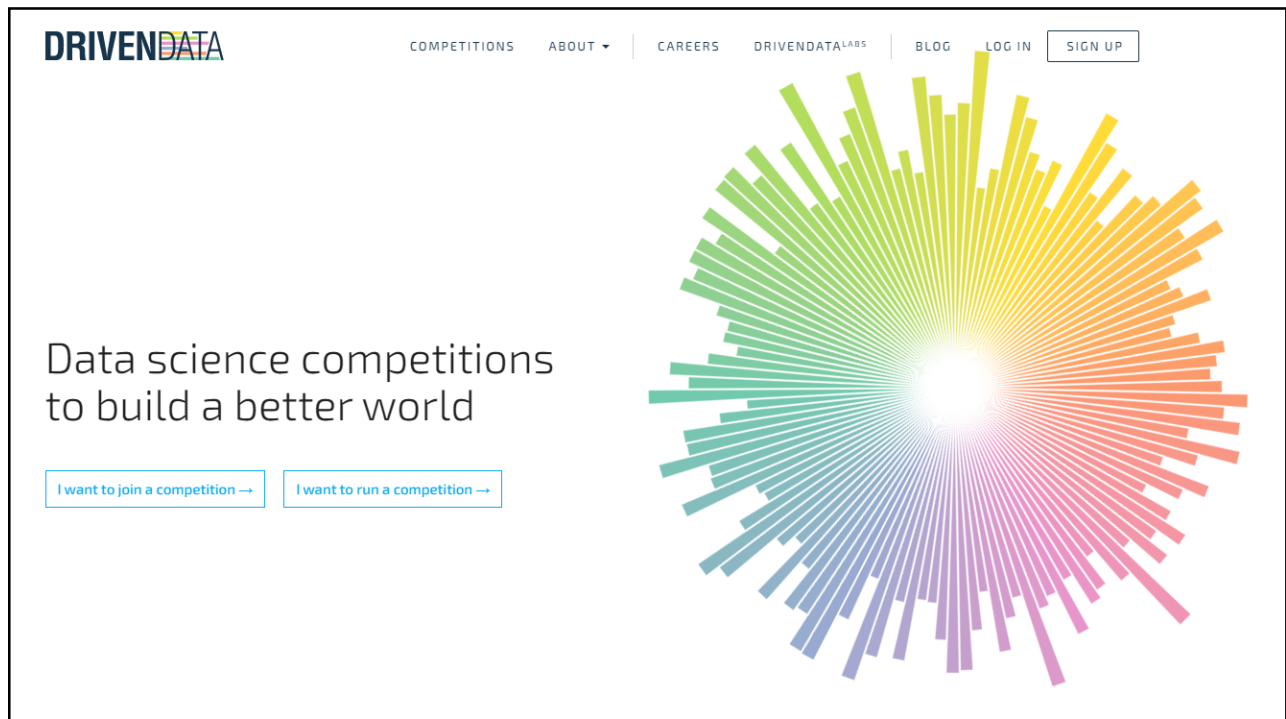


1



2

Problem Statement

- “Using data from [Taarifa](#) and the [Tanzanian Ministry of Water](#), can you predict which pumps are functional, which need some repairs, and which don't work at all? This is an intermediate-level practice competition. Predict one of these three classes based on several variables about what kind of pump is operating, when it was installed, and how it is managed. A smart understanding of which waterpoints will fail can improve maintenance operations and ensure that clean, potable water is available to communities across Tanzania.” – DrivenData

3

What is a well point?



