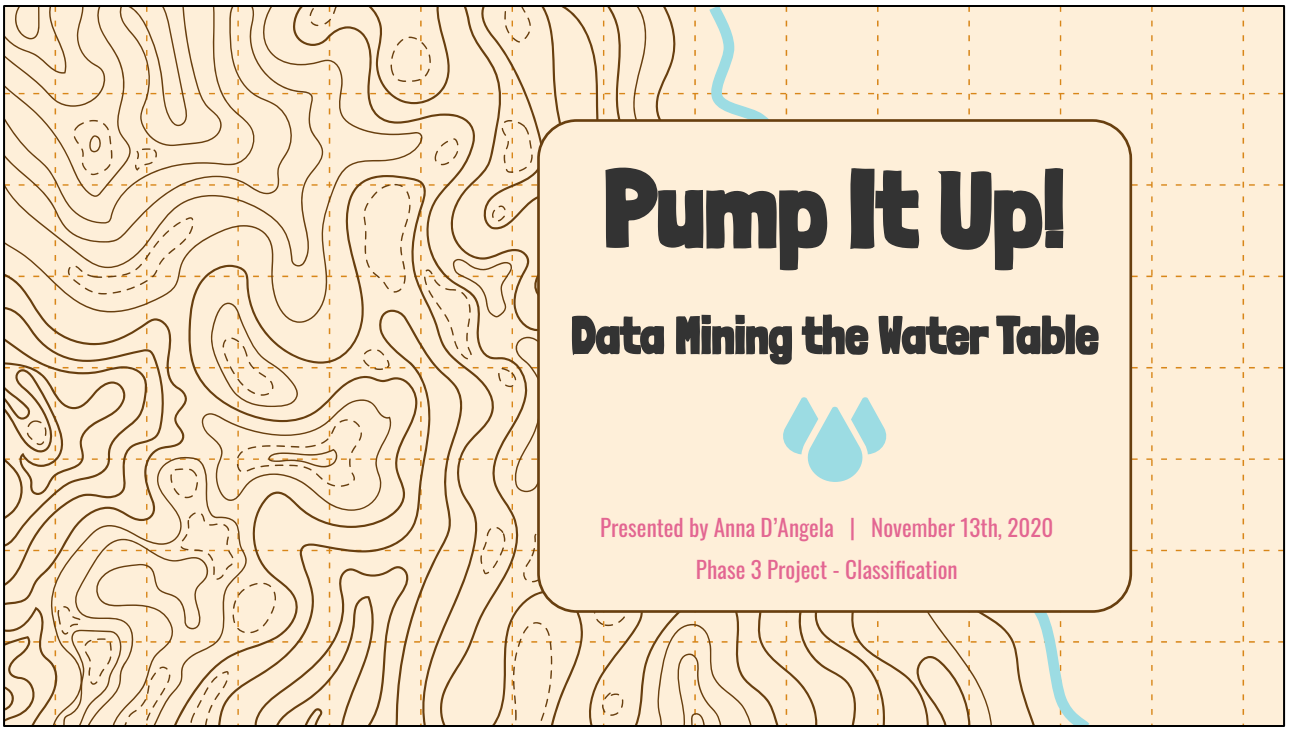

Pump It Up!
Data Mining the Water Table
Presented by Anna D'Angela | November 13th, 2020
Phase 3 Project - Classification

Tanzania Water Crisis

"Tanzania, as a developing country, struggles with providing clean water to its citizens. There are many water points established in the country, but some are in need of repair while others have failed altogether. A smart understanding of which waterpoints will fail can improve maintenance operations and ensure that clean, potable water is available to communities across Tanzania."

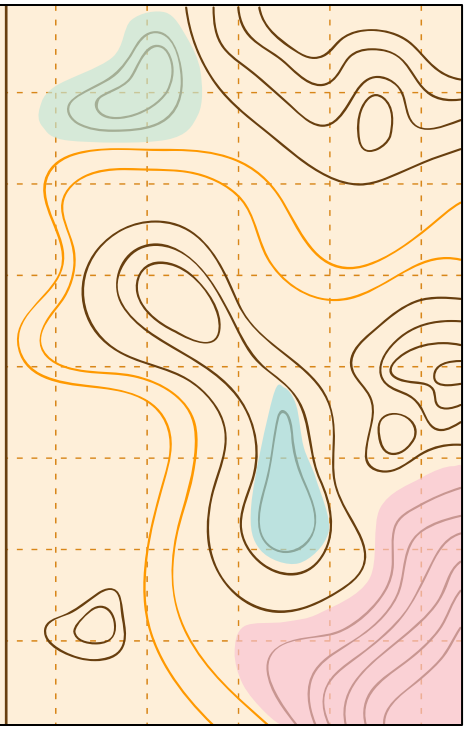
Competition hosted by DRIVENDATA
Project for Flatiron School, Data Science Program

