



Tanzanian Water Wells

Michael Mahoney

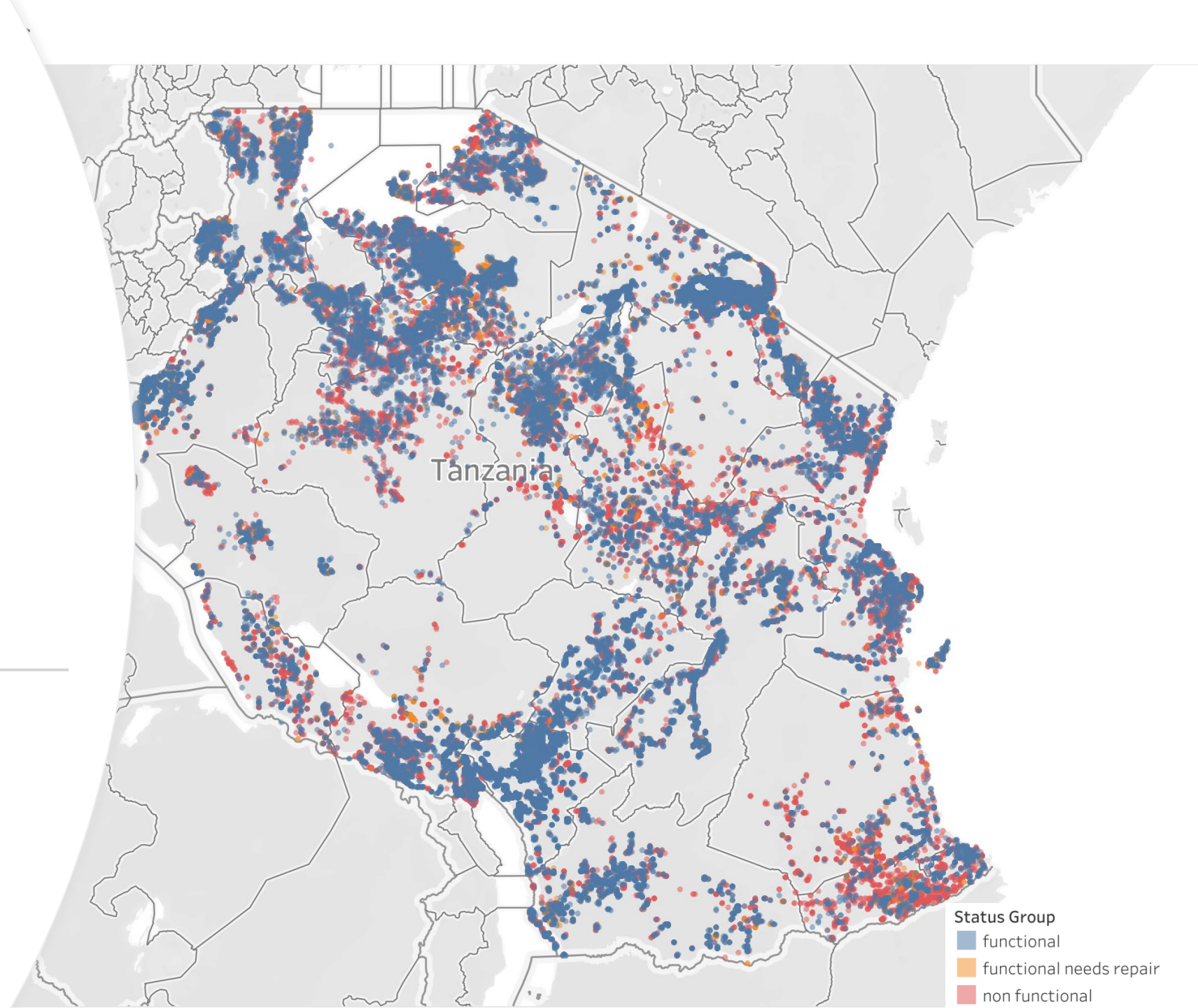


Business Understanding

- The big picture
 - How should the program organize new construction and rehabilitation going into the future
 1. Understand the overall status of wells
 - What do we have and how did we get here.
 - Our target variable in this case is the functionality of individual wells
 - Modeling as a pseudo-auditing system
 2. Managing maintenance and new construction
 - Finding patterns that will give insight into what wells are likely to fail in the short term
 - When well construction must occur, how to mitigate the risk of non-functionality

There's Water Everywhere...

This is an overview of the various status groups by geographic location



Data Understanding

Data Details:

Size: Originally 59000 x 40

Missing Data : Less than 5% missing overall

Data Quirks:

Features: Vast Number of Unique Categorical Features

Place Holder values: Several Columns had existing imputed data

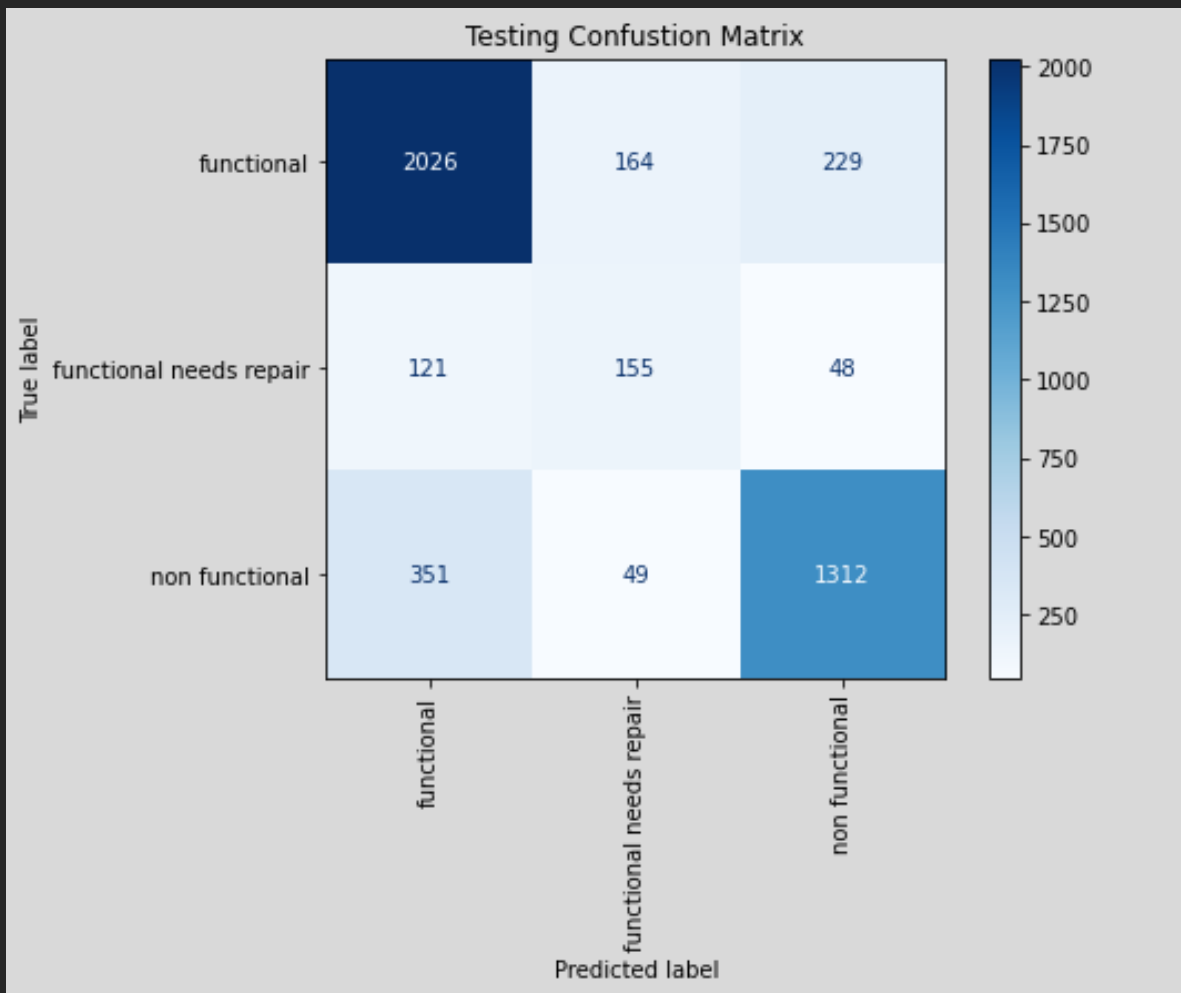
Feature Engineering:

Categorical Columns with thousands of unique entries were binned into groups with similar frequencies

Rare Labels: Incredibly rare labels were absorbed into existing “other” classes within each column

Datetime: Datetime objects were re-engineered into separate columns. Categorical columns for months, day of the week and continuous columns for the calendar day and year.

Model Evaluation



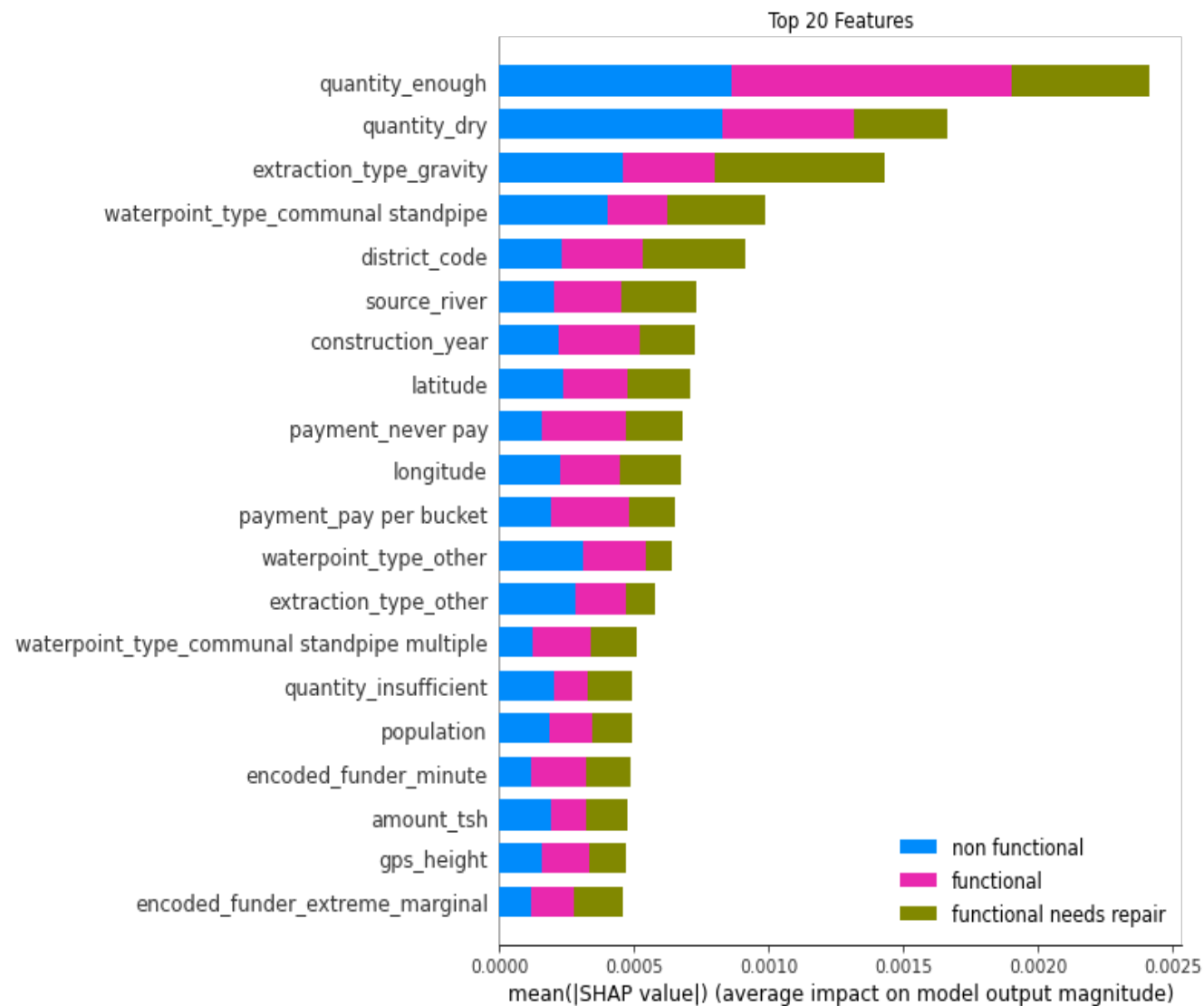
Stack Train Report

| | functional | functional needs repair | non functional | accuracy | macro avg | weighted avg |
|-----------|--------------|-------------------------|----------------|----------|--------------|--------------|
| precision | 0.901453 | 0.930289 | 0.949718 | 0.926776 | 0.927153 | 0.927153 |
| recall | 0.910757 | 0.960623 | 0.908948 | 0.926776 | 0.926776 | 0.926776 |
| f1-score | 0.906081 | 0.945213 | 0.928886 | 0.926776 | 0.926727 | 0.926727 |
| support | 29840.000000 | 29840.000000 | 29840.000000 | 0.926776 | 89520.000000 | 89520.000000 |

