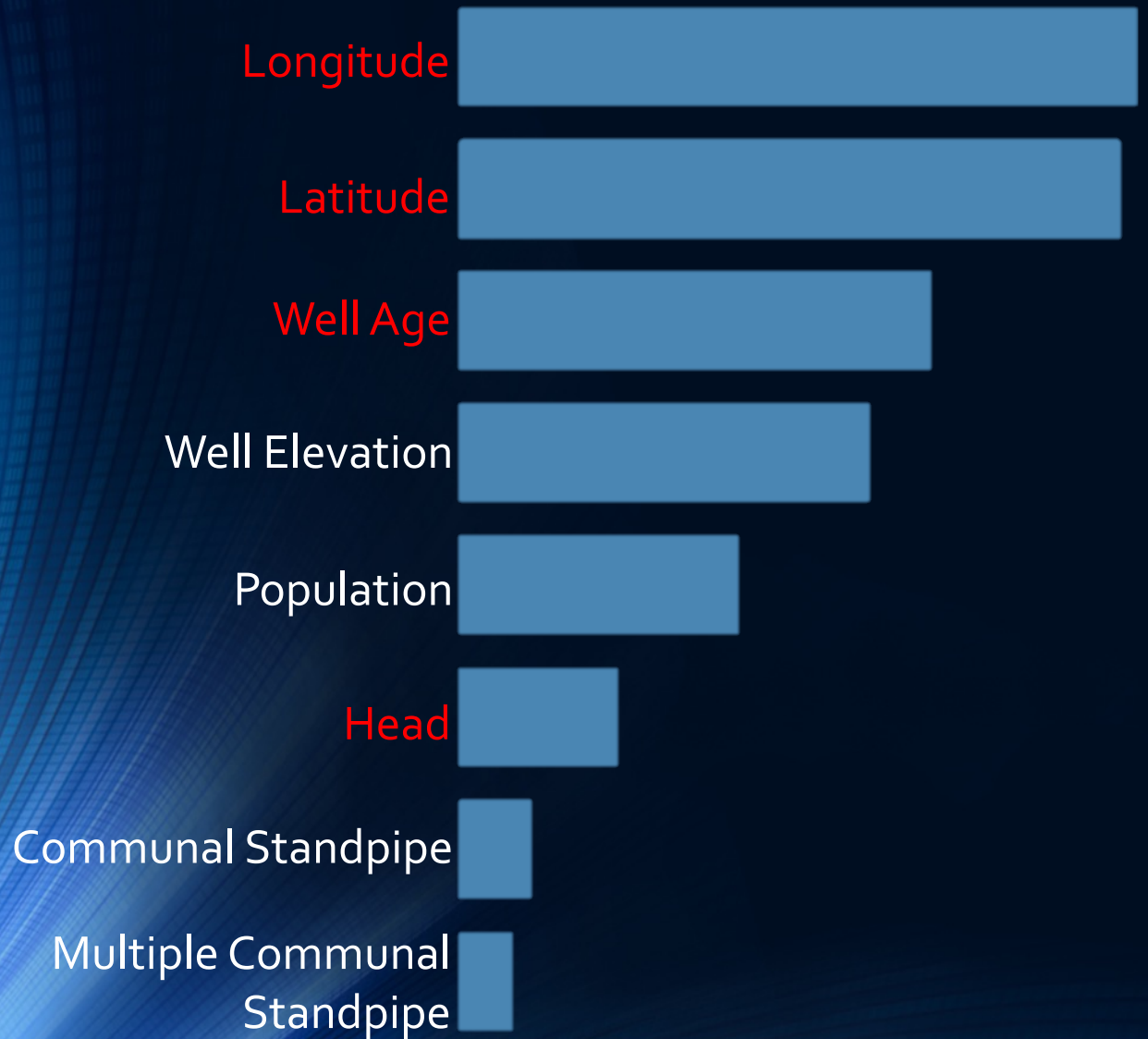
# Best Classification Modeling of Well Status

