Water in Africa: Technical and Equipment Researchers, LTD. (WATER)

Water Point Repair in Tanzania

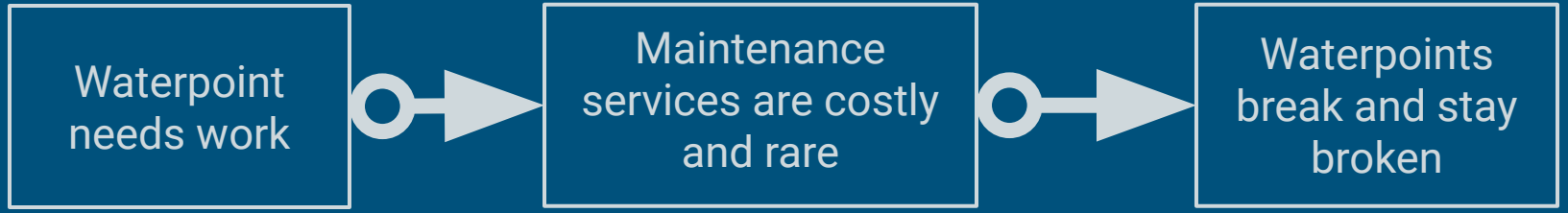
Goal:

Reduce overhead costs by anticipating water pump maintenance status

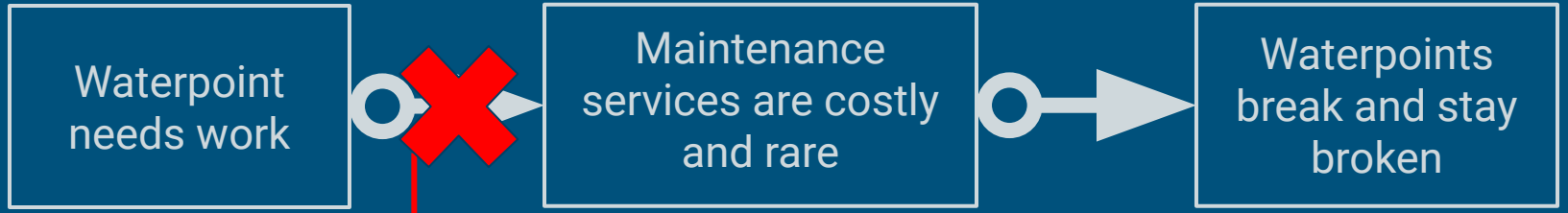
Executive Summary

- Current Landscape
  - The Problem
  - Data Understanding
  - Data Limitations
- Predictive Analysis
  - Our Model
  - Feature Importance
  - Model Performance
- Conclusion
  - Improvement Areas
  - Growth Opportunities
  - Takeaways

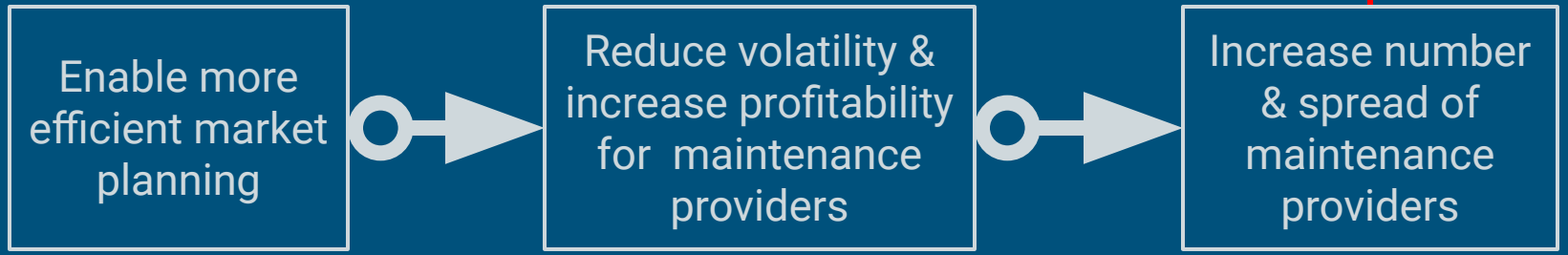
The Problem


Waterpoint needs work
Maintenance services are costly and rare
Waterpoints break and stay broken