Stack Test Report

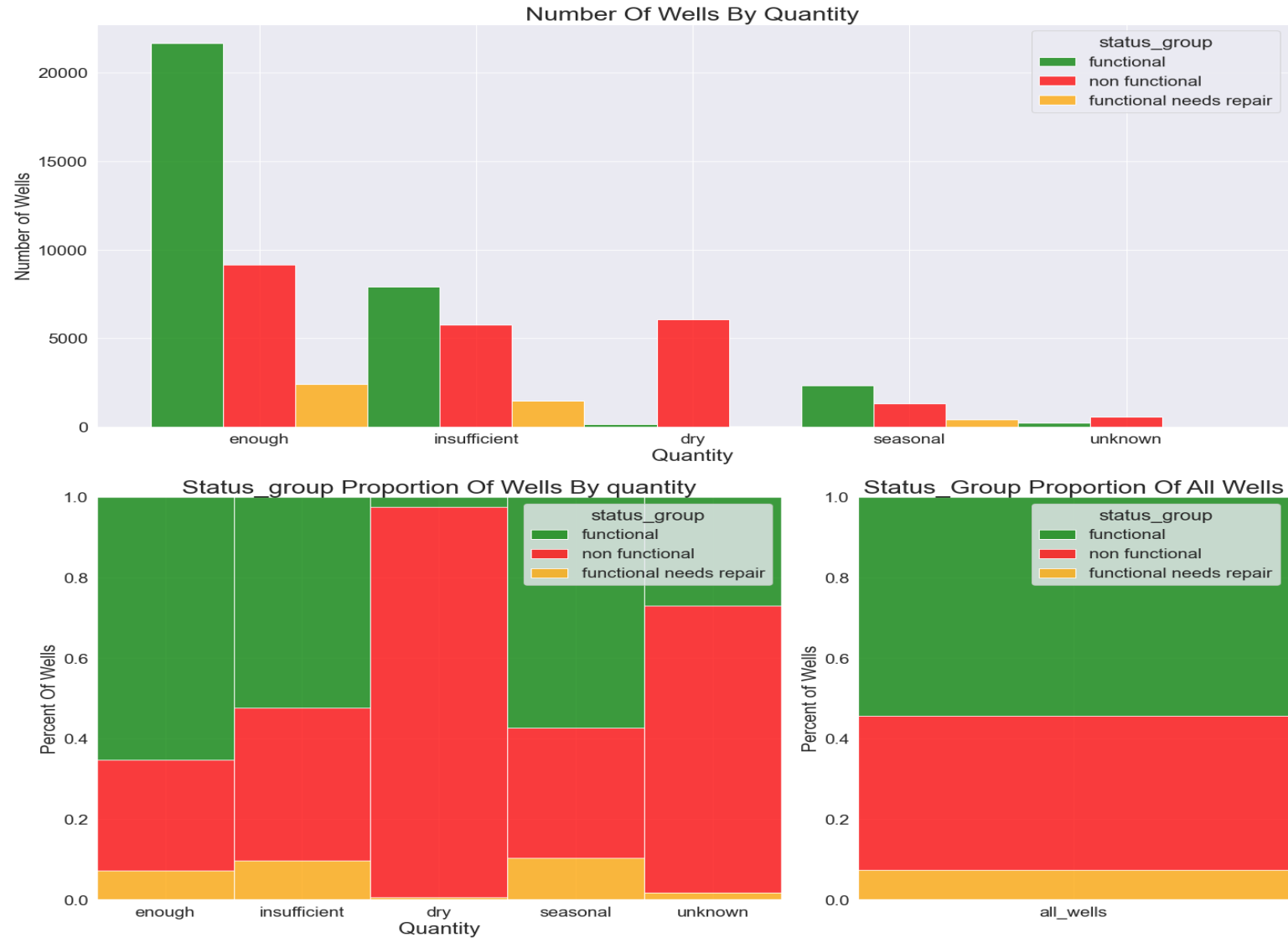
| | functional | functional needs repair | non functional | accuracy | macro avg | weighted avg |
|-----------|-------------|-------------------------|----------------|----------|-------------|--------------|
| precision | 0.826522 | 0.403341 | 0.818681 | 0.783838 | 0.682848 | 0.792732 |
| recall | 0.819347 | 0.521605 | 0.783294 | 0.783838 | 0.708082 | 0.783838 |
| f1-score | 0.822919 | 0.454913 | 0.800597 | 0.783838 | 0.692809 | 0.787577 |
| support | 2419.000000 | 324.000000 | 1712.000000 | 0.783838 | 4455.000000 | 4455.000000 |

Overall Feature Importance



Recommendations:

1. Additional site analysis before building or rebuilding wells that are labelled (or likely to be labelled) dry.
2. If wells must be built in areas that are likely to run dry, perhaps alternative water sources should be explored