Random Forrest



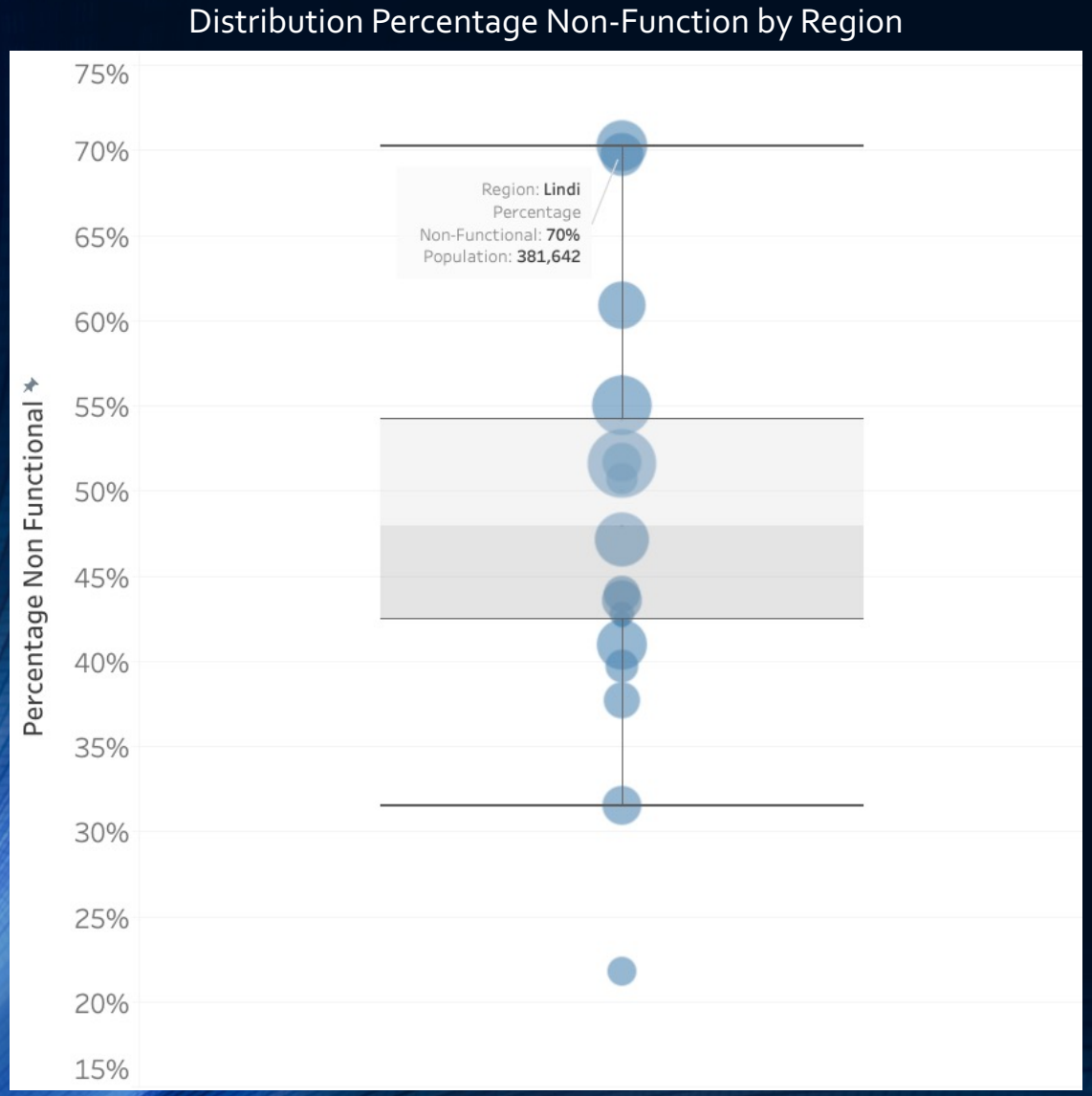
- Prioritized Recall which prioritizes finding all well failures while maybe responding to wells which haven't failed
- Increased maintenance spend and resources required
- Focused on models with high interpretability to understand what drives well failures
- Best model was a Random Forrest with a Recall for detecting failures of 77%