<https://www.thefourthcorner.com/wp-content/uploads/hand-pump-water-well.jpg>

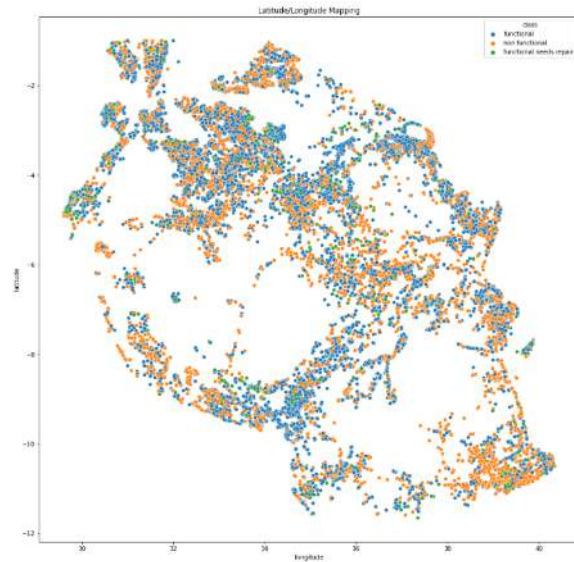
4

Locations of Wellpoints

- Wellpoints are distributed throughout the country

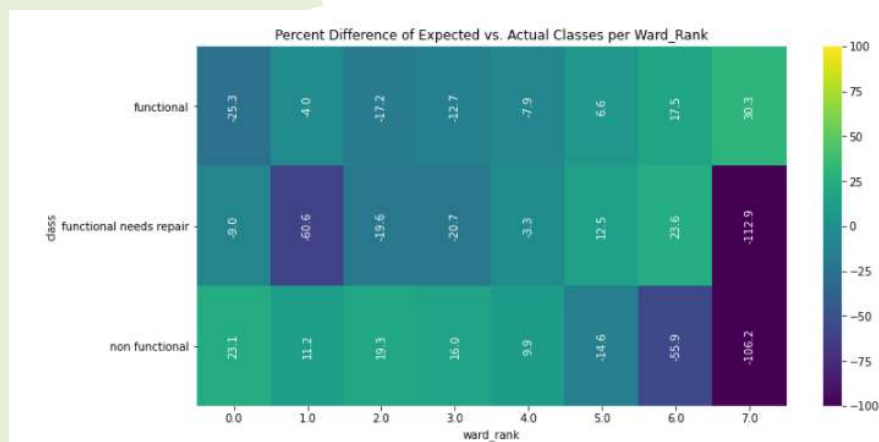
- Location columns include:

Latitude
Longitude
Subvillage
Ward
LGA
Region Code
District Code
Basin



5

Locations of Wellpoints – Ward Rank



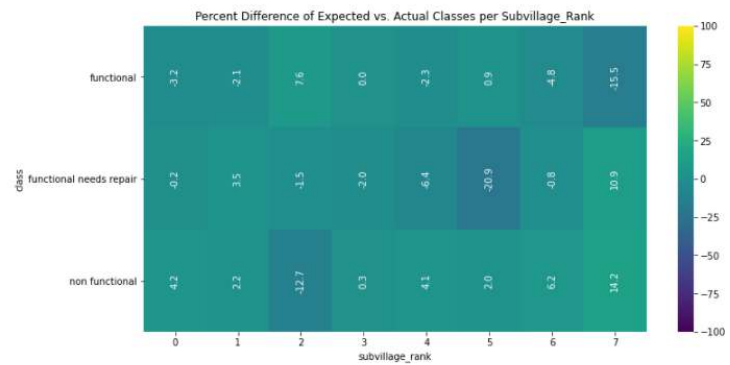
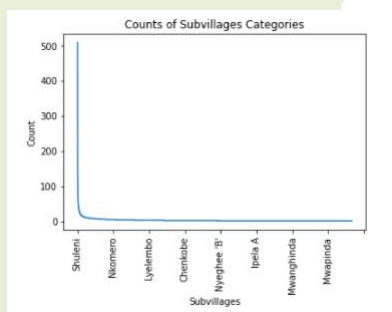
6