Feature Importance

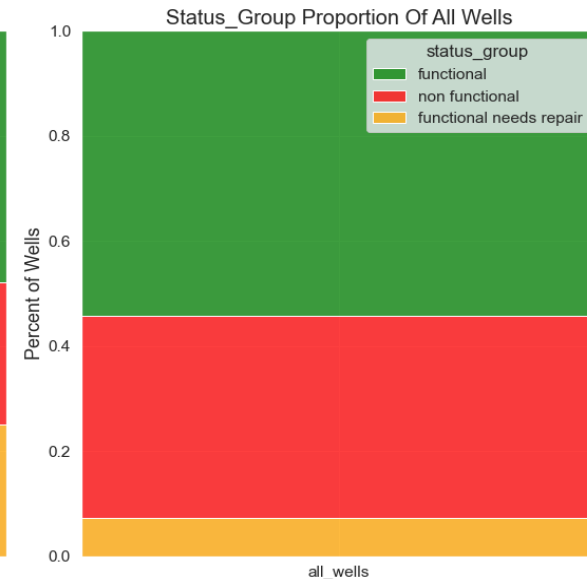
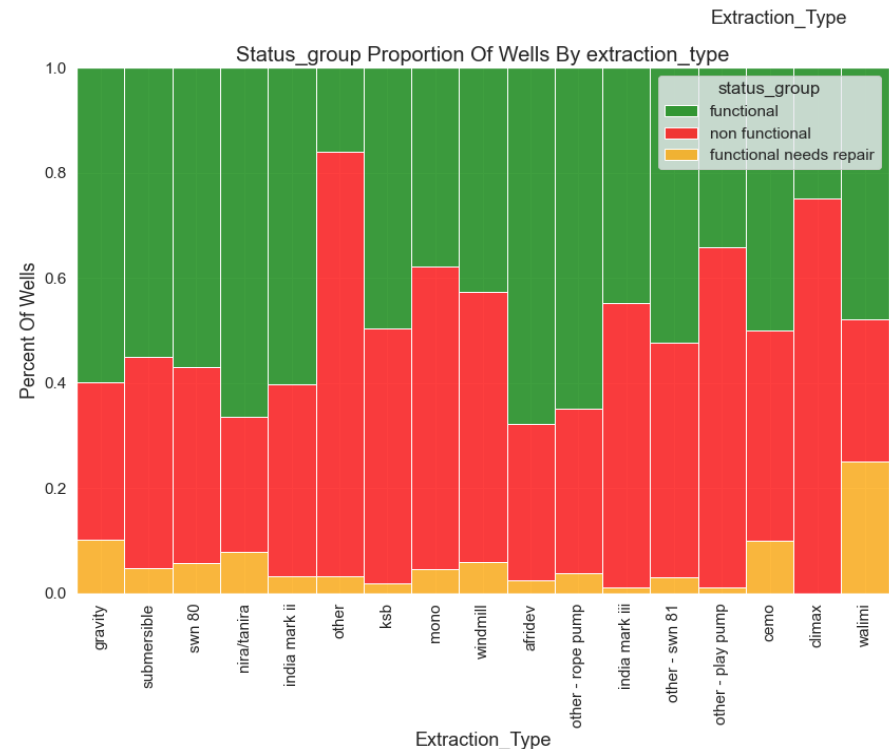
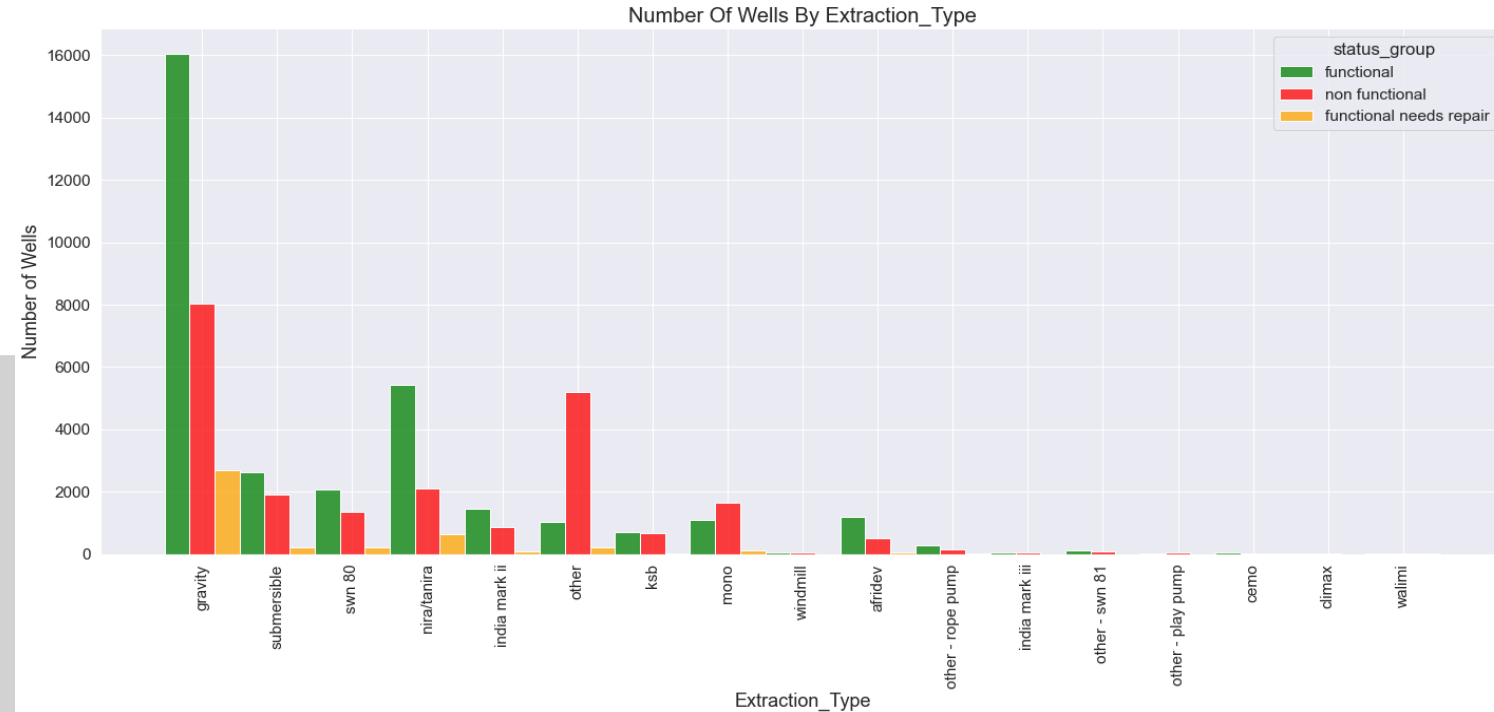
- Dry wells are highly likely to be non-functional
- Enough quantity wells lean towards functionality when compared to the average

Feature Importance

- Gravity wells are slightly more likely to be functional than average, however, they are also more likely to require more maintenance.
- Other type wells are incredibly likely to be non-functional

Recommendations:

1. New builds should avoid well types that would fall under the other category.
2. Retro-fit old wells that are other type if possible.
3. Gravity wells should be reconsidered as the primary well extraction type due to maintenance issues.