# Most Important Features from Model



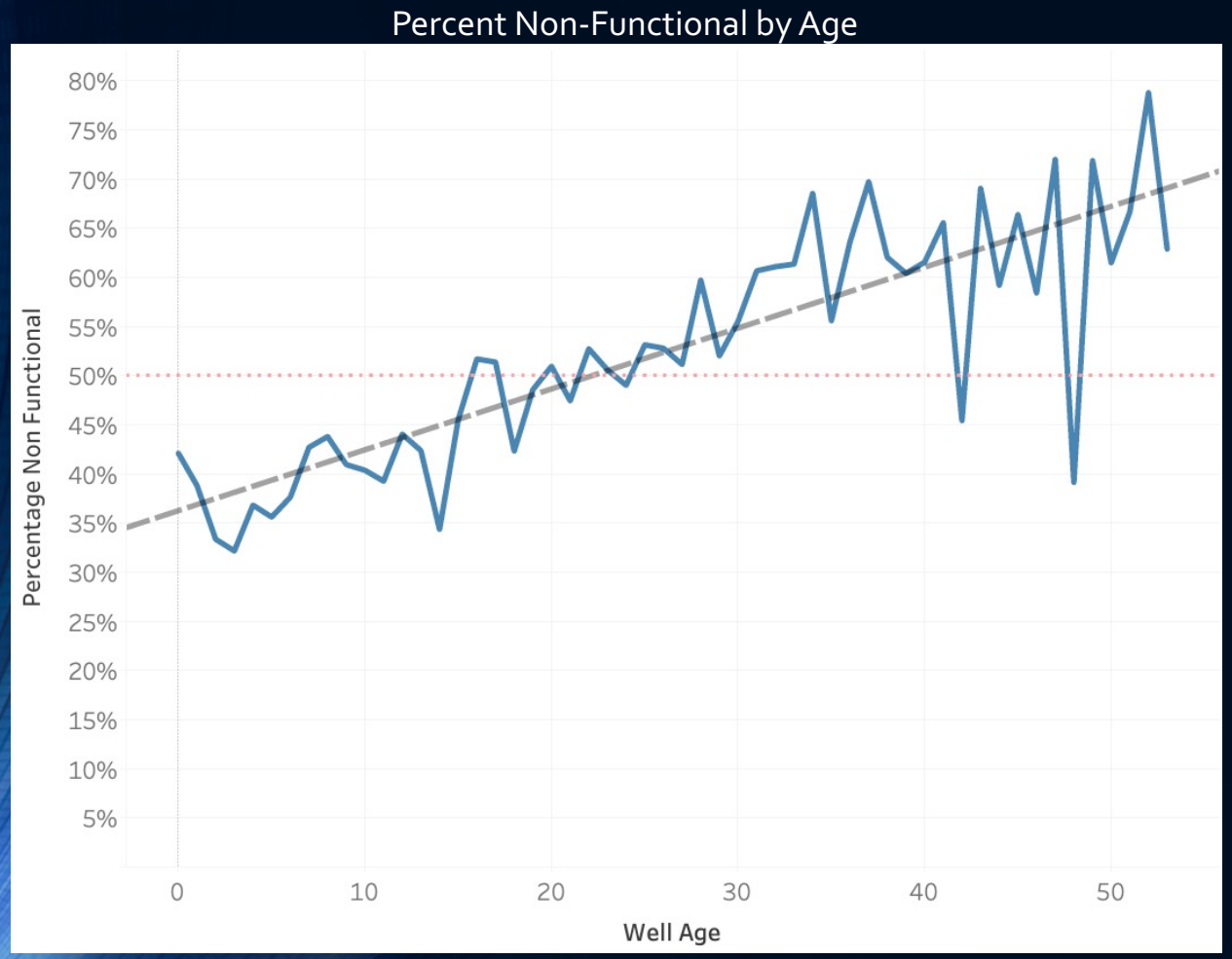

- Location (longitude and latitude) is the most important feature for predicting well function
- Well age directly affects reliability
- Well head directly affects water quantity
- Well Elevation and Population are important features for prediction, although analysis did not highlight specific relationships

