Tanzanian Water Wells

By: Andrew J Moy

Goals & Methodology

- Our goal is to predict the functionality of well in Tanzania

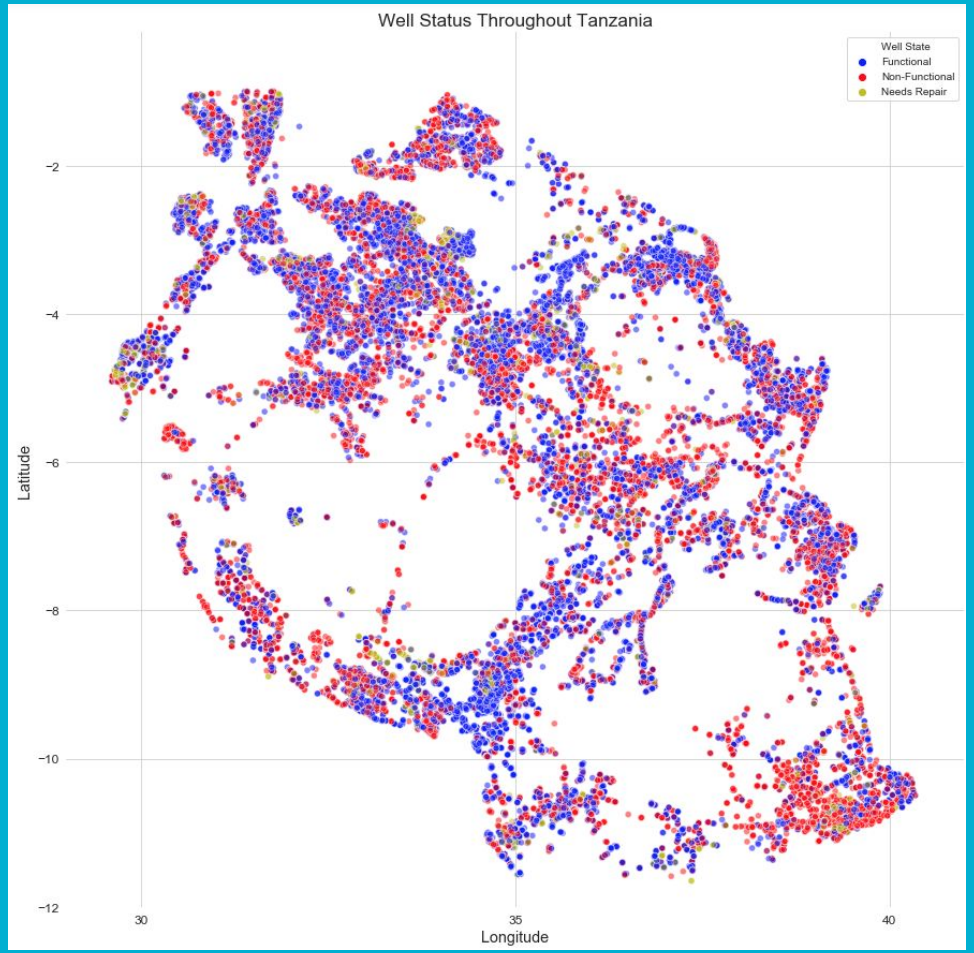
Using the OSEMN methodology.