The Problem



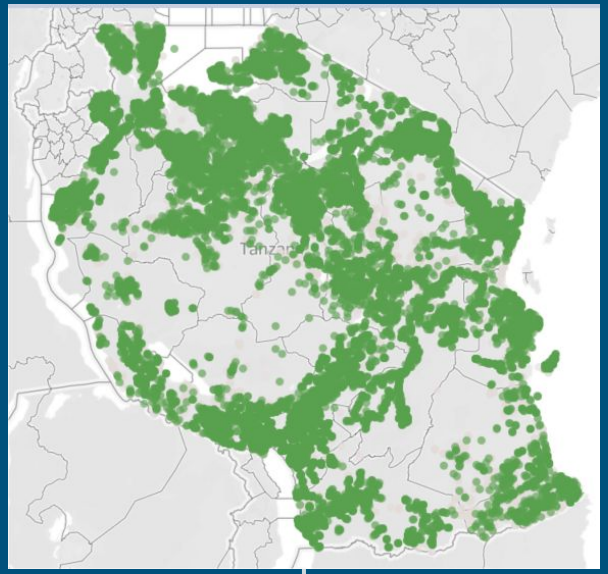
The Solution



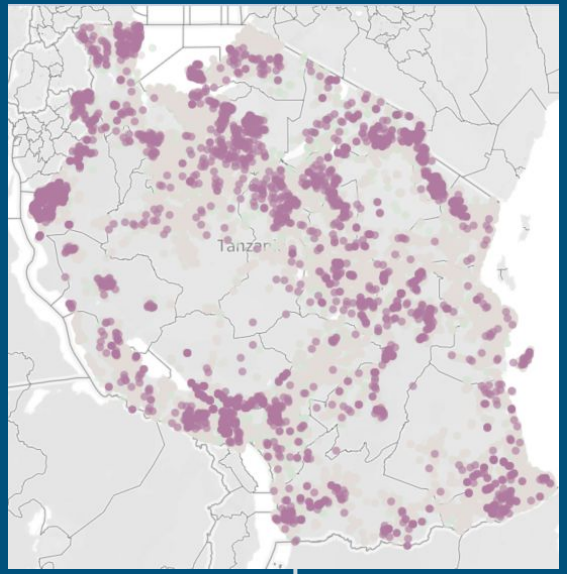
Current Landscape

Data Understanding:

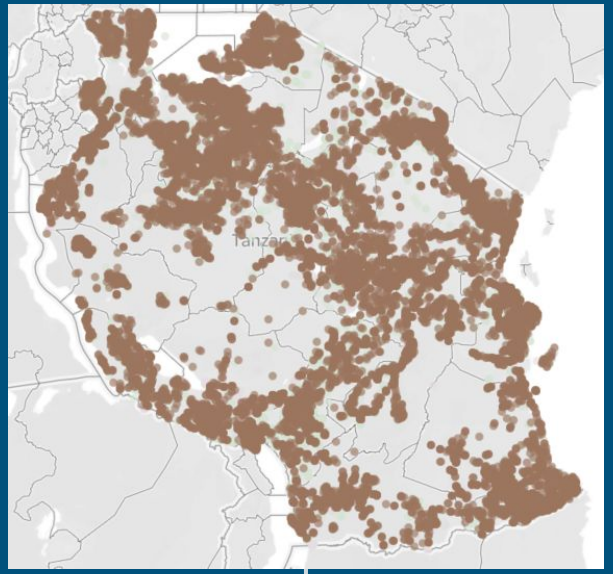
Geographical



Functional



Needs Repair

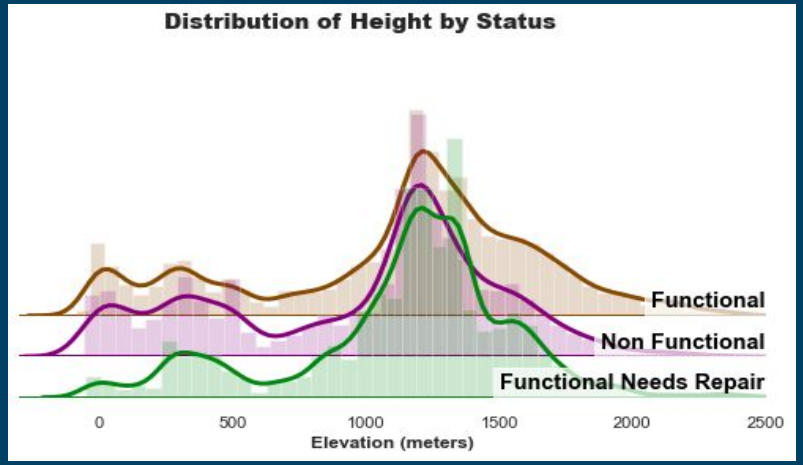


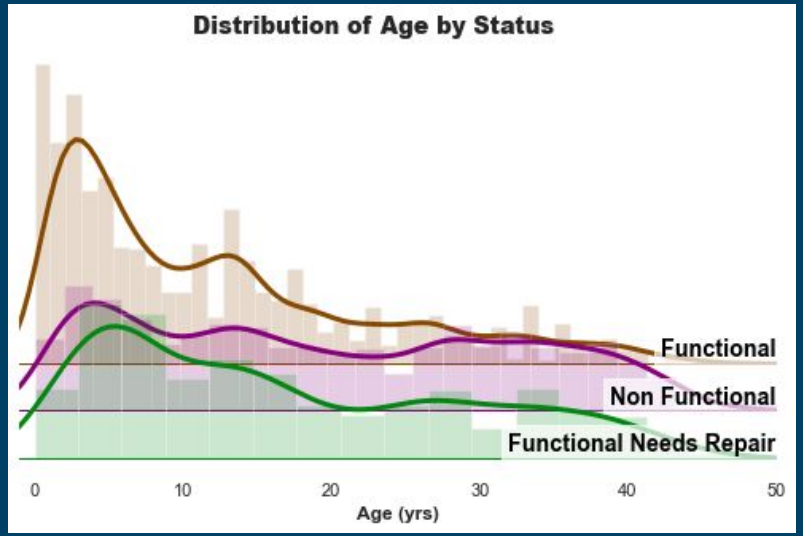
Non-Functional

Data Understanding

Obtained from Taarifa and Tanzanian Water Ministry

- 59,400 entries
- 38 properties per entry
- Properties cover:
  - Geography
  - Geology
  - Community Engagement
  - Operational habits
  - Age
  - Resource availability
  - Waterpoint specifics