# Location Affects Reliability



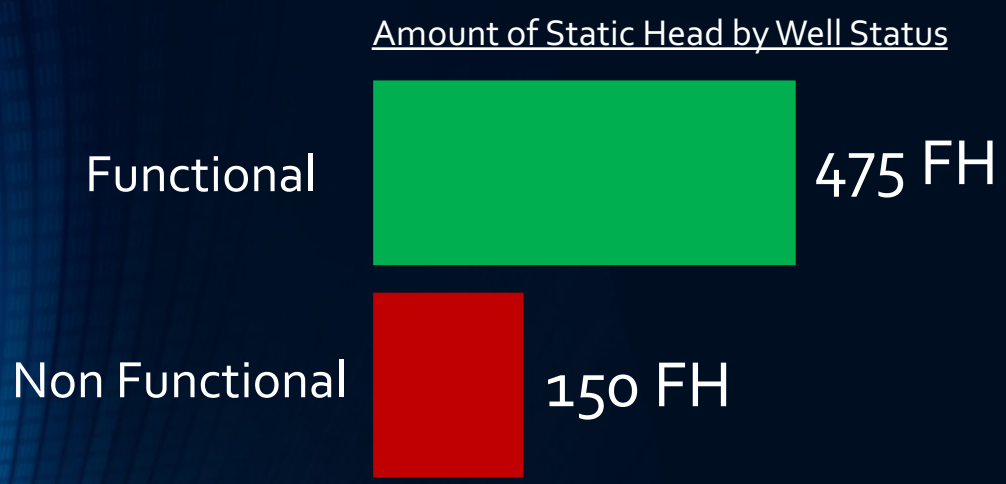

- High variability in well function across regions
- Many Regions with high failure percentage and high population
- Recommendation: Focus on regions which have historically high failure percentage

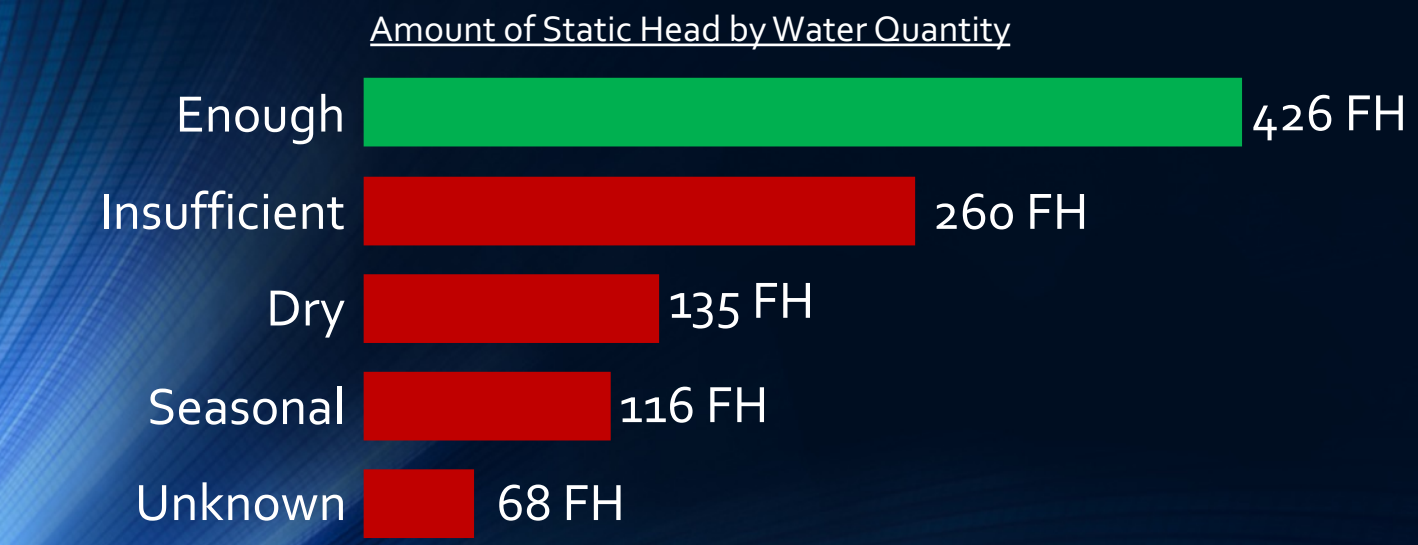
# Well Age Negatively Affects Reliability




- As well ages increase, the percentage of non-functional wells increases
- If a well is over age 24-25, it is more likely that well will be non-functional rather than functional
- 2.5M people are supported by older wells (>24 years old)
- Recommendation: Focus maintenance on older wells to maintain supply of water