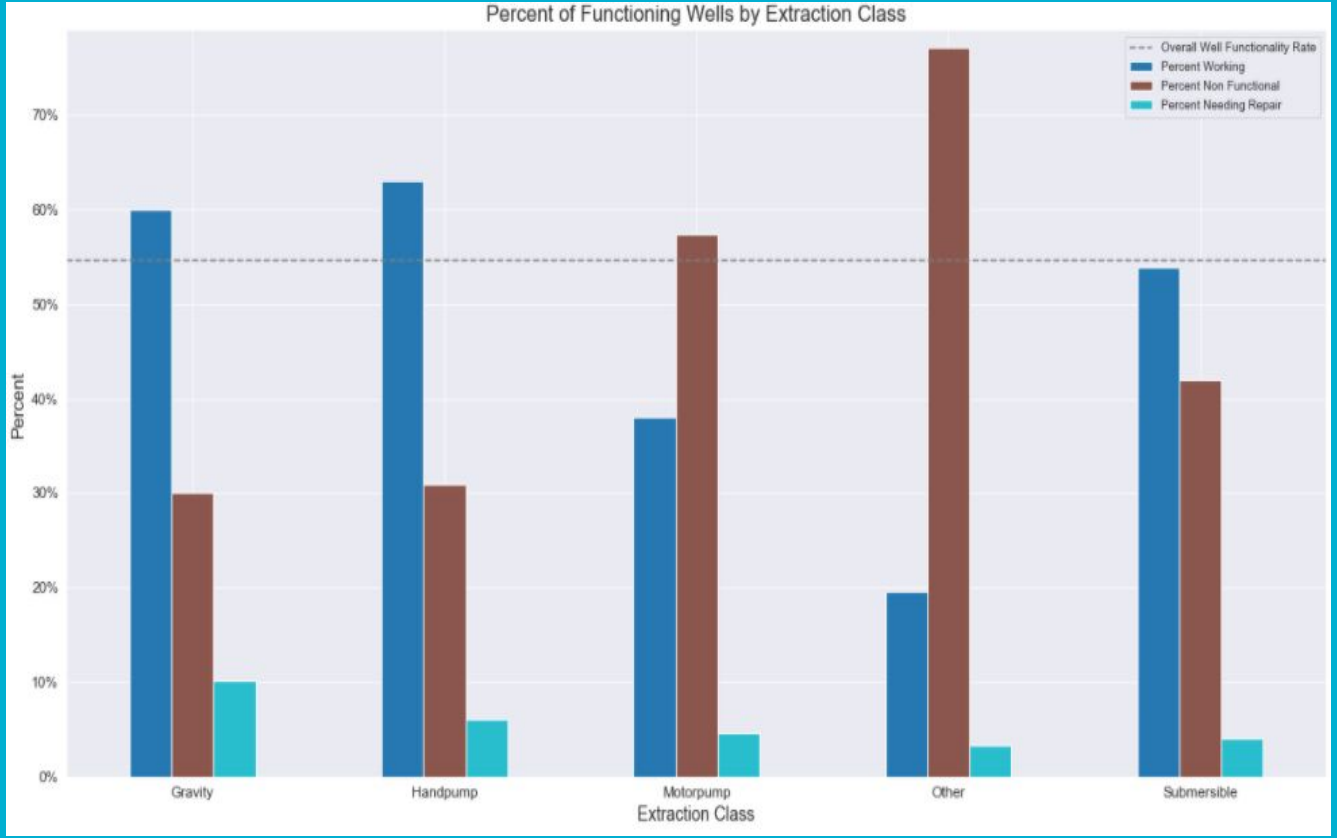
1. We obtained the data from Taarifa
2. Scrubbed data - removed, replaced, imputed certain values to various place-holders
3. Explore - Understand the data and look for patterns or trends
4. Model & Interpret - Built and test models and examine feature importance

Is there any geographic pattern for functioning wells?

Slightly. It does appear that the wells tend to cluster based on functionality.