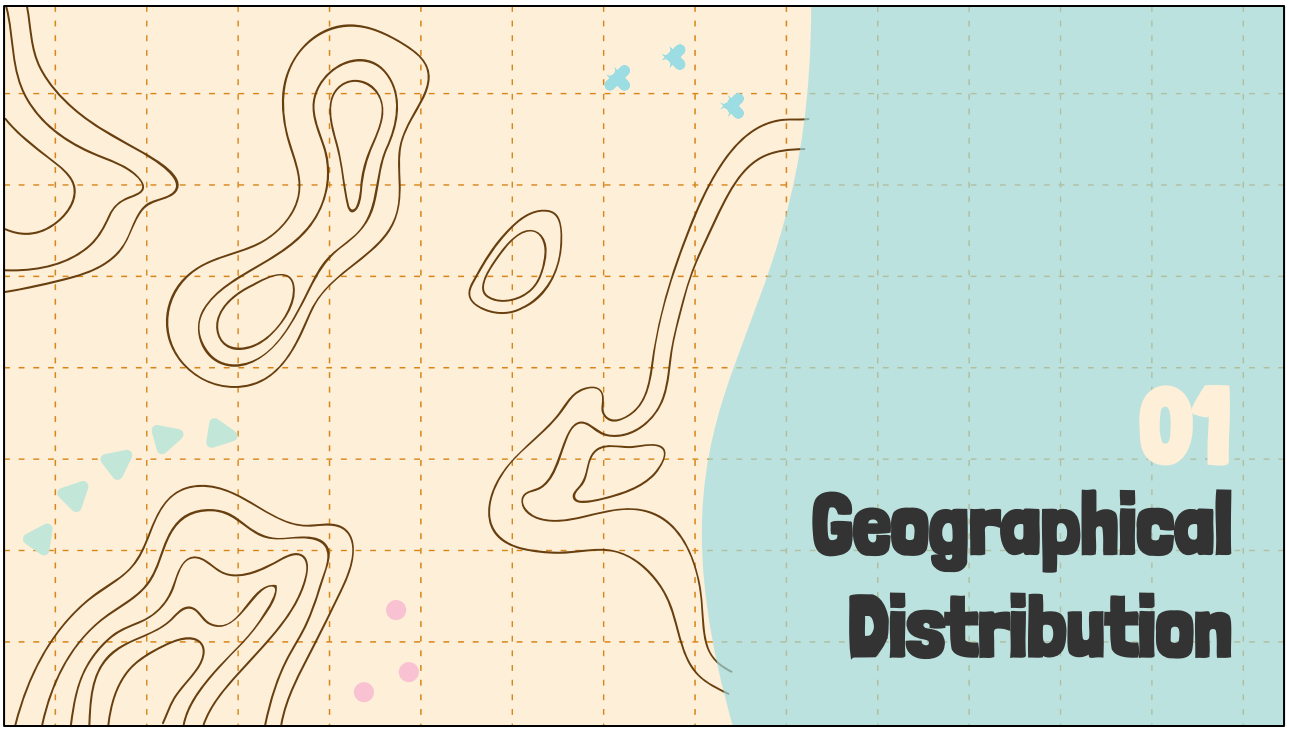
Objective:

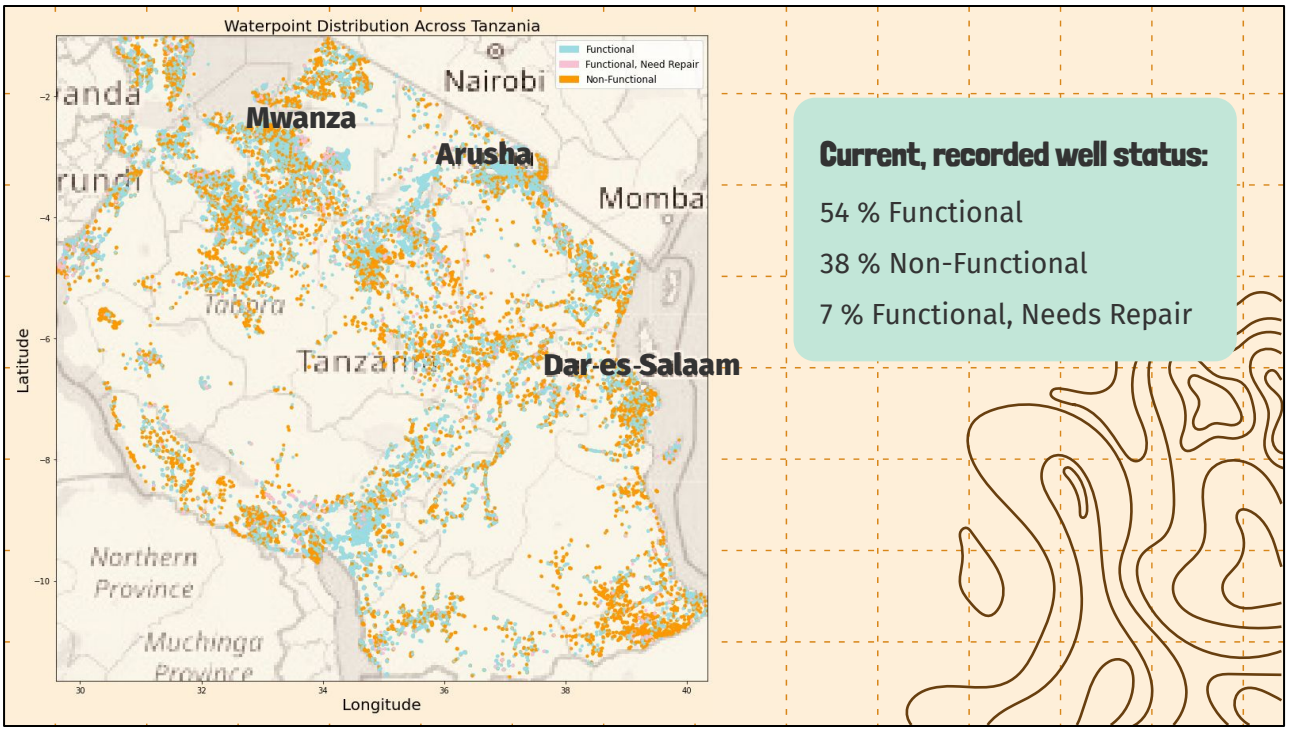
Predict the functionality of a water point

Examine the effect of the following on water point status:

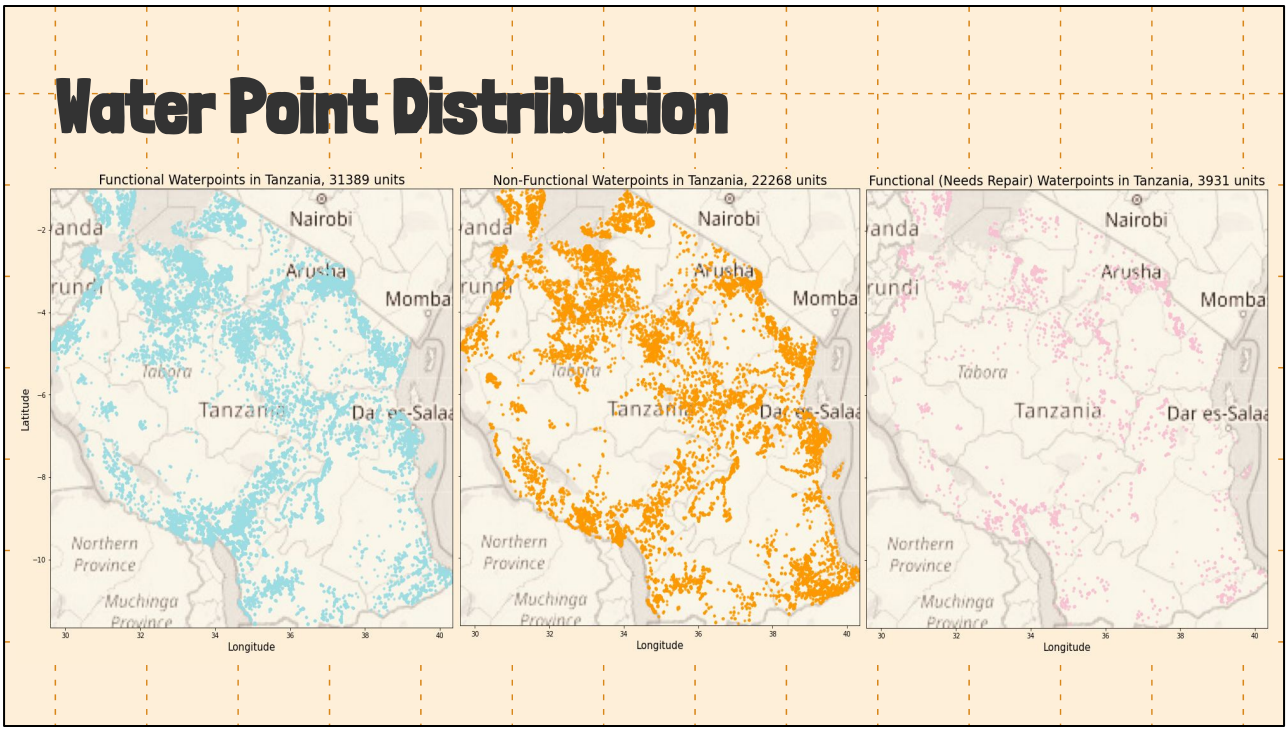
- Geographical Distribution
- Seasonality
- Waterpoint Management
- Payment Type