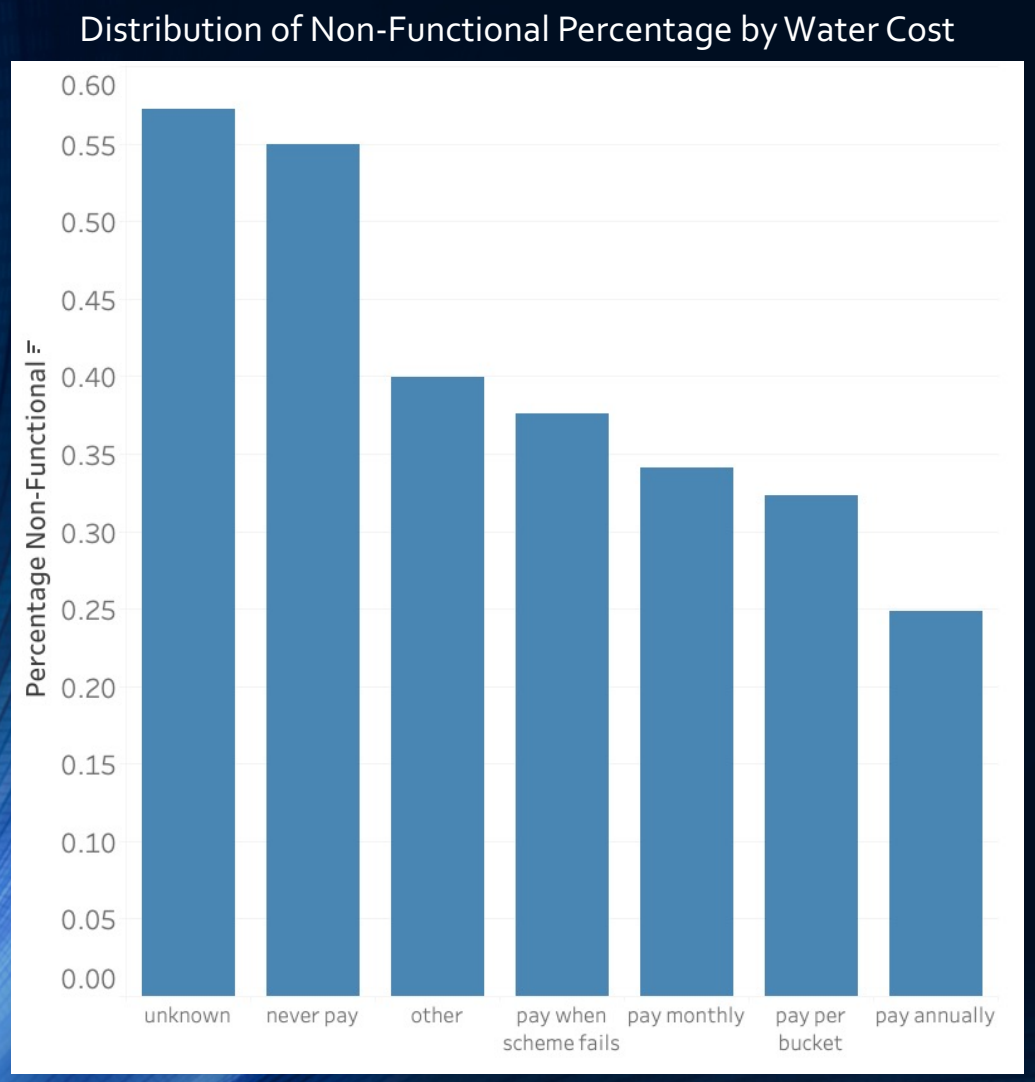
# Lower Static Head Indicates Failure




- As well ages increase, the amount of static head on the well will decrease, lowering water quantity
- This can be artificially improved by technology such as a pump
- Recommendation: Keep a close eye on static head as it directly correlates with water quantity

# You Get What You Pay For




- While not specifically important for the random forest model, there is a clear trend between showing that if you pay for the water, the wells reliability is higher
- Recommendation: Focus on supporting the populations which cannot afford to pay for water

# Recommendations & Next Steps

## Recommendations

- Develop a relationship with locals in order to establish a frequent communication protocol to ensure if water quantity drops, maintenance management companies get notified as soon as possible.
- Develop support model for population areas that do not pay for water since analysis shows paying for water brings better reliability
- Improve data governance to ensure better data quality and better predictions