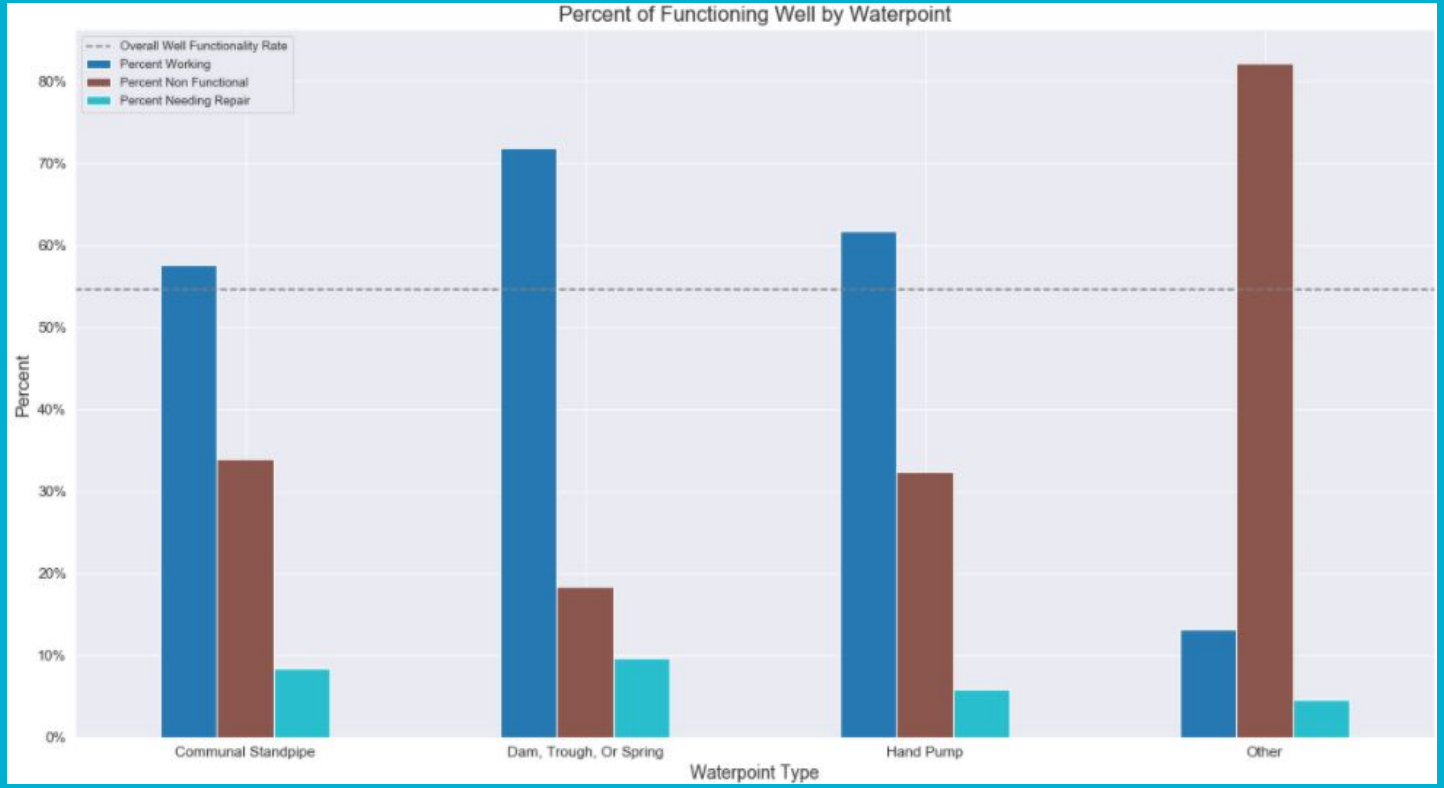
It can just correlate to the amount of available water.



