

Tanzania Waterpoint Functionality Classification using Random Forest

Author: Chi Bui
05/2021



Overview

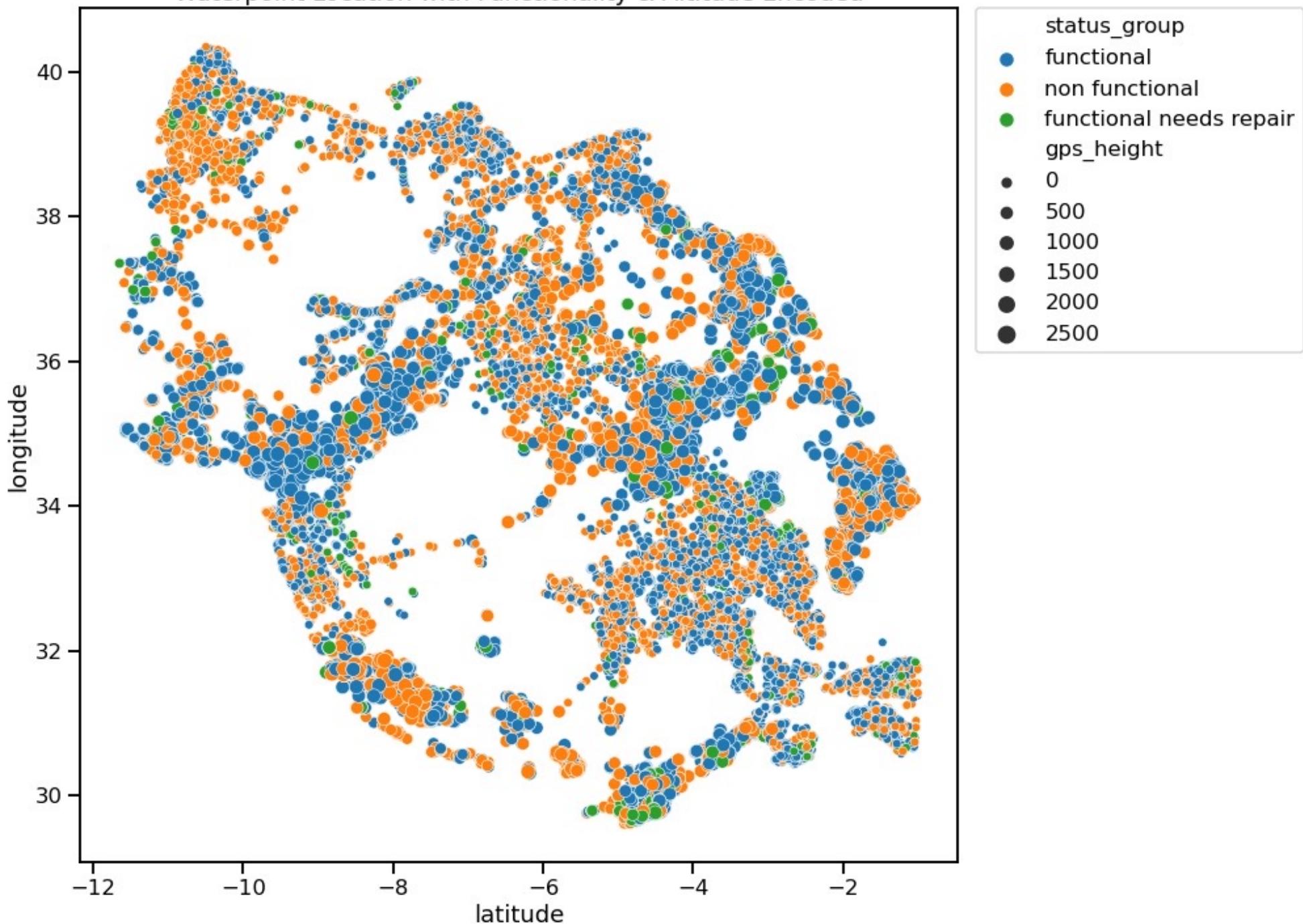
Business Problem

- Better understanding of which features might be contributing to waterpoints functionality
- Improve maintenance operations
- Ensure that safe, clean water becomes available to more people across the country