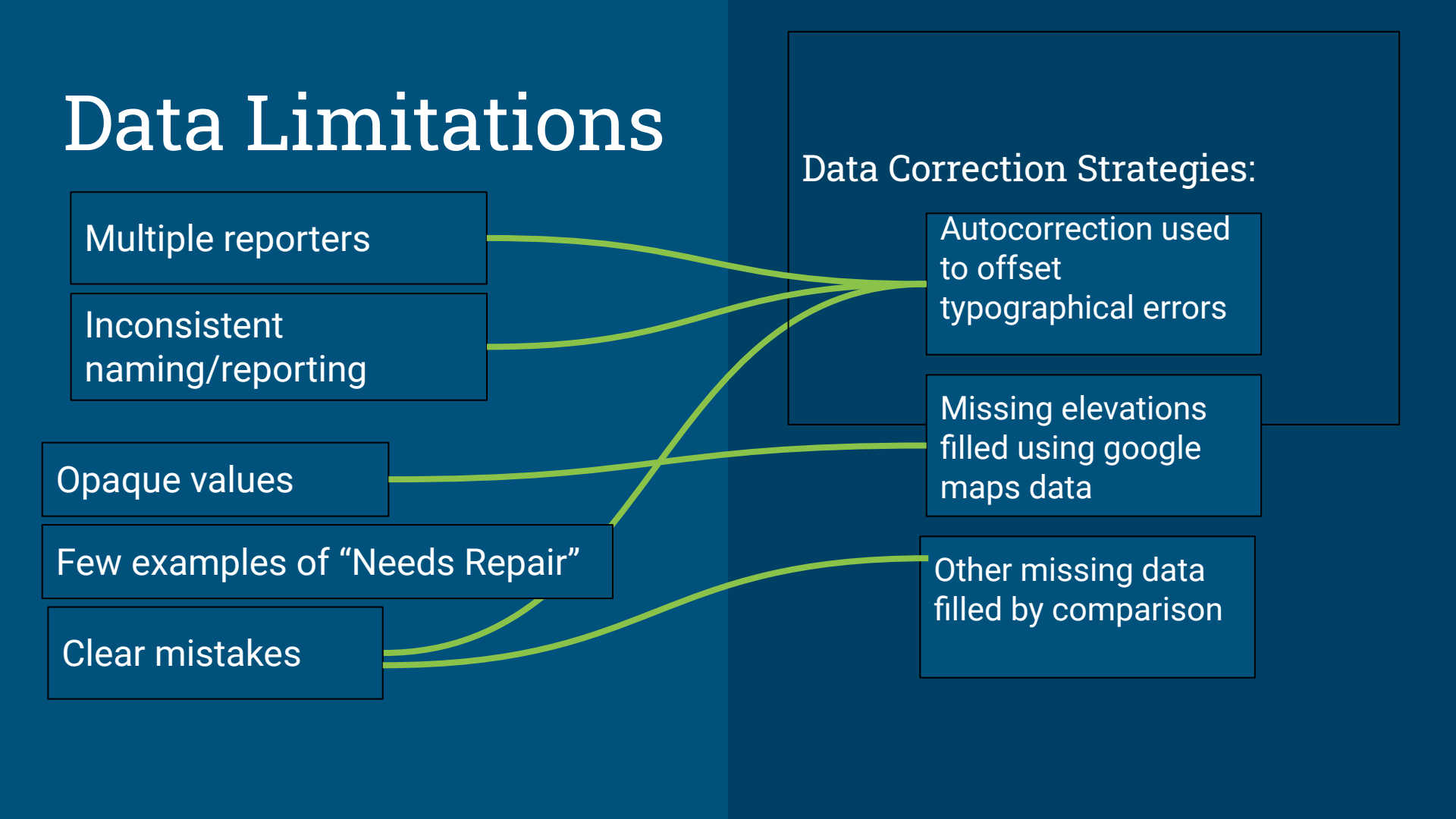


Data Limitations

- Multiple reporters
- Inconsistent naming/reporting
- Opaque values
- Few examples of “Needs Repair”
- Clear mistakes

Data Correction Strategies:

- Autocorrection used to offset typographical errors
- Missing elevations filled using google maps data
- Other missing data filled by comparison



Predictive Analysis

Our Model

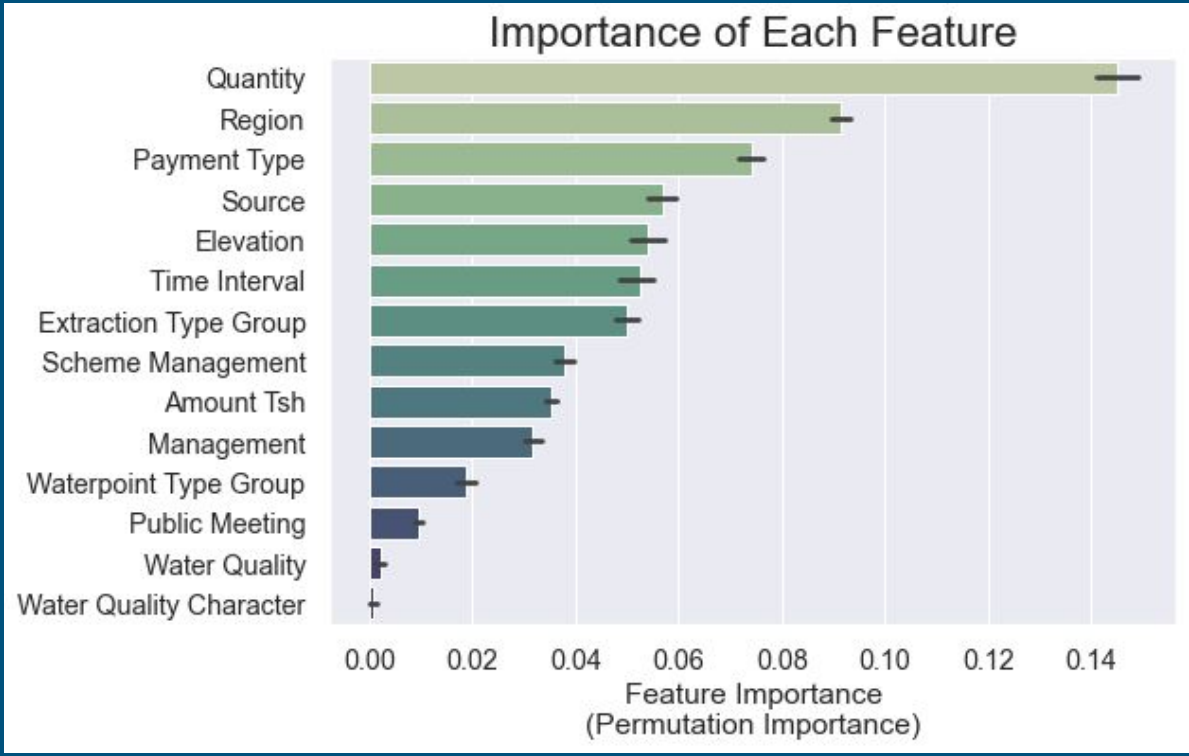
- Implements “Random Forest Classification”
- Optimized to reduce over-extending expectations

Model looks at:

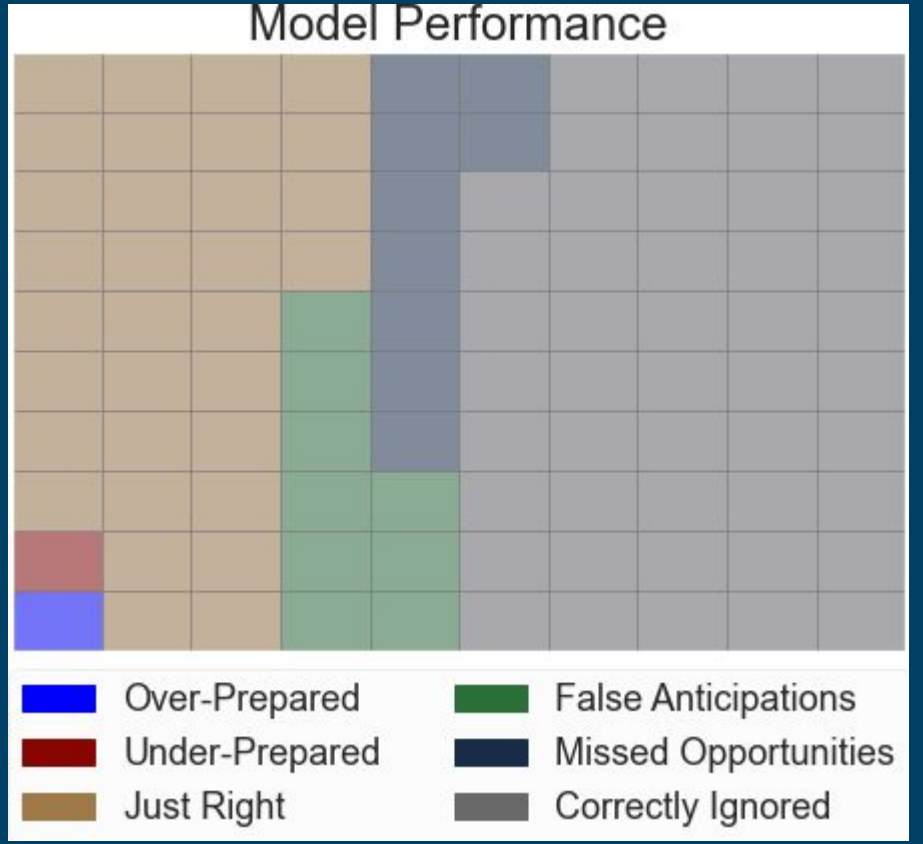
- Region & Elevation
- Water Quantity, Quality, and Source
- Payment Strategy, Community Engagement, and Management
- Well & Pump types & Capacity
- Age

Importance of Features

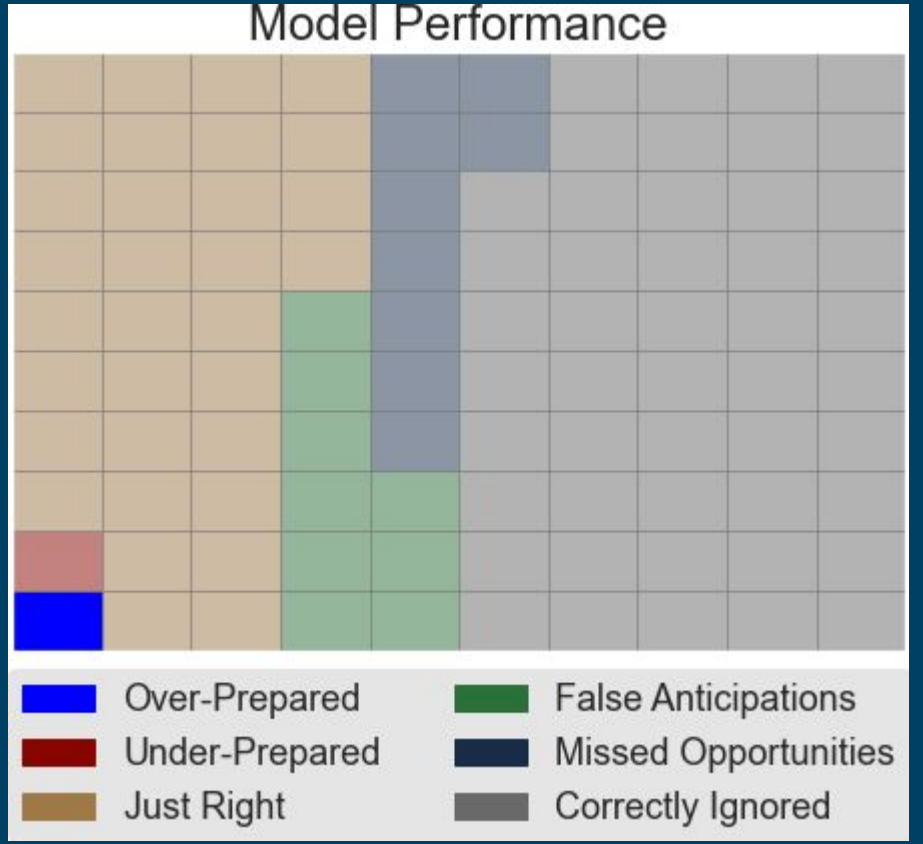
- Empirical approximation of feature importances
- Most Important:
  - Availability
  - Location
  - Payment Strategy
- Quality features may not be as unimportant as they appear