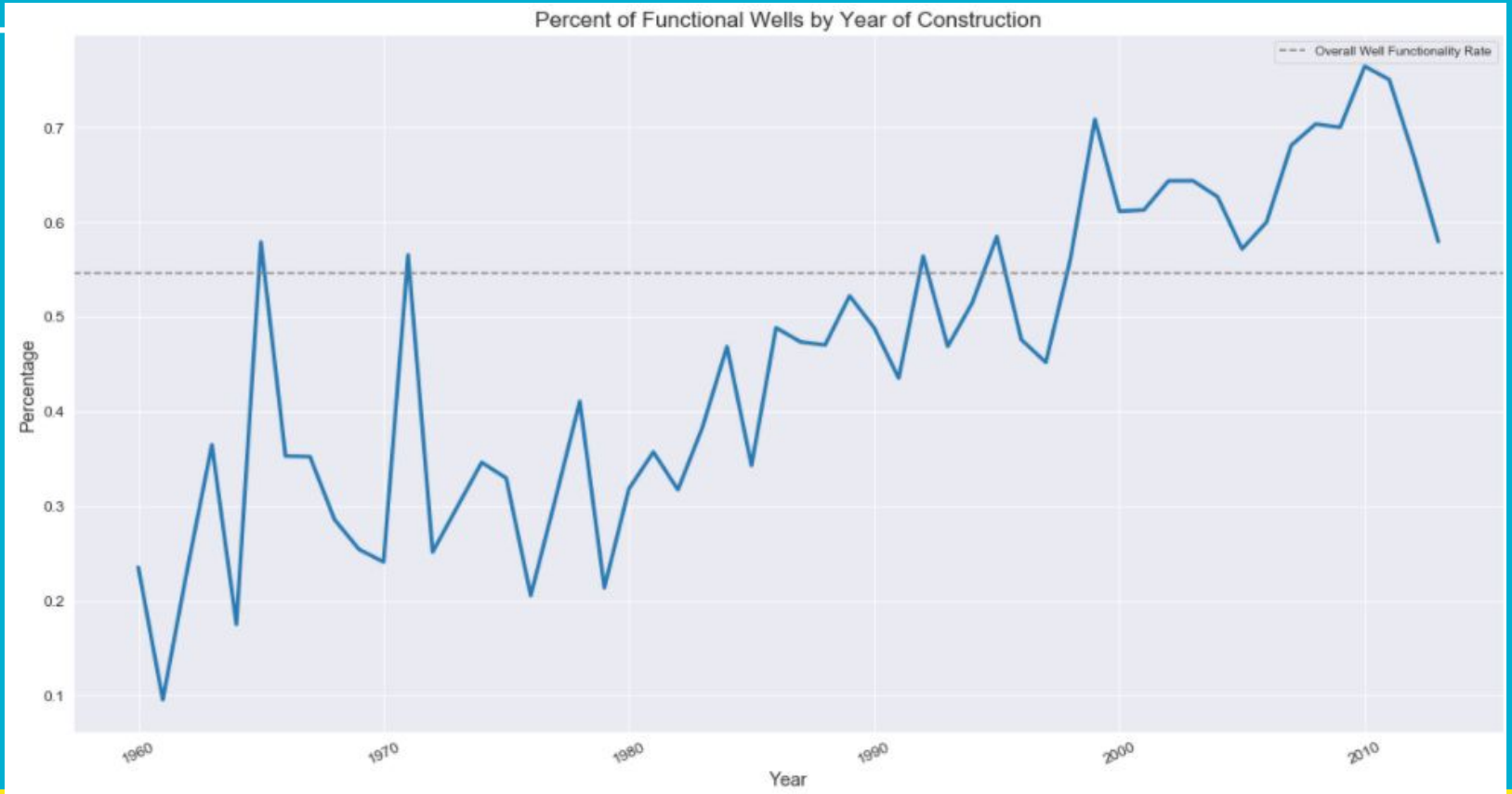


Does the functionality vary depending on the way its extracted?

What about the waterpoint?