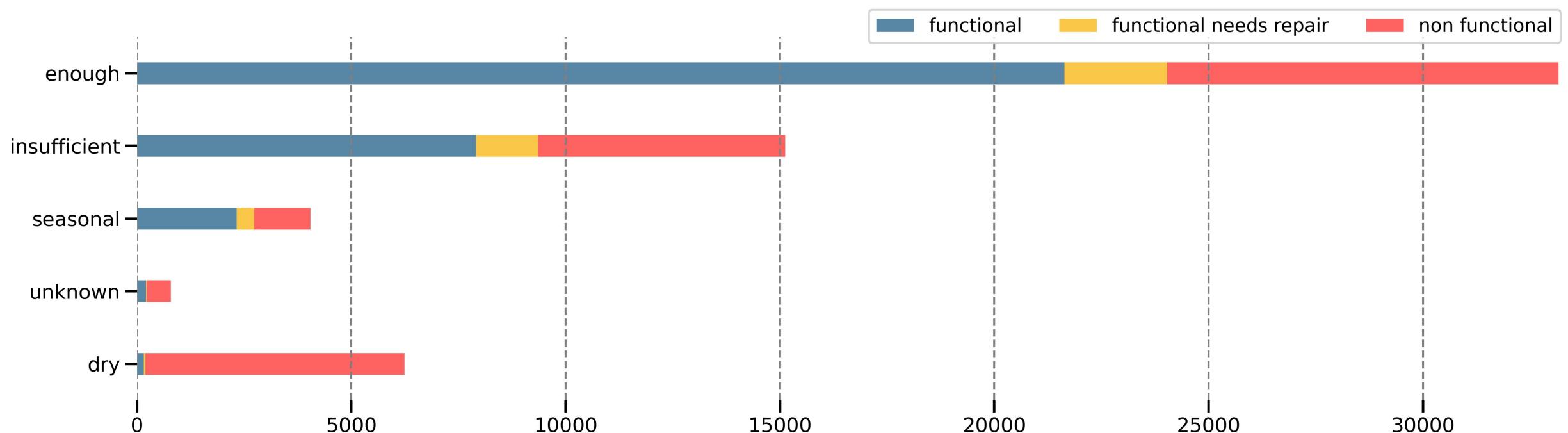
Dataset

- <https://www.drivendata.org/>
- **Taarifa** and the **Tanzanian Ministry of Water**
- 59,364 datapoints and 40 columns
- **status_group**

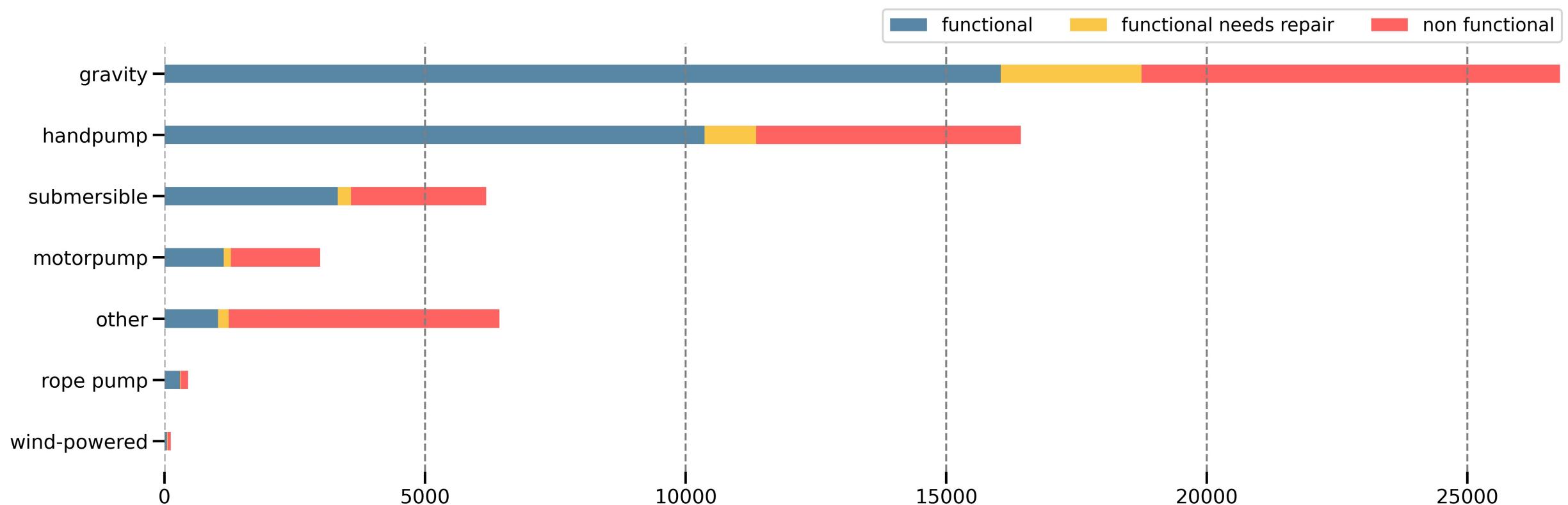
Waterpoint Location with Functionality & Altitude Encoded