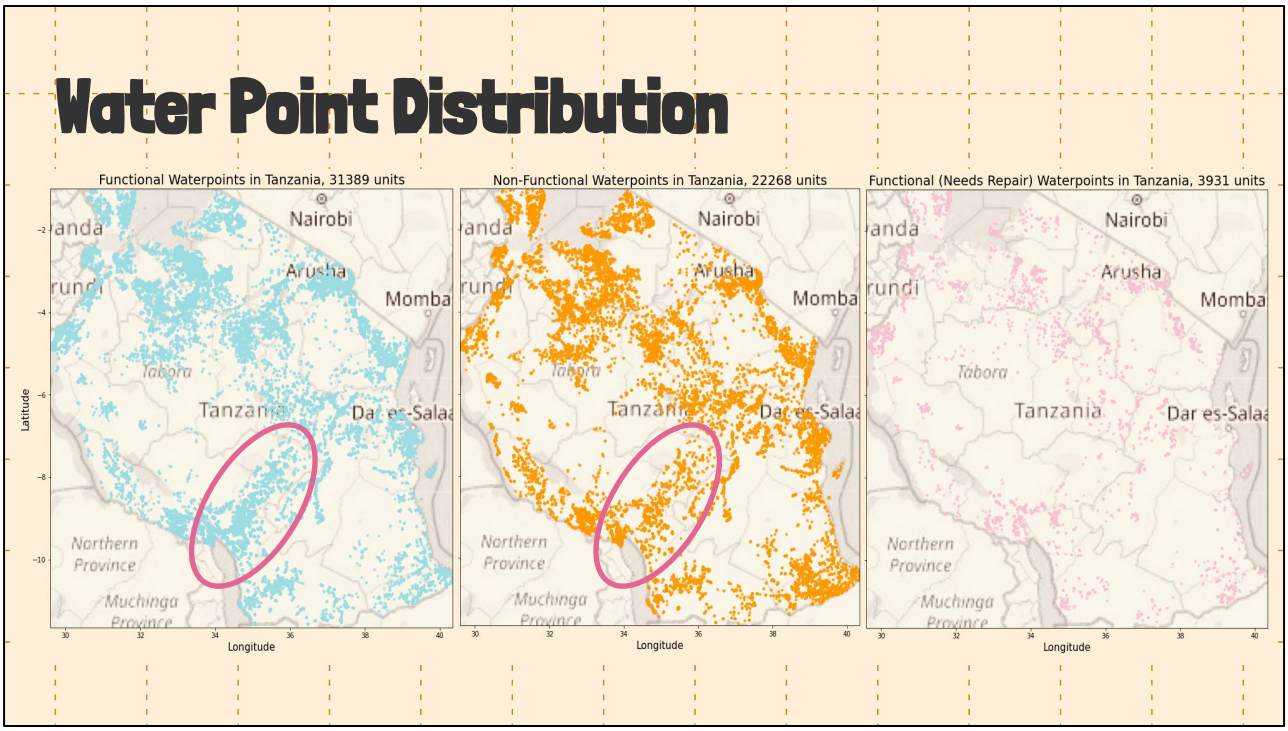

01
Geographical
Distribution



Note: the three most populous cities in Tanzania.