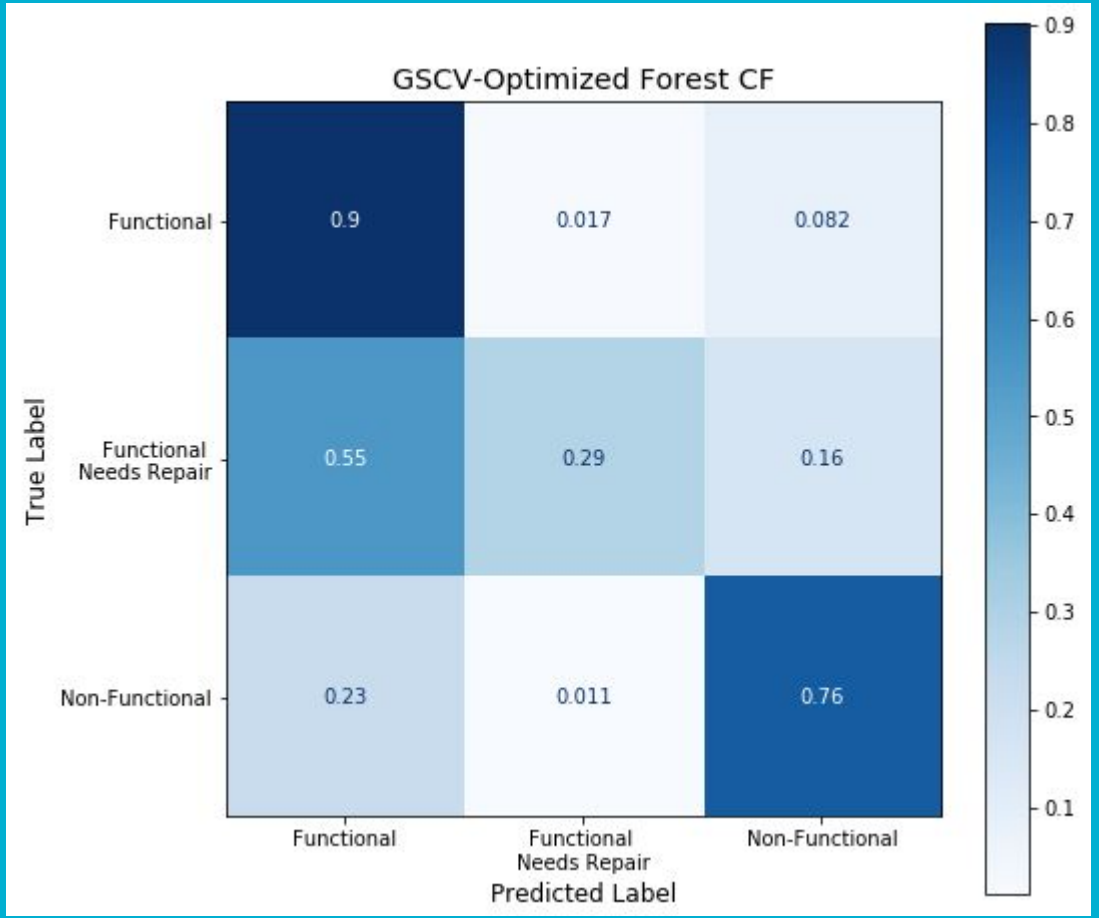
How does the functionality factor with the age of the well?