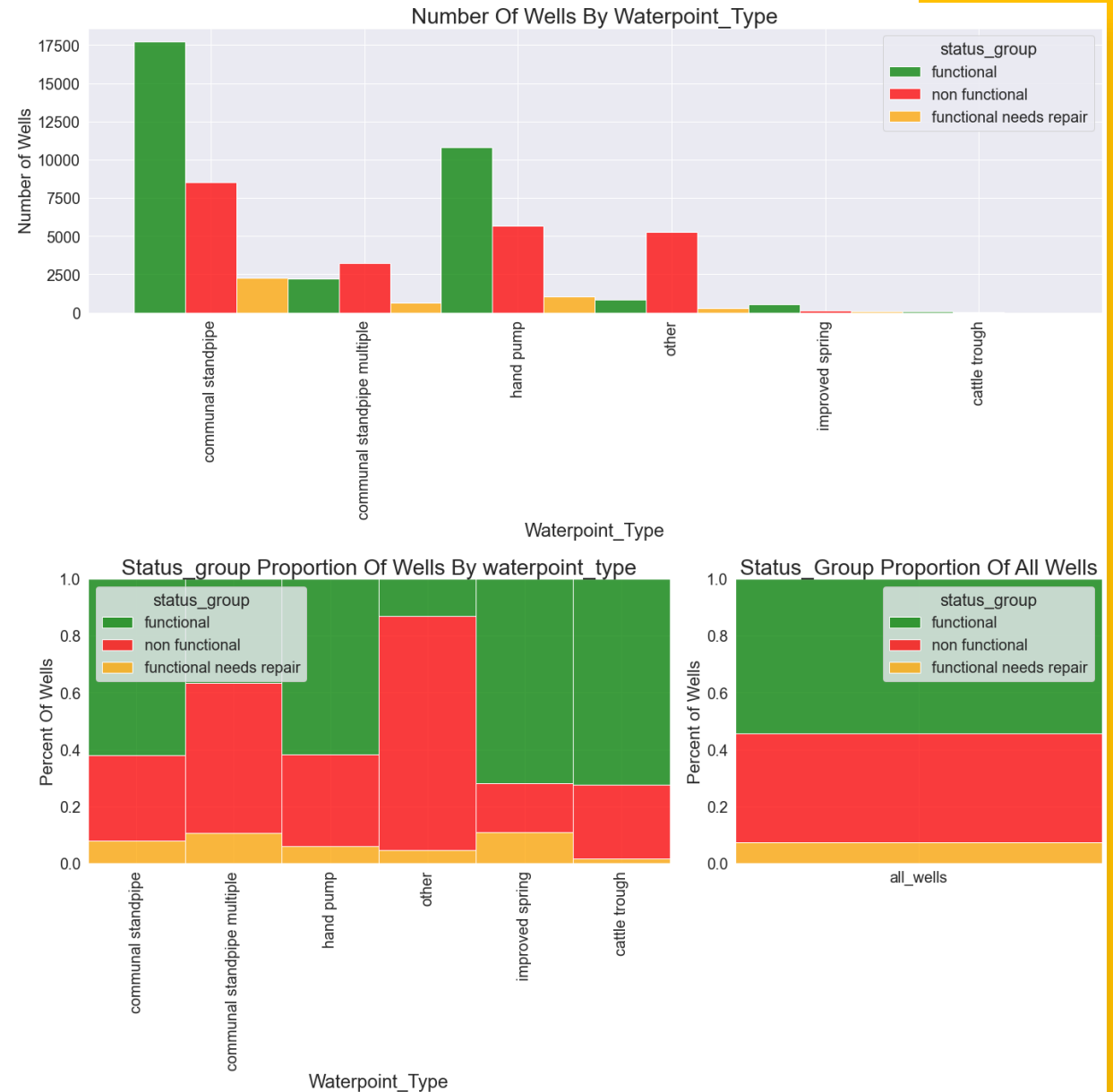


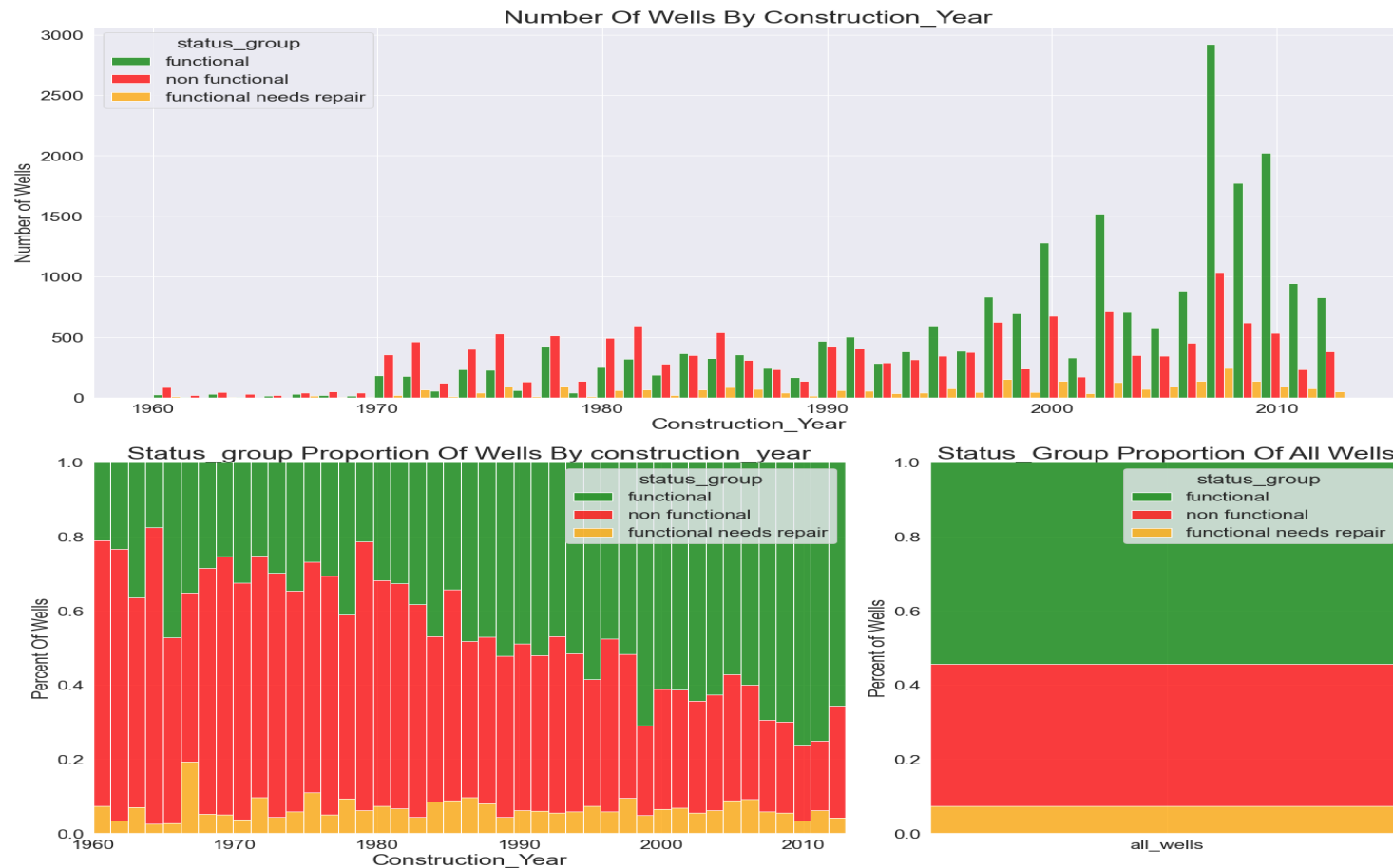
Features Importance

- Communal standpipe appears to have a slight inclination towards functionality
- Communal standpipe multiple and other types are heavily biased towards non-functional.