Locations of Wellpoints – Urban / Rural



7

Locations of Wellpoints – Subvillage



8

Water & Equipment Information - Source



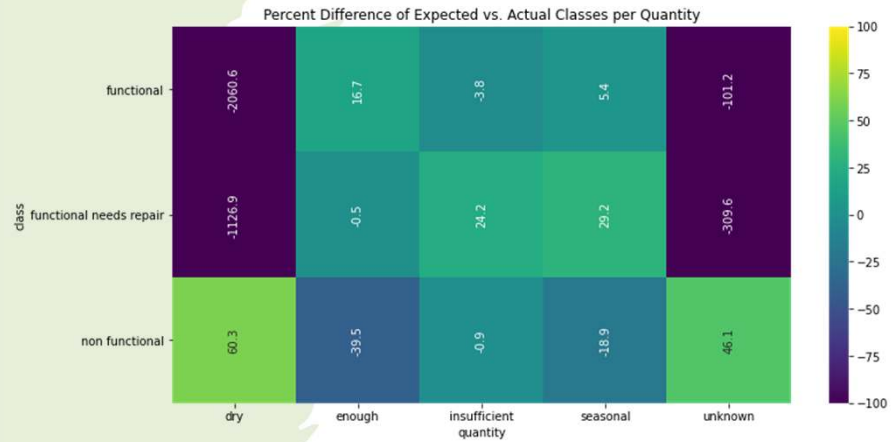
9

Water & Equipment Information – Water Quality



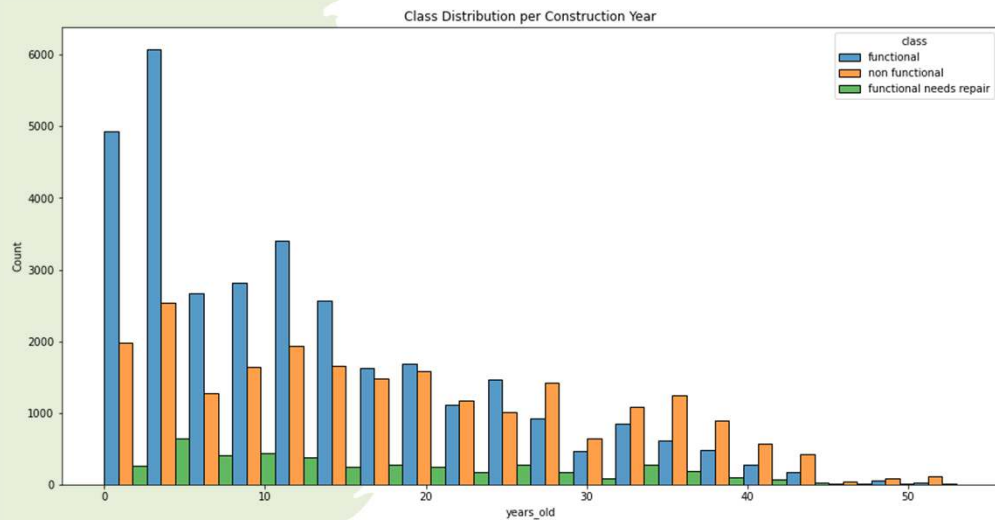
10

Water & Equipment Information – Water Quantity



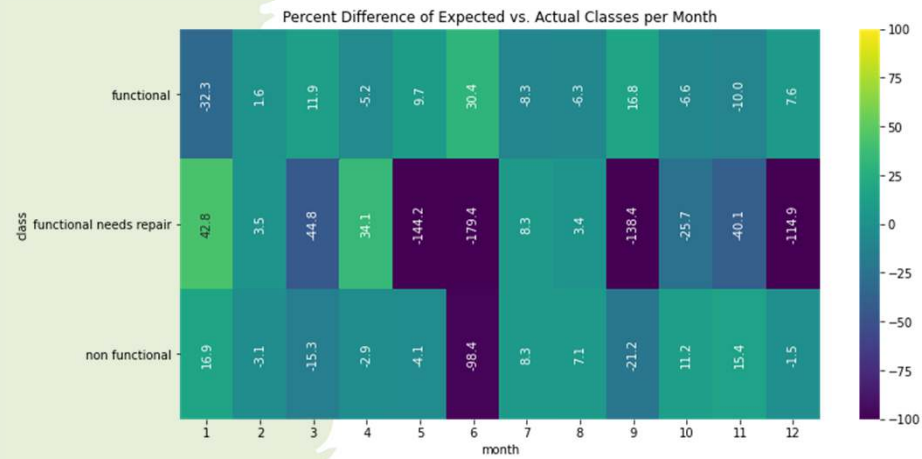
11

Date / Time – Years Old



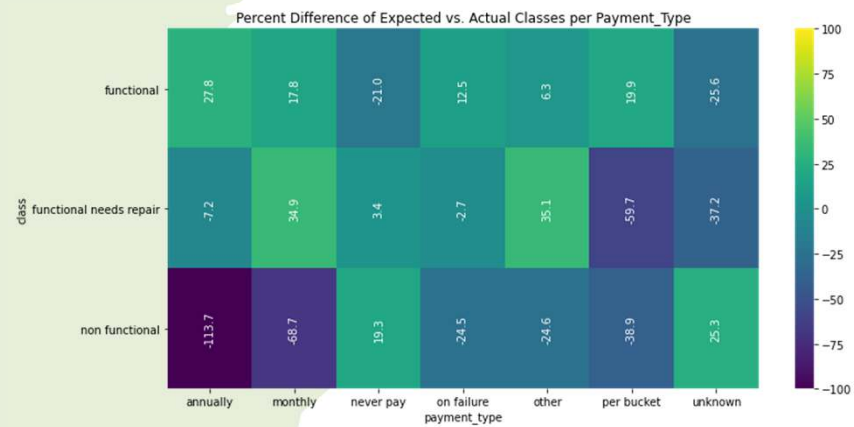
12

Date / Time – Month Recorded



13

Administrative – Payment Type



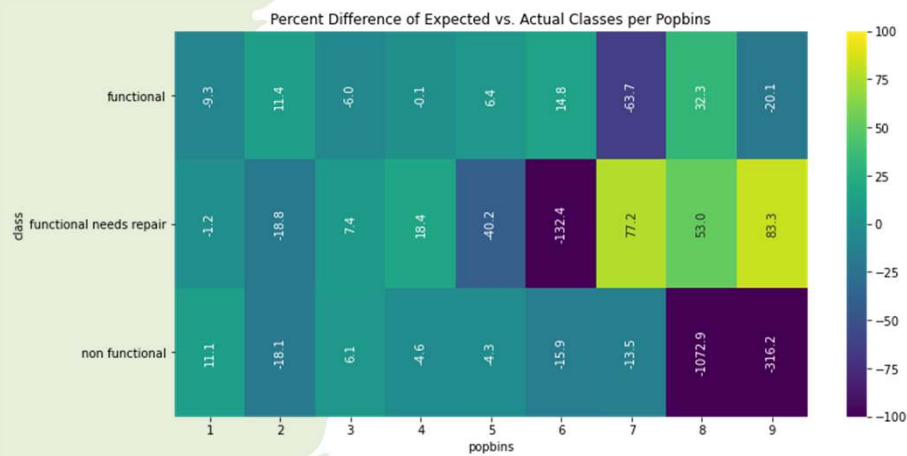
14

Administrative – Installer / Funder



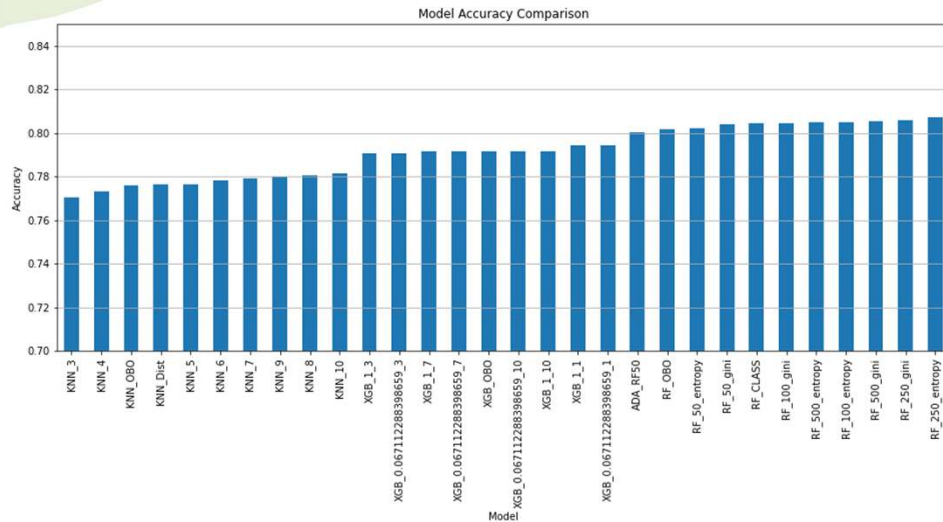
15

Administrative – Population