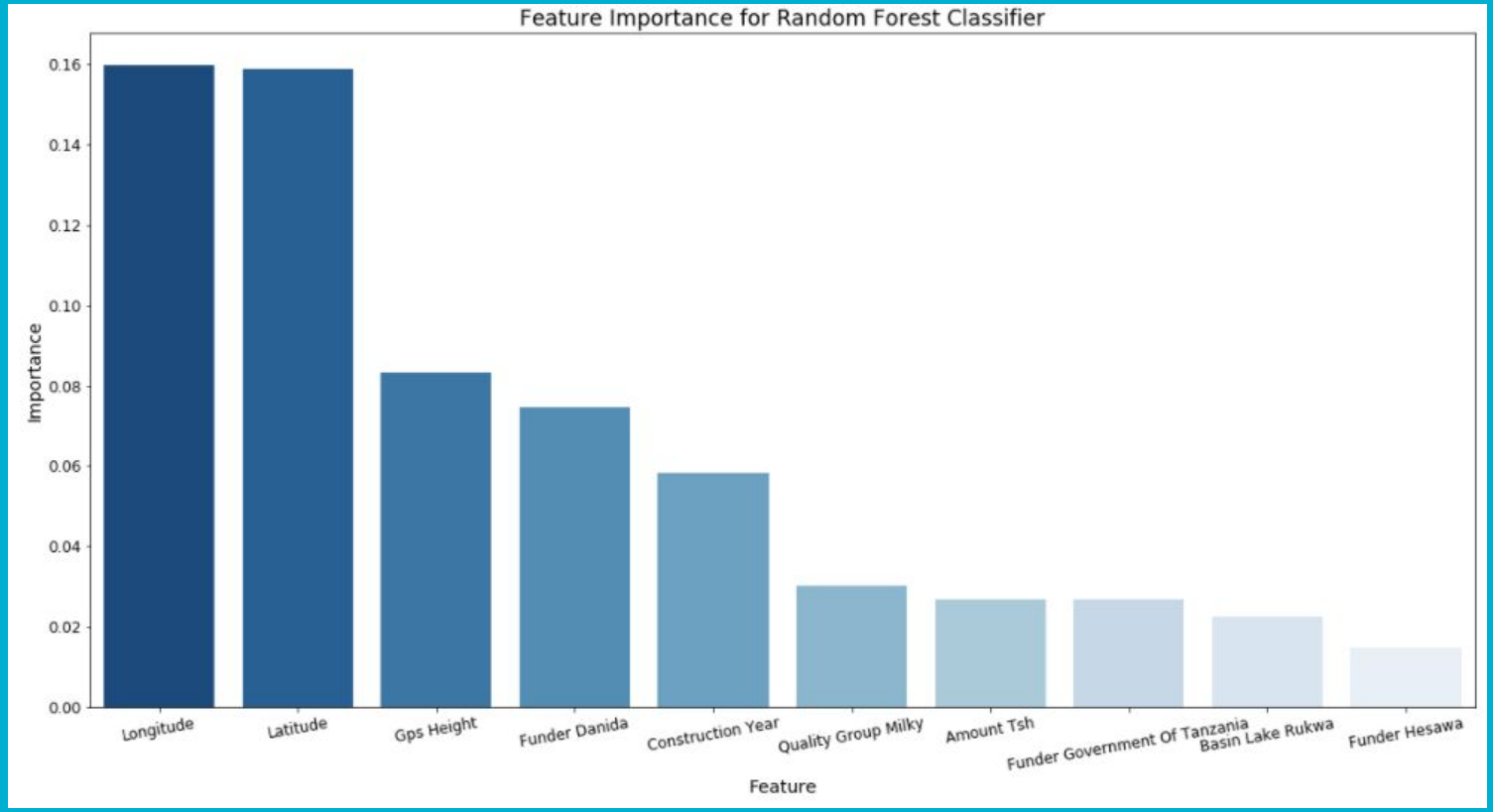
Modeling the Data

A Random Forest classifier was built and optimized for accuracy.

It did well classifying the functional well correctly. However, it does make more mistakes with non-functional and need repair.



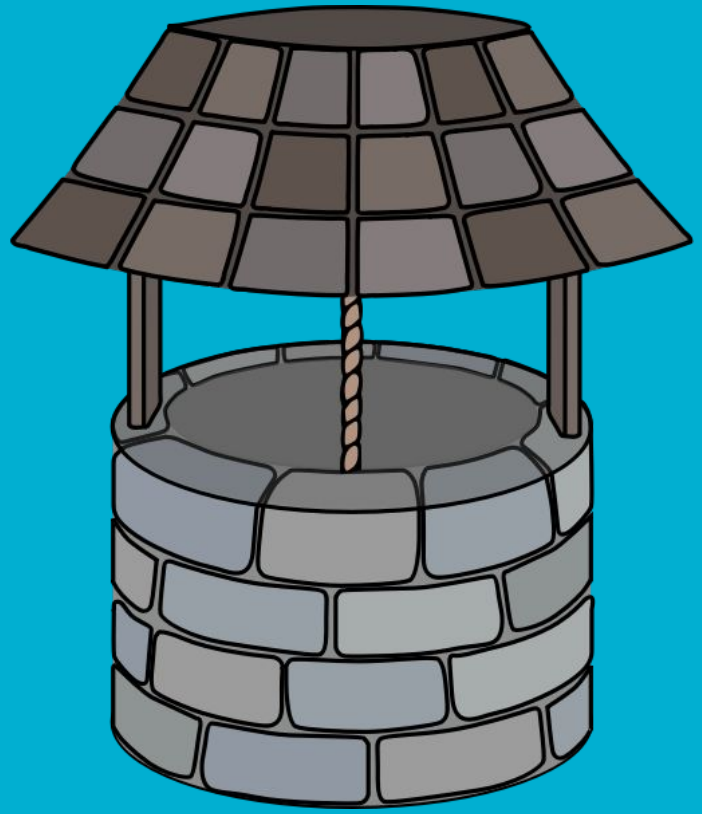
Important Features



Recommended Actions

The most important features were geographic. Find the problem areas in need of repairs and direct resources to those regions.

Focus on gravity and hand-pump extraction methods



Future Work

Dig deeper into the features that would identify well in need of repair rather because those need to be taken care of in a timely manner moving forward.

Geographically, figure out where water is scarce and needed more than others.

Combine more models to increase accuracy.

Thank You!

Taarifa

Tanzanian Ministry of Water

Drivendata

Flatiron