Let’s examine the distribution of these statuses separately. Functional, needs repair is diffuse and rare. Let’s focus on comparing the two main classes: functional and non-functional.

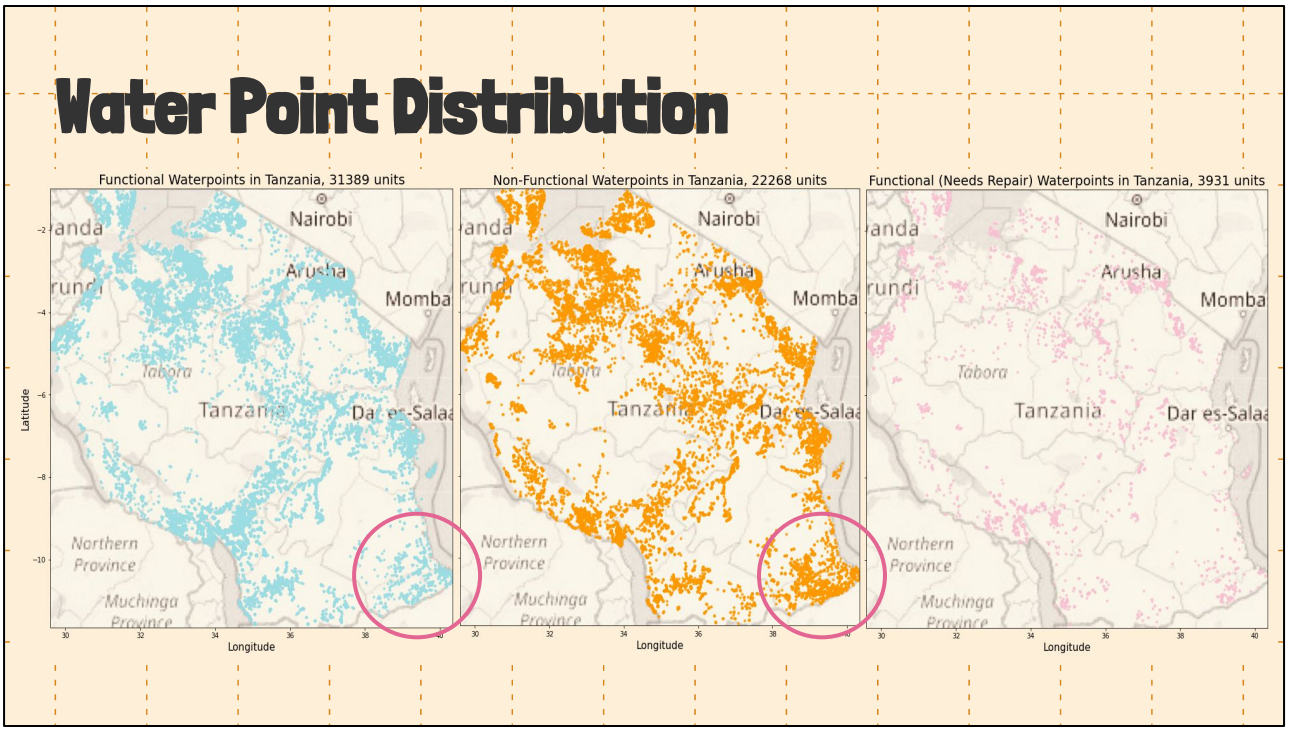


RURAL ACCESS INDEX

"The RAI measures the proportion of people who have access to an all-season road within 2 kilometers (km), considered a reasonable distance for people's normal economic and social purposes."

RAI is calculated based on population, road distributions, and road quality. It is scaled between 0 and 100.

While functional wells are seen as generally more prevalent than not, I have highlighted a cluster that is visually much more dense. This highlight corridor has ranks in highest and second highest level of RAI for Tanzania: a range of 34.2 to 93.0. Indicating above average access to roads. We can see a heavier density of functioning wells as compared to the same region in non-functional and FNR (functional, needs repair).