## Next Steps

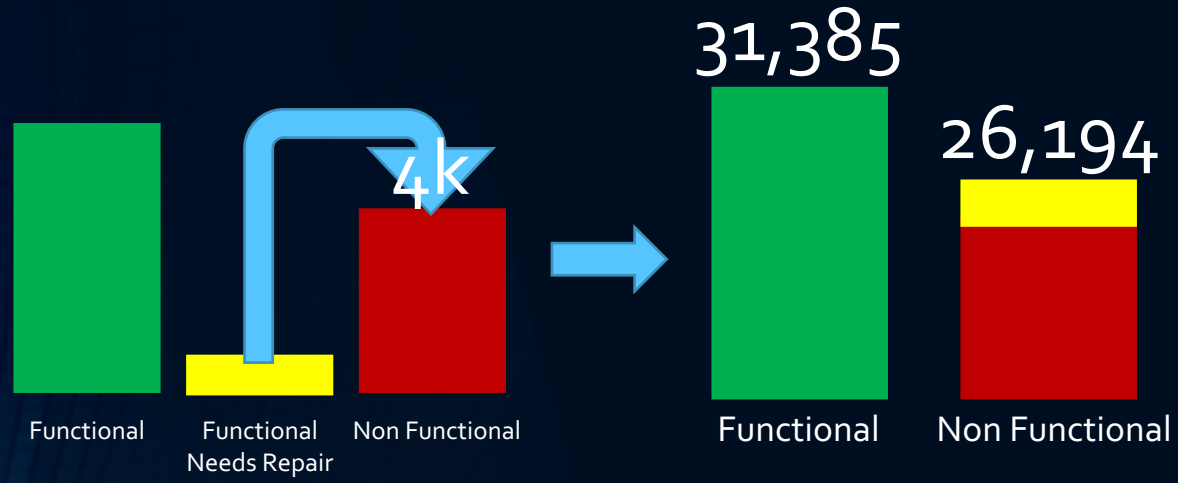
- Use more sophisticated algorithms that may perform better at finding well failures but does not tell you why they are failing
- Develop Impact Rating Dashboard

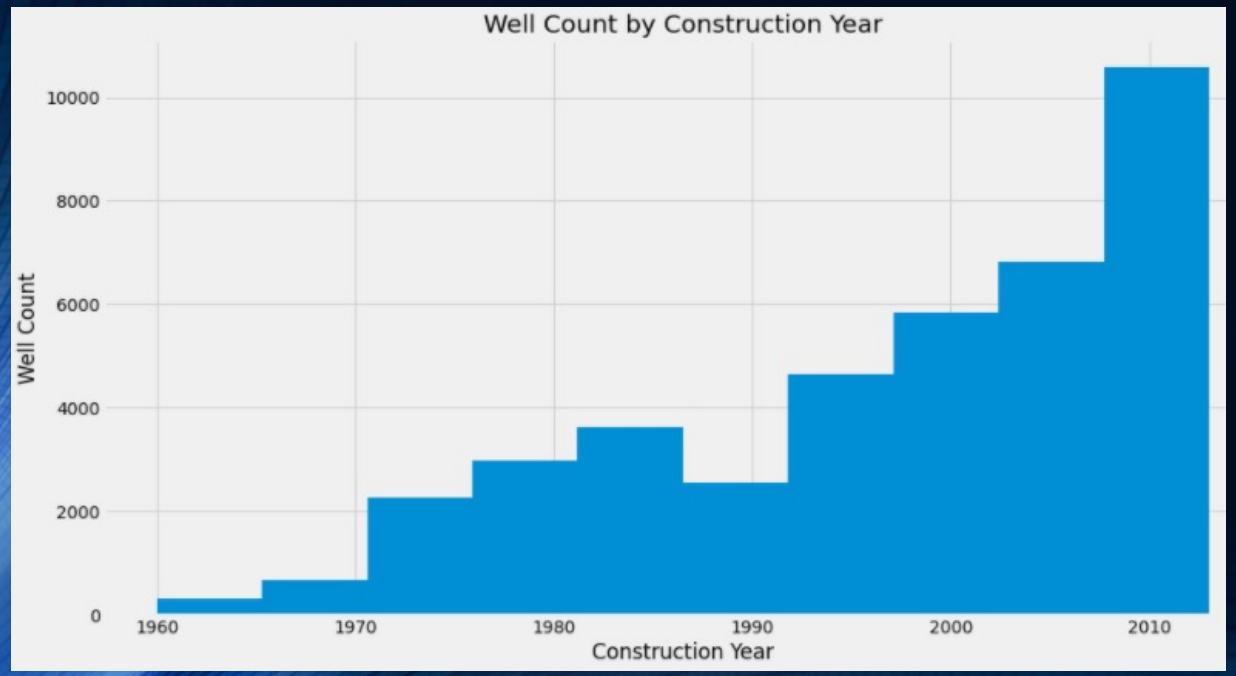
# Thank you

Questions?

# BACKUP

# Analysis Overview




- Focused on wells not functioning or functioning at a reduced capacity (functional needs repair)
- Created well_age feature
- 20,000 wells with unknown construction_year. Filled values keeping identical distribution
- Classification modeling to use well features to predict non-functional water wells can save lives by increasing reliability and maintenance response time