16

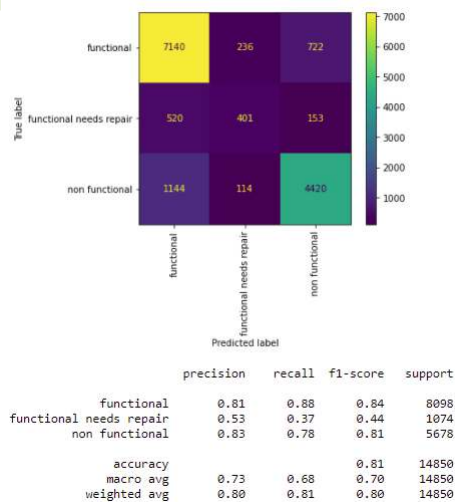
Model Comparisons



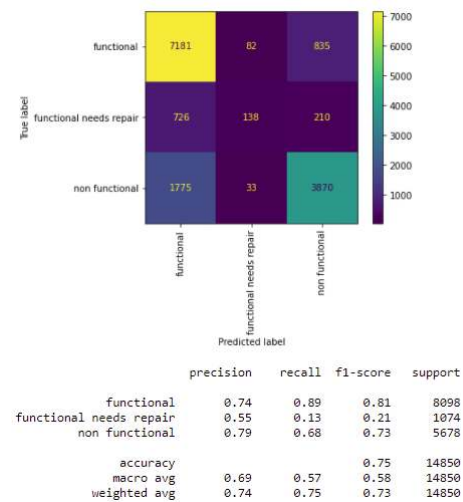
17

Trees vs. Lines

Best Random Forest



Best Logistic Regression



18

Conclusions

- The achievement of over 80% accuracy should allow for the Tanzanian government to use this model as an accurate way to plan ahead for maintenance issues as well as choose which areas to look deeper at for regular up keep and or replacement. The year over year simulation available on Github shows a continued improvement of the model as new information from new years are add in so the tool will only continue to grow in it's ability to predict wellpoint statuses.