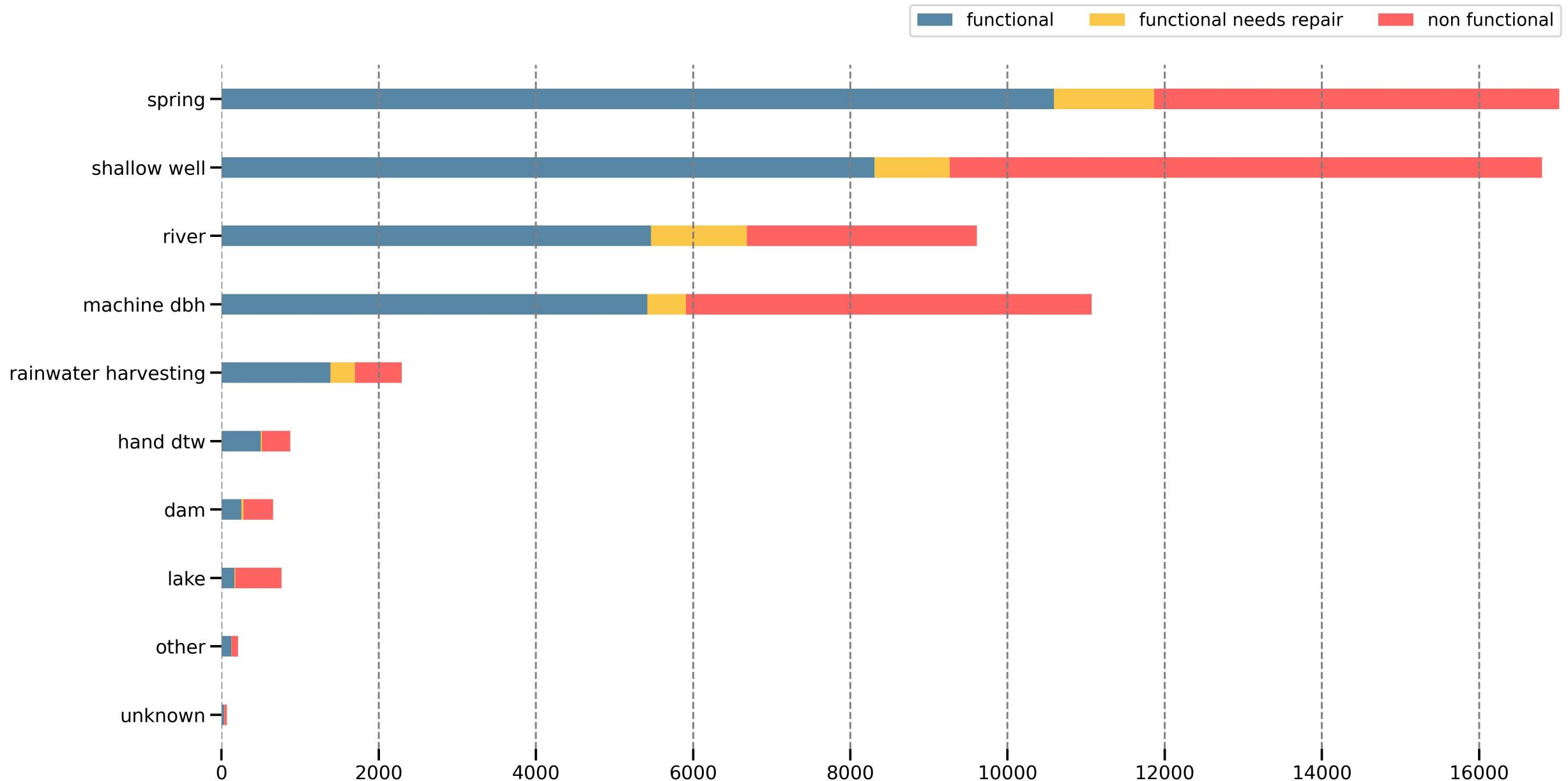
Proportion of Well Functionality Statuses Grouped by Quantity