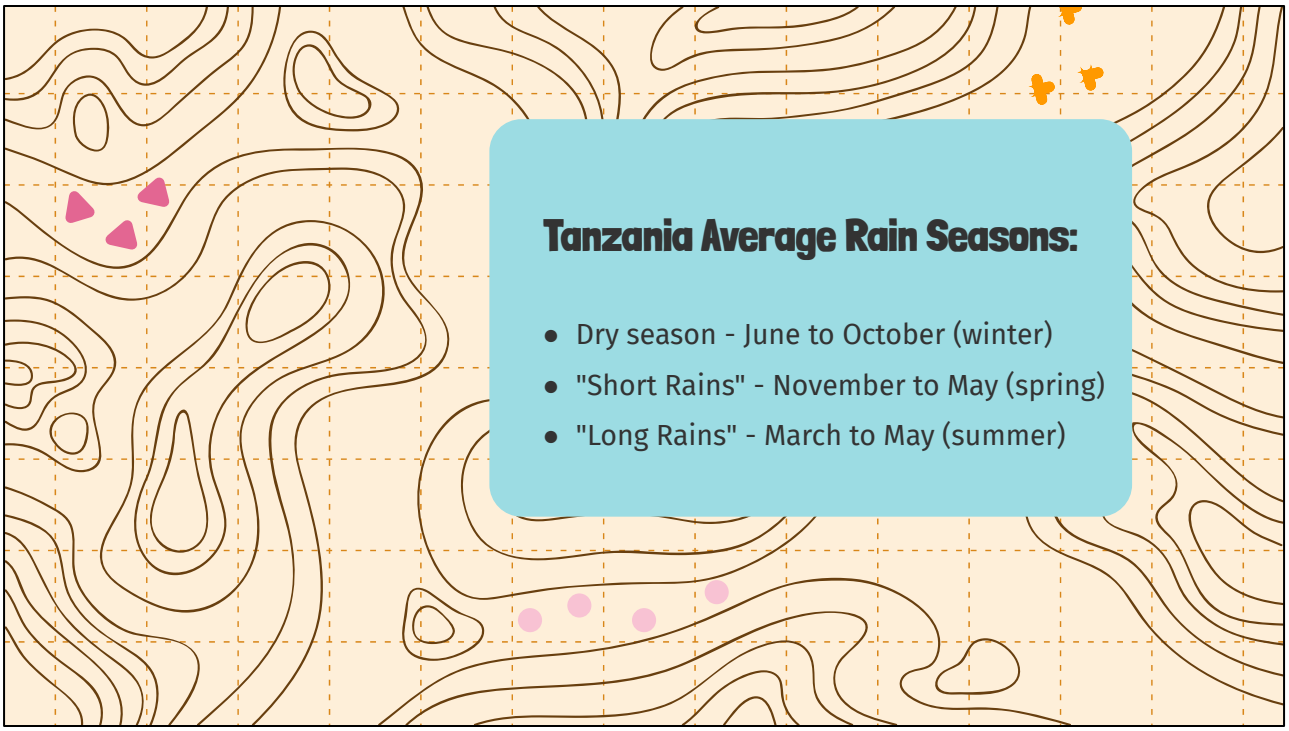


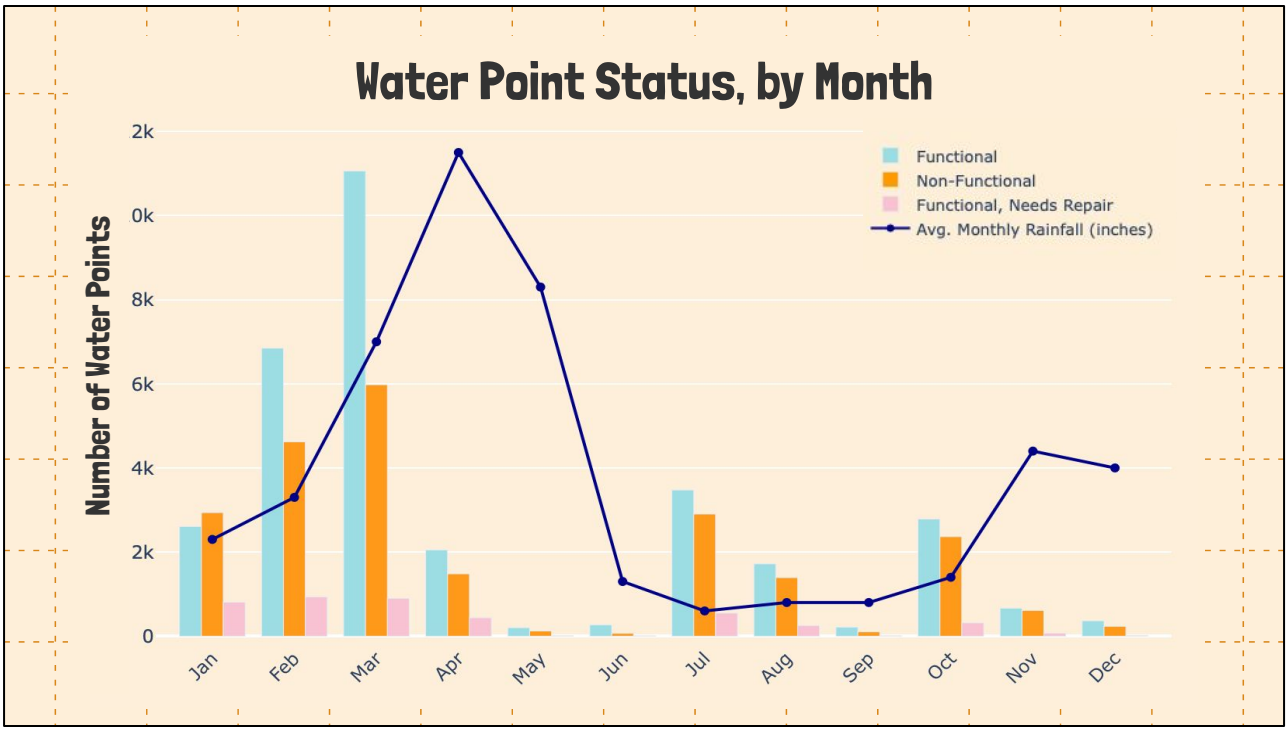
The south east corner of Tanzania however, is at the lowest level of RAI, ranging 0 - 22.6. We can see a more concentrated cluster of non-functional wells in this area. This would suggest that ease of access contributes to a water points functionality.