Recommendations

1. New builds and retro-fits should avoid the communal standpipe multiple and other types as the waterpoint.





Feature Importance

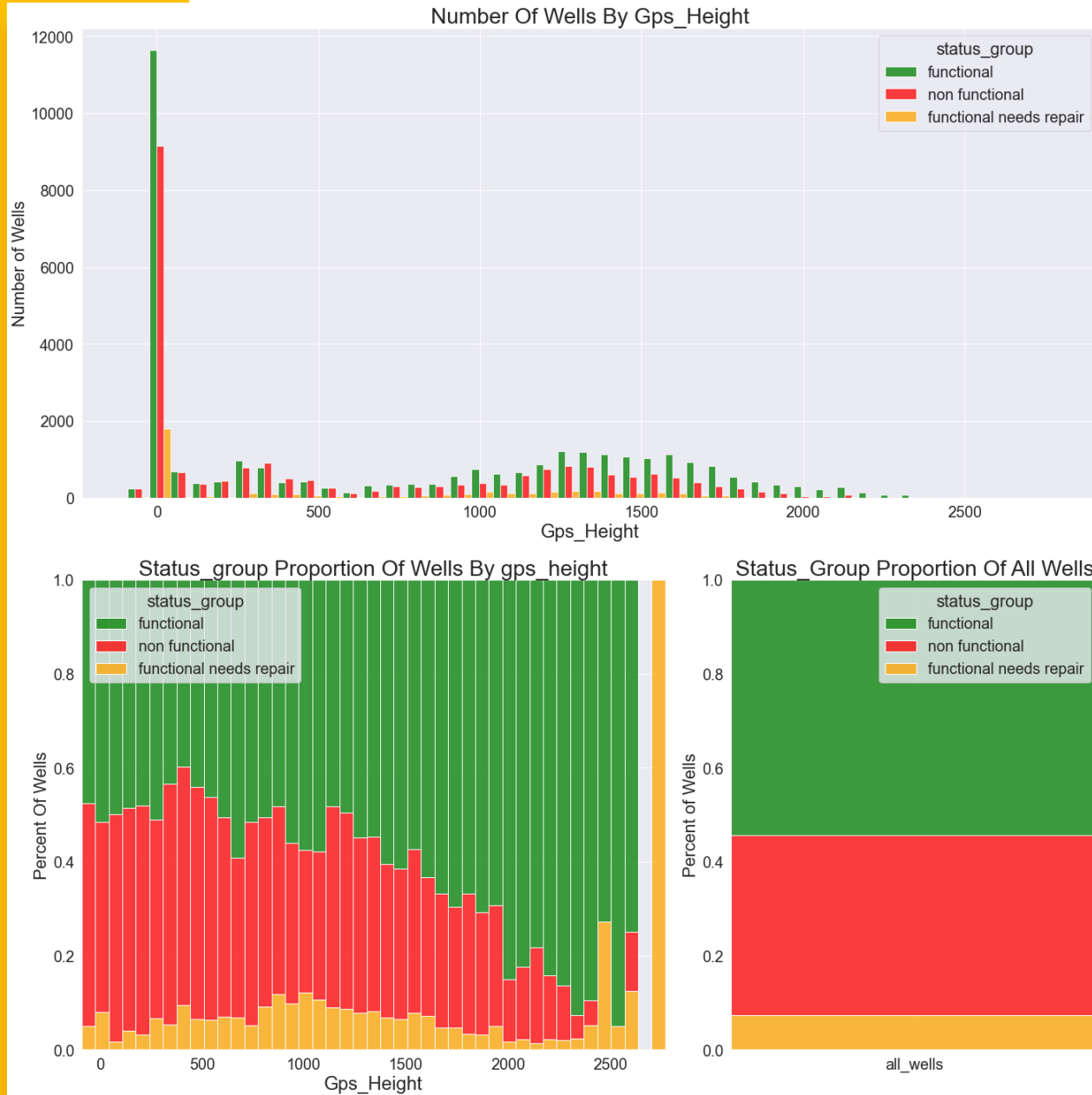
- Older wells are significantly more likely to be non-functional than newer wells.