Well Functionality Status by Extraction Type Class

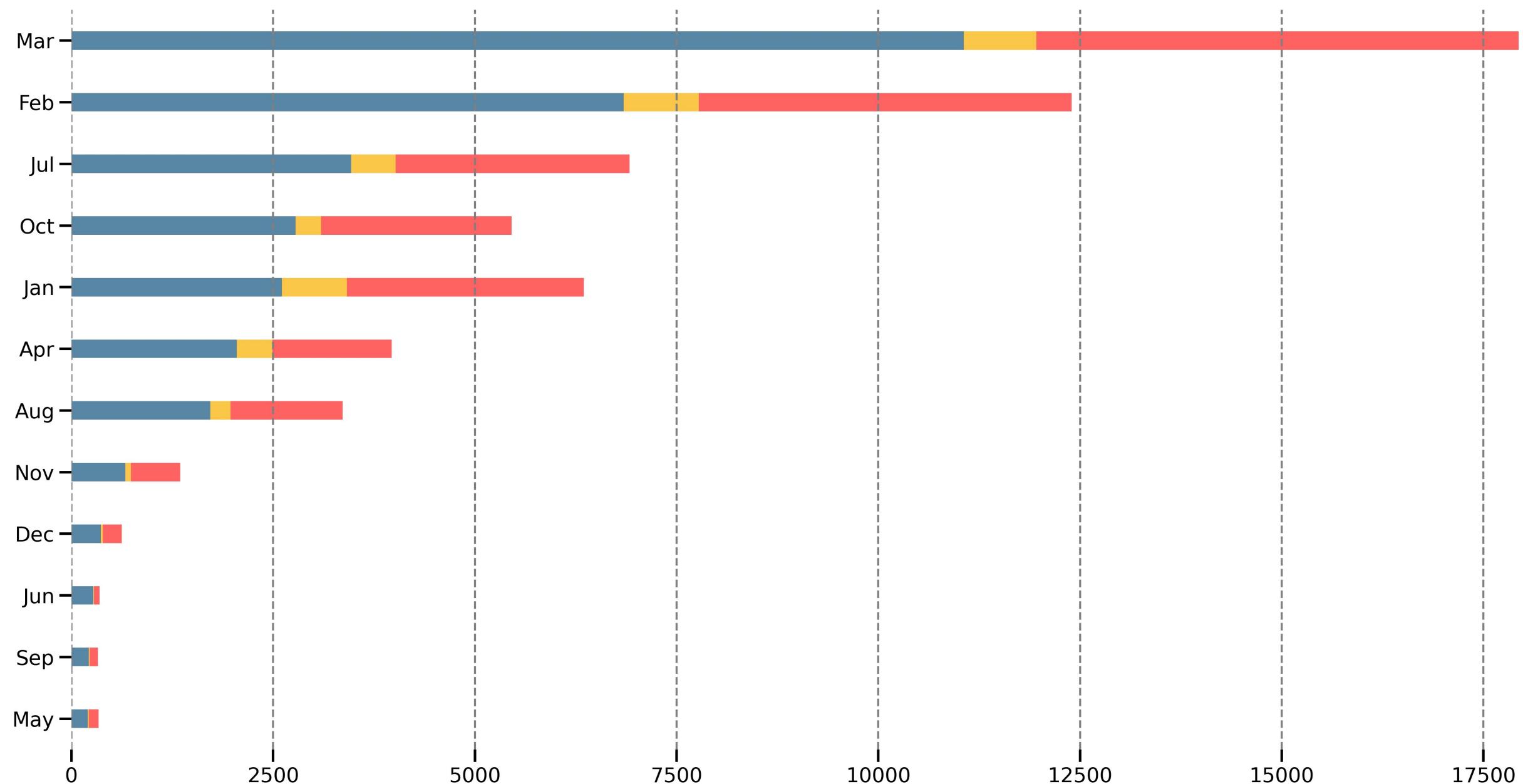


Well Functionality Status by Water Source



Proportion of Well Functionality Status Grouped by Month of Recording