02

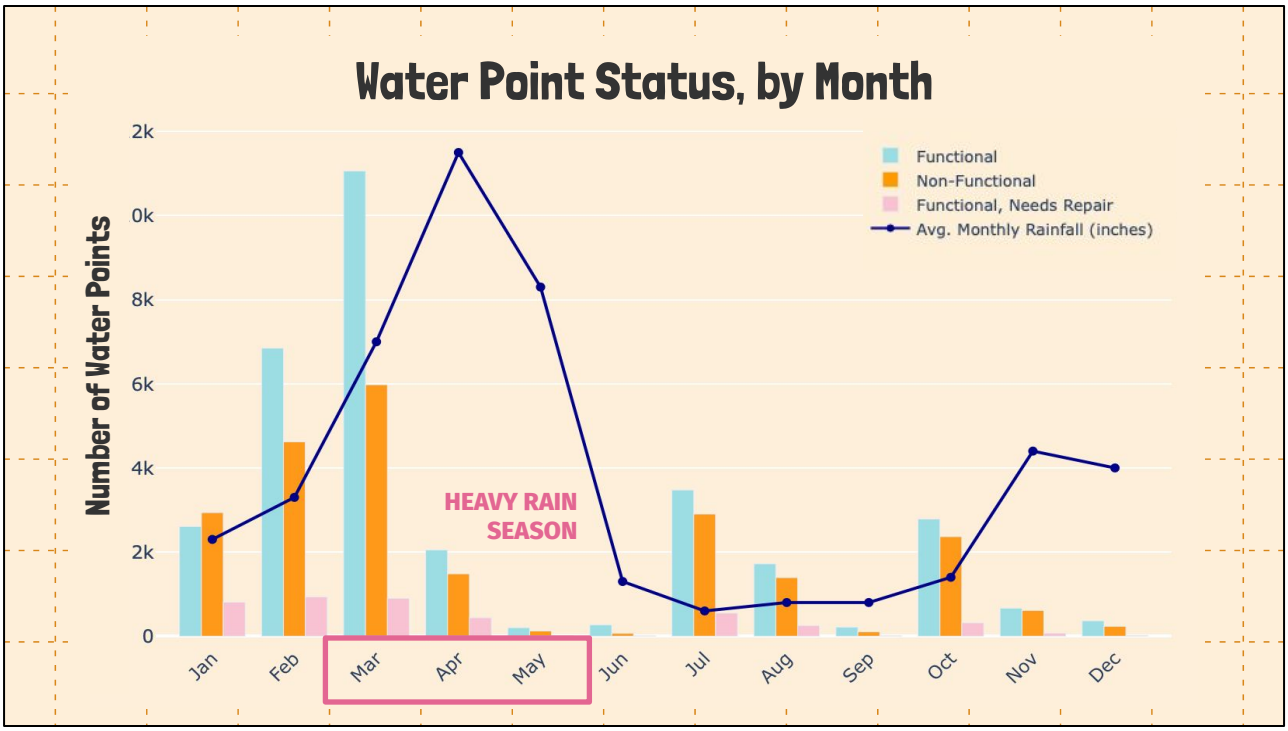
Seasonality

Winter dry season in Tanzania. Good for safaris, bad for hydration.


Tanzania Average Rain Seasons:
Dry season - June to October (winter)
"Short Rains" - November to May (spring)