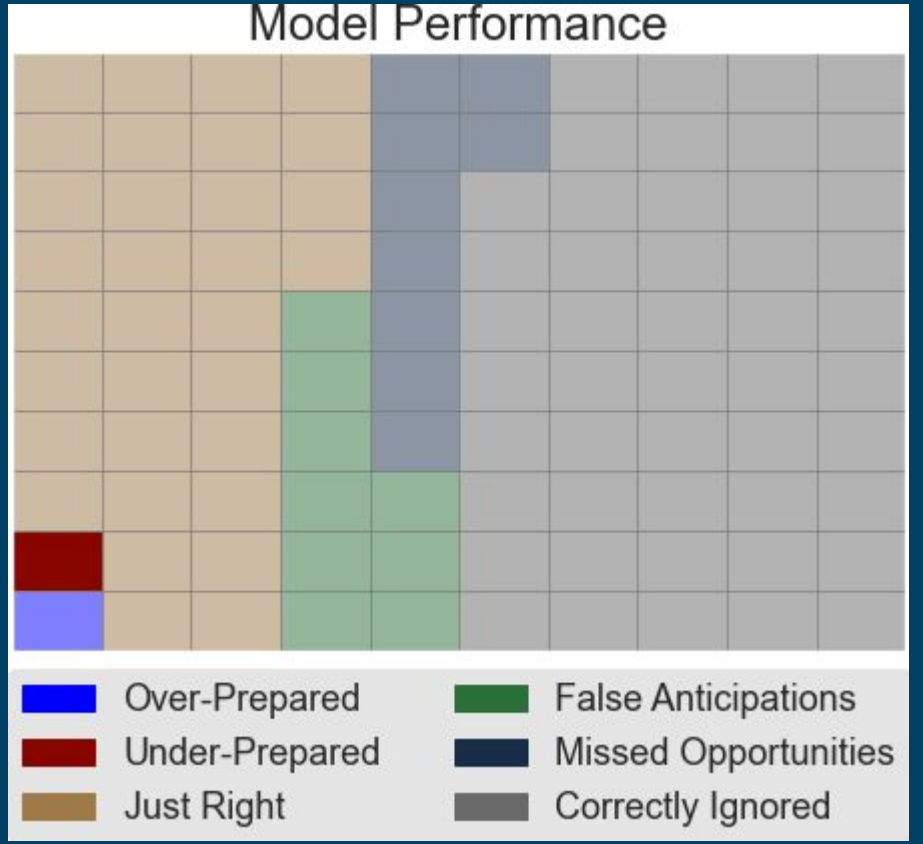
Model Performance



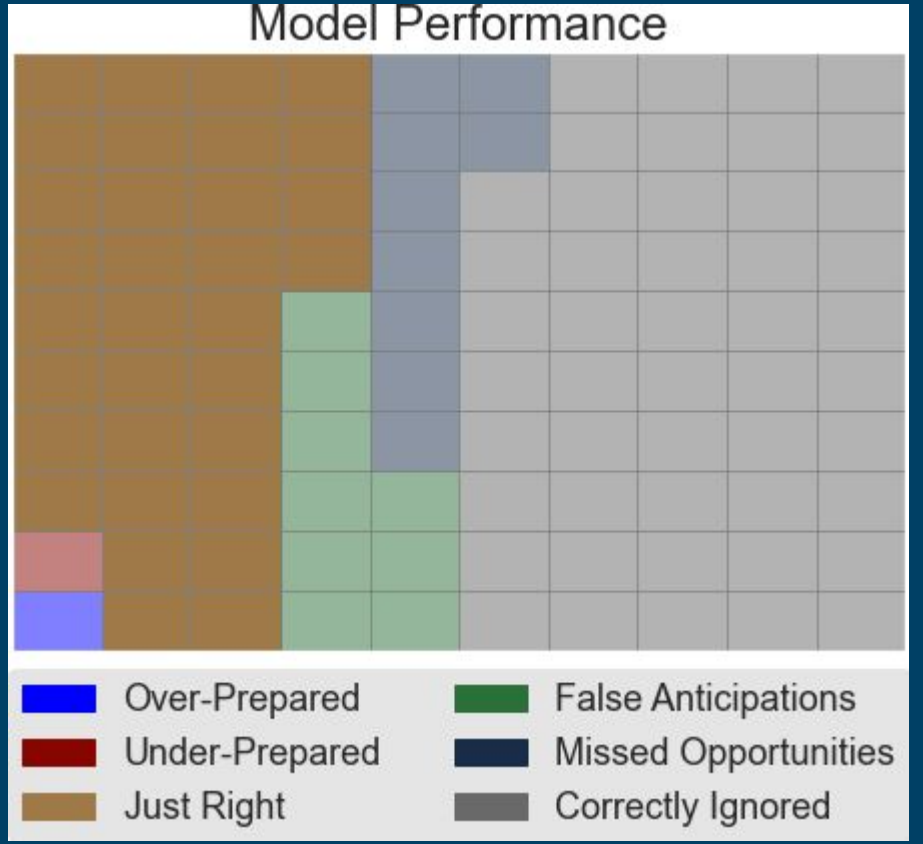
Model Performance



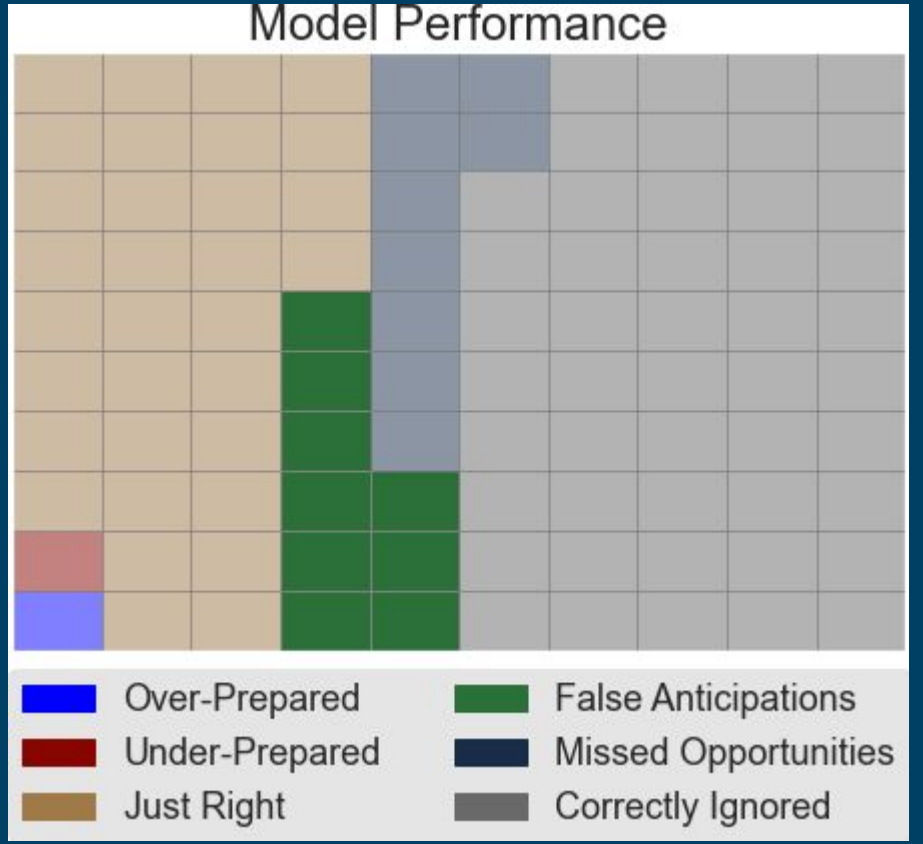
Model Performance



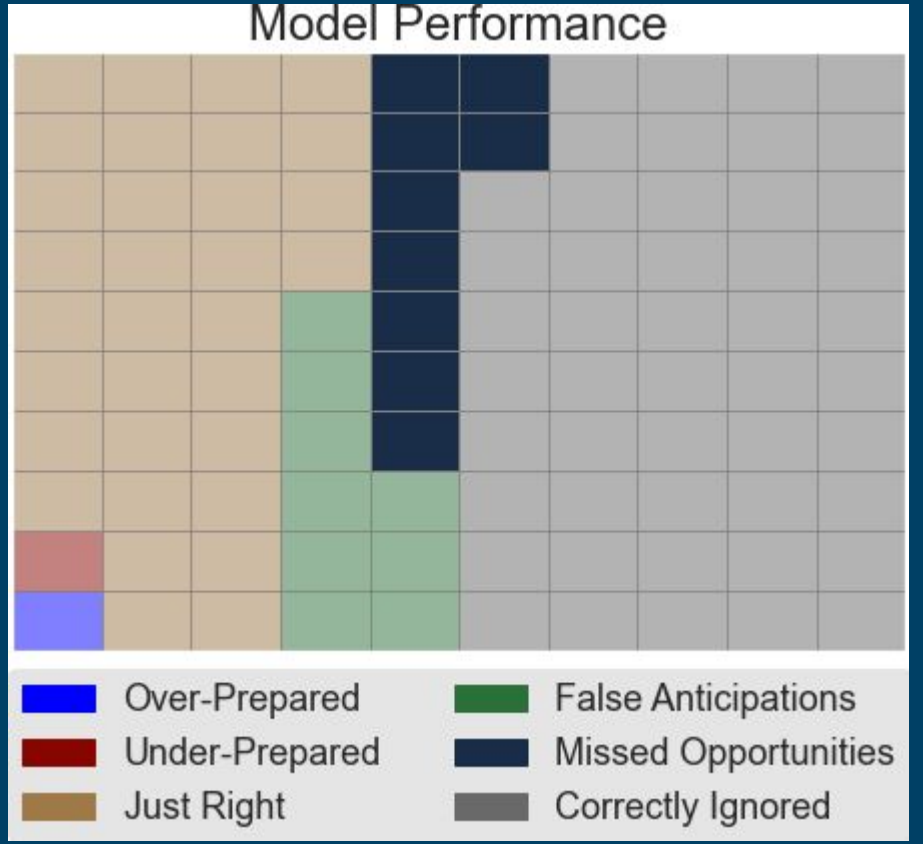
Model Performance



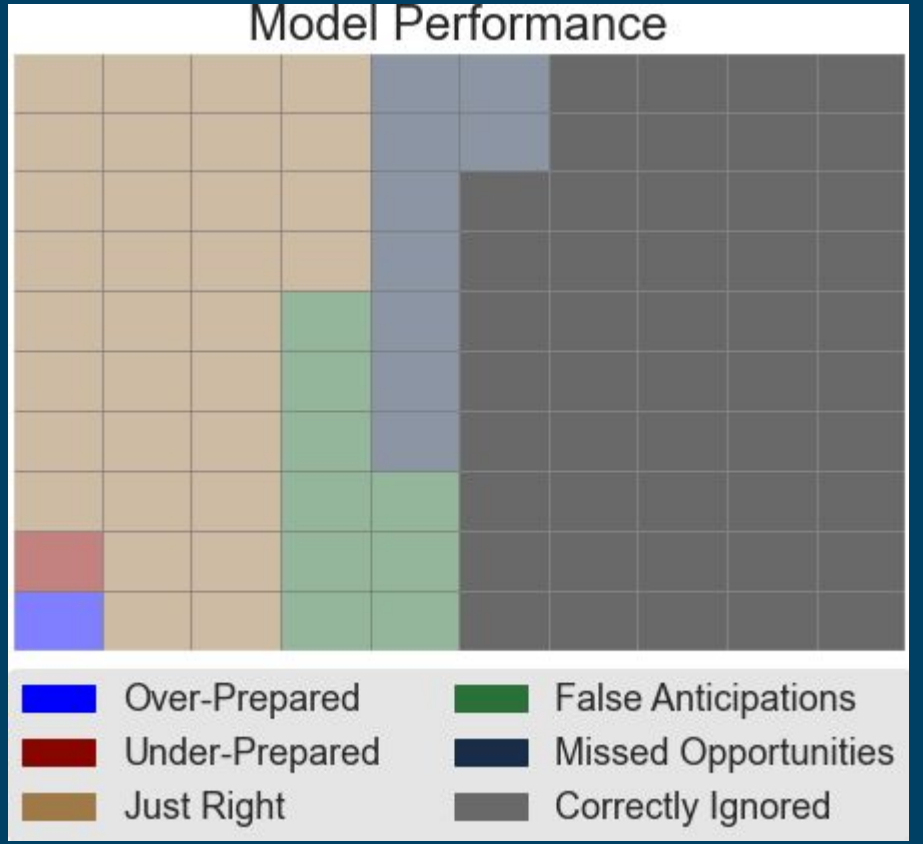
Model Performance



Model Performance



Model Performance



Conclusion & Takeaways

Areas to Improve

- Better **computational** resources
- Underlying **data quality** correlations
- Obtain **more data** for integrity control
- More sophisticated **data correction**
- Further **feature optimization**
- Informed **feature engineering**

Growth Opportunities

- Cautiously investigate applicability to other areas
- Develop lifetime estimations
- Invert model to identify demand for particulars
- Combine with economic models

Takeaways

Goal: Reduce overhead costs by anticipating water pump maintenance status

Improve maintenance planning & budgeting by:

- Using readily available categories of data
- Applying data to our model
- Model gives desired performance:
  - 91% for pumps needing any maintenance
  - 78% for identifying level of maintenance

“ASANTE!”

James Shaw

Alexander Newton


Questions?