functional functional needs repair non functional



Challenges

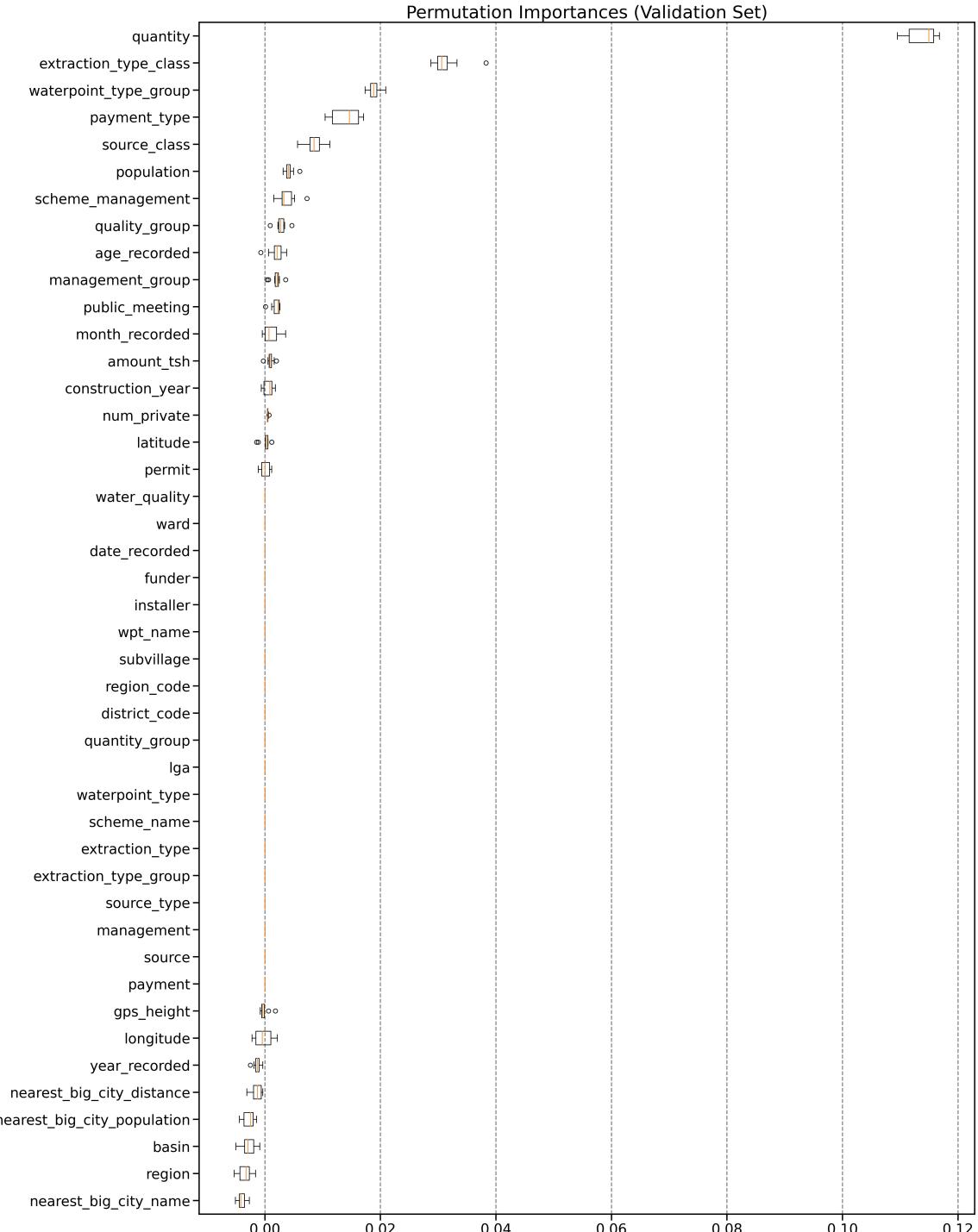
Numerical Features

- Zeros (0) as placeholder

Categorical Features

- unknown