Recommendations

- Wells older than 1990 should be targeted for rebuilds
- Wells constructed in the 1990s should be placed on maintenance schedules to limit additional non-functionality

Observation

- Provides a loose roadmap for the aging pattern of modern wells.

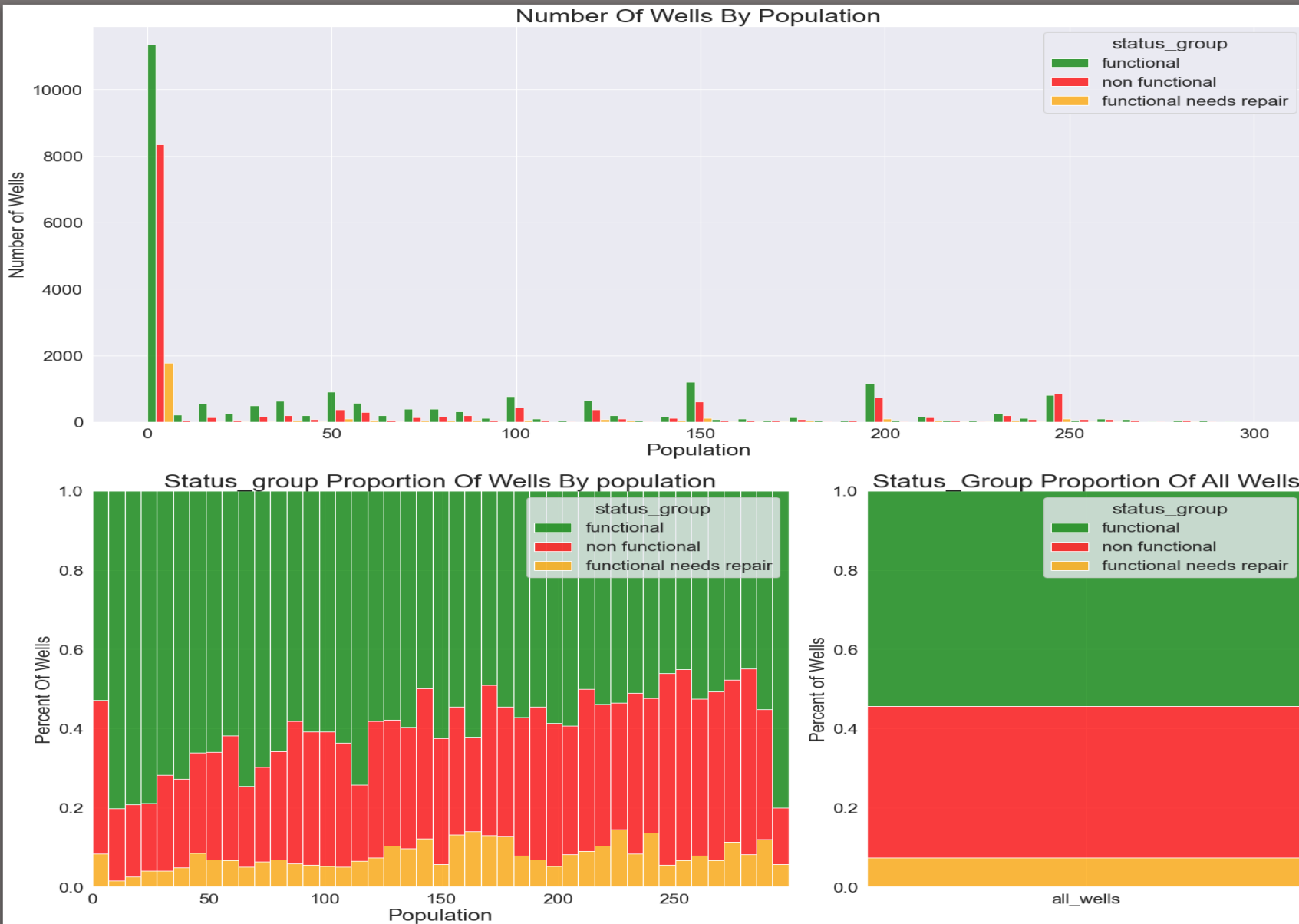


Feature Importance

- Wells at higher elevation are much more likely to be functional than average.

Recommendations

- When planning a well in any given area. Lean towards picking locations with higher elevations.



Feature Importance

- Overall, the probability of wells being non-functional increases as population increases.
- The notable exception is with the population being 0 or 1.

Recommendations

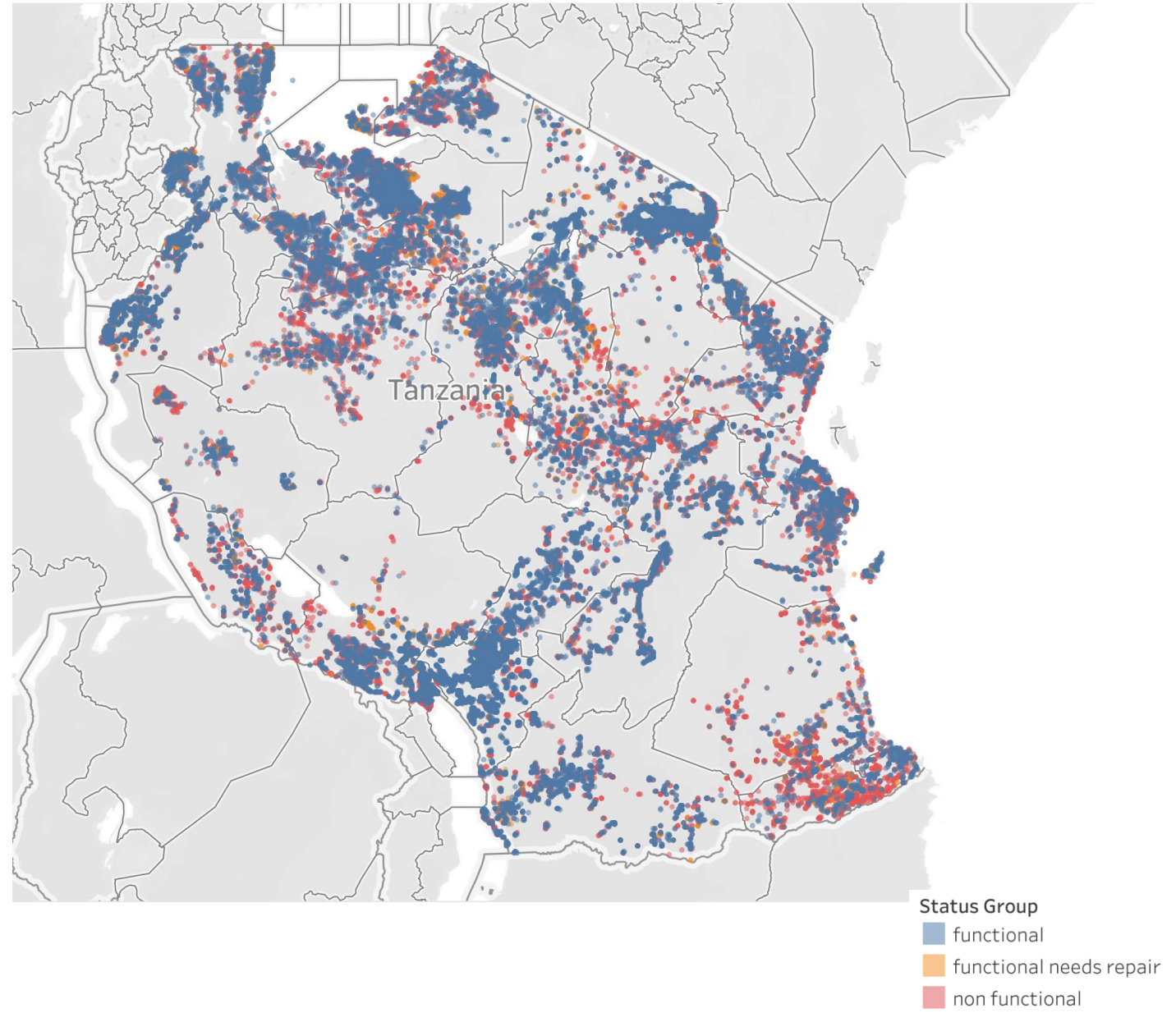
1. The sweet spot appears to be, $115 > \text{populations} > 8$. Build wells near populations of this range.

Feature Importance

- Certain clusters around the country are more or less likely to be functional.

Recommendations

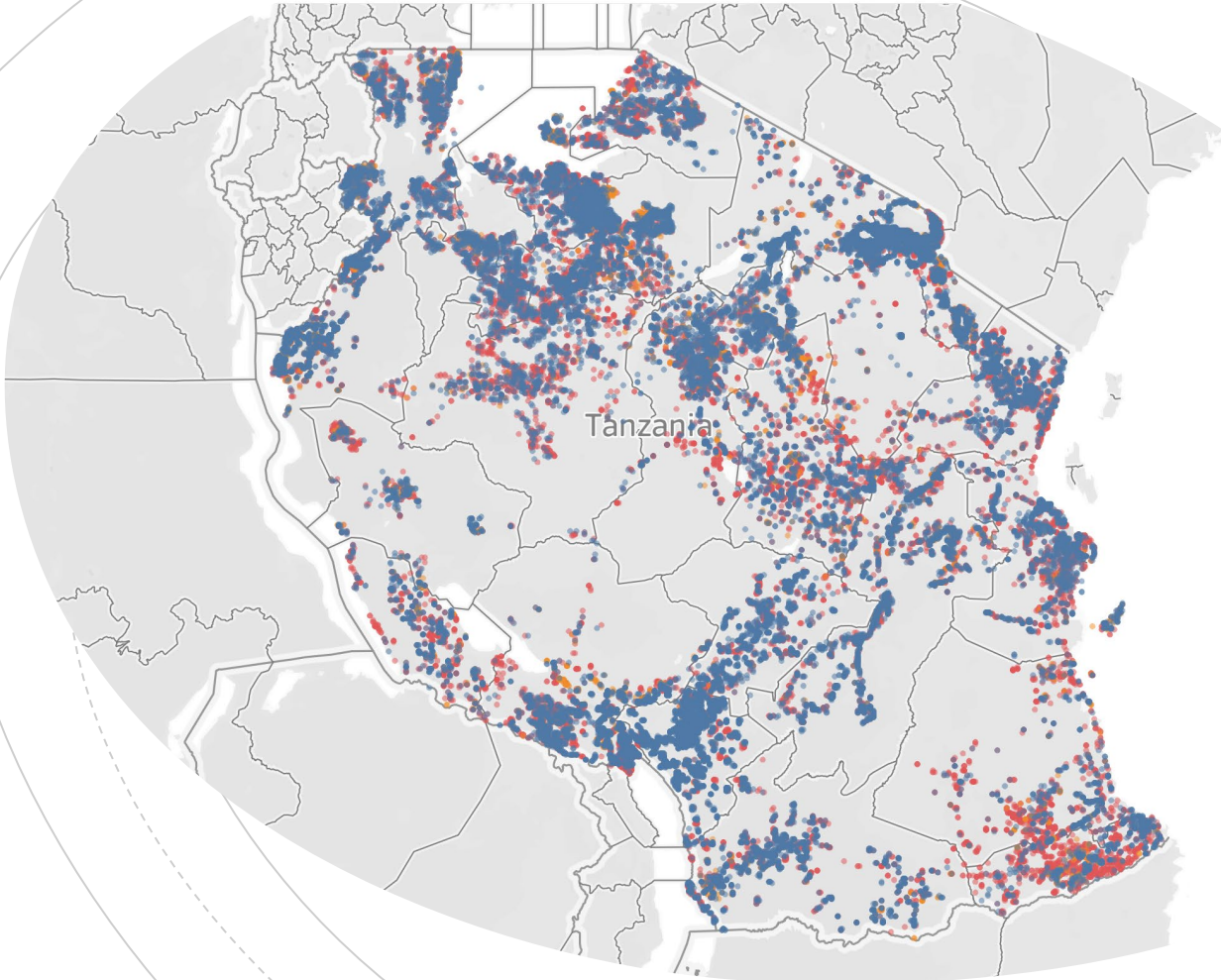
1. Further investigate before adding or rehabilitating wells in non-functional clusters.



Future Work

There's still lots to do

1. Model Improvements
 1. Use additional business and program insights to adjust feature engineering.
 2. Experiment with additional imputation techniques rather than using a static fill value.





Thank you