- other

Inferential Model

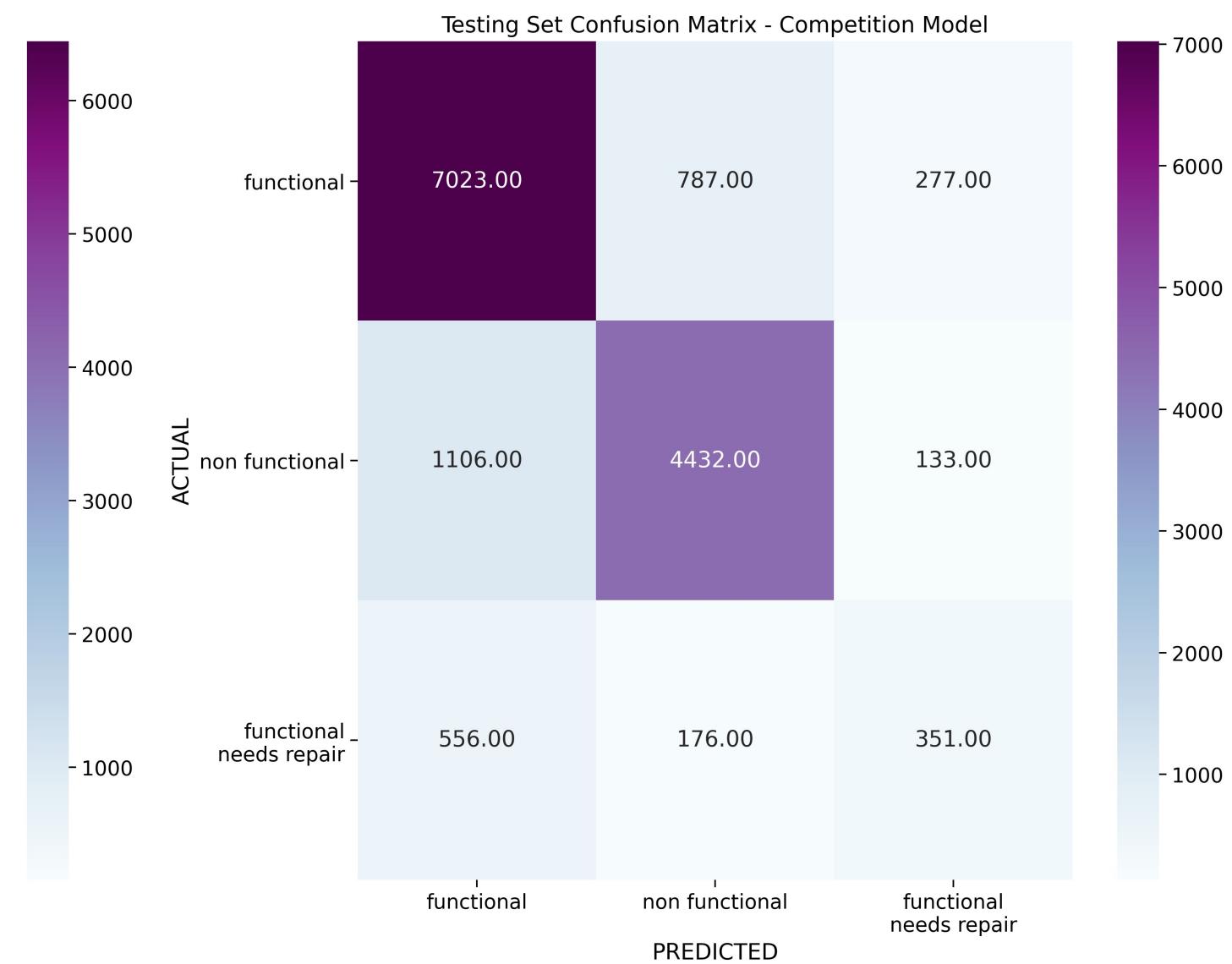
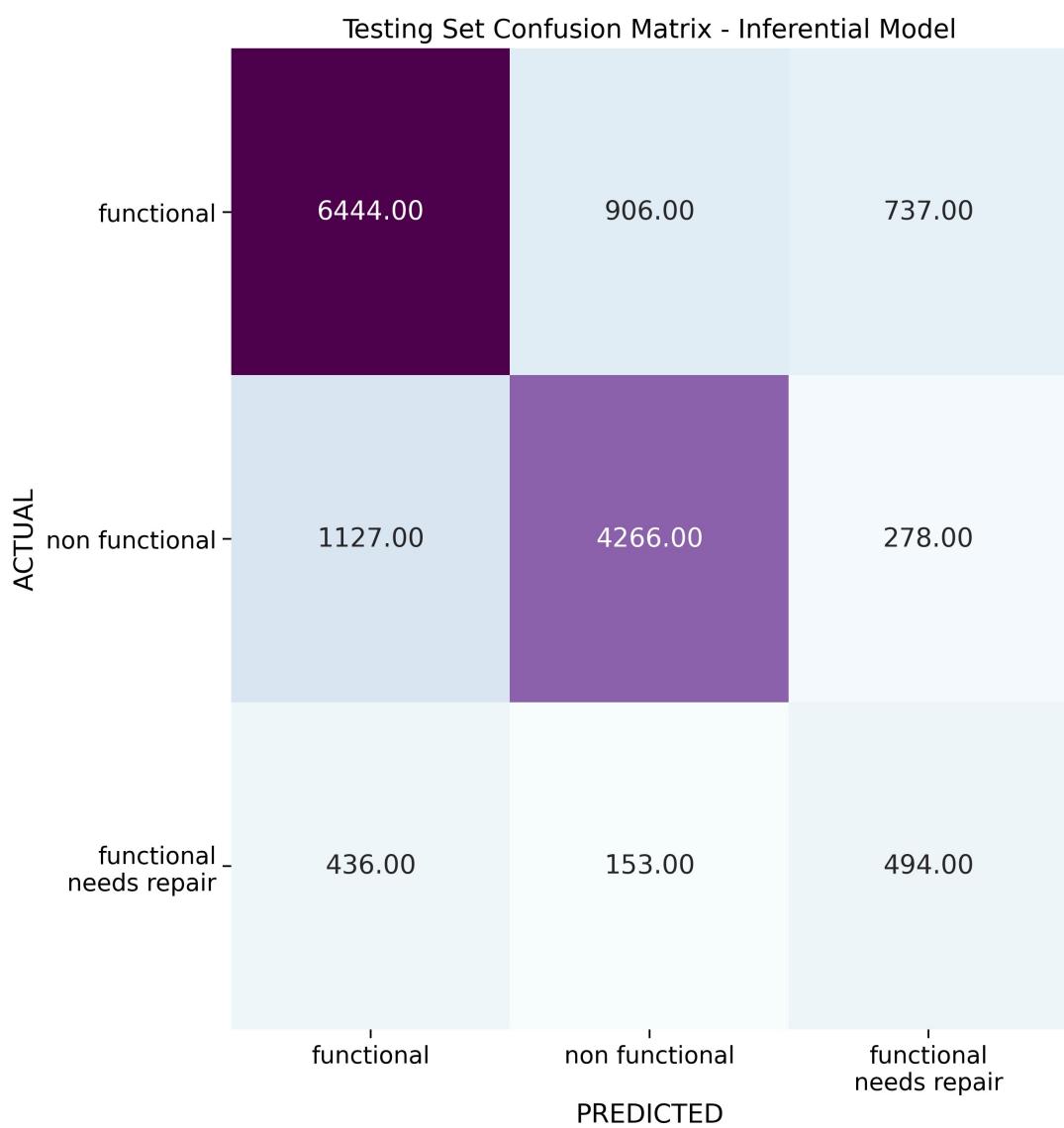


Testing Set Confusion Matrix - Inferential Model



Competition Model

- Maximize Accuracy
- No Oversampling
- Competition Score: 81.11%



Next Steps

- Other oversampling methods
- Different hyperparameters

Thank you!

Email: chibui191@gmail.com

Phone: +1(813) 607-7476

Tampa, Florida