"Long Rains" - March to May (summer)



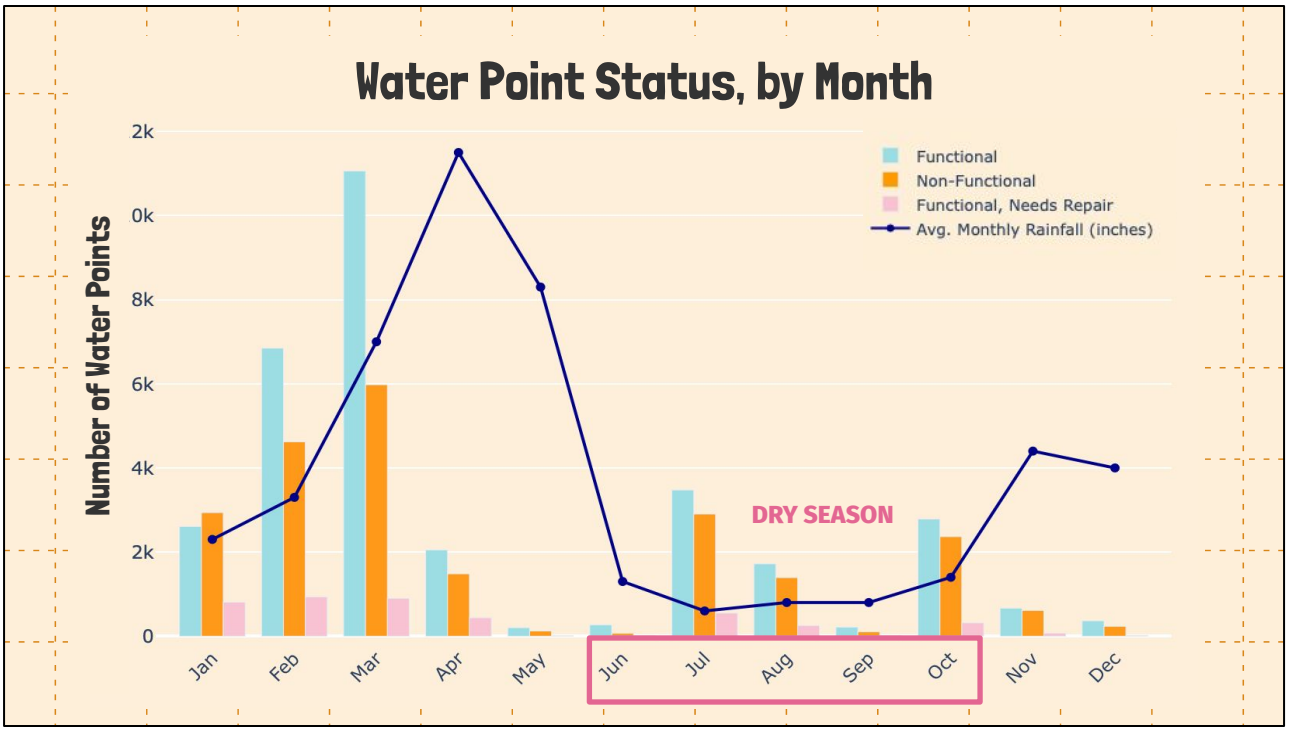
Tanzania average rainfall seasons:
Dry season – June to October (6 - 10)
Wet season – "Short Rains" November to May (11 - 2),
and "Long Rains" March to May (3 - 5)



SUMMER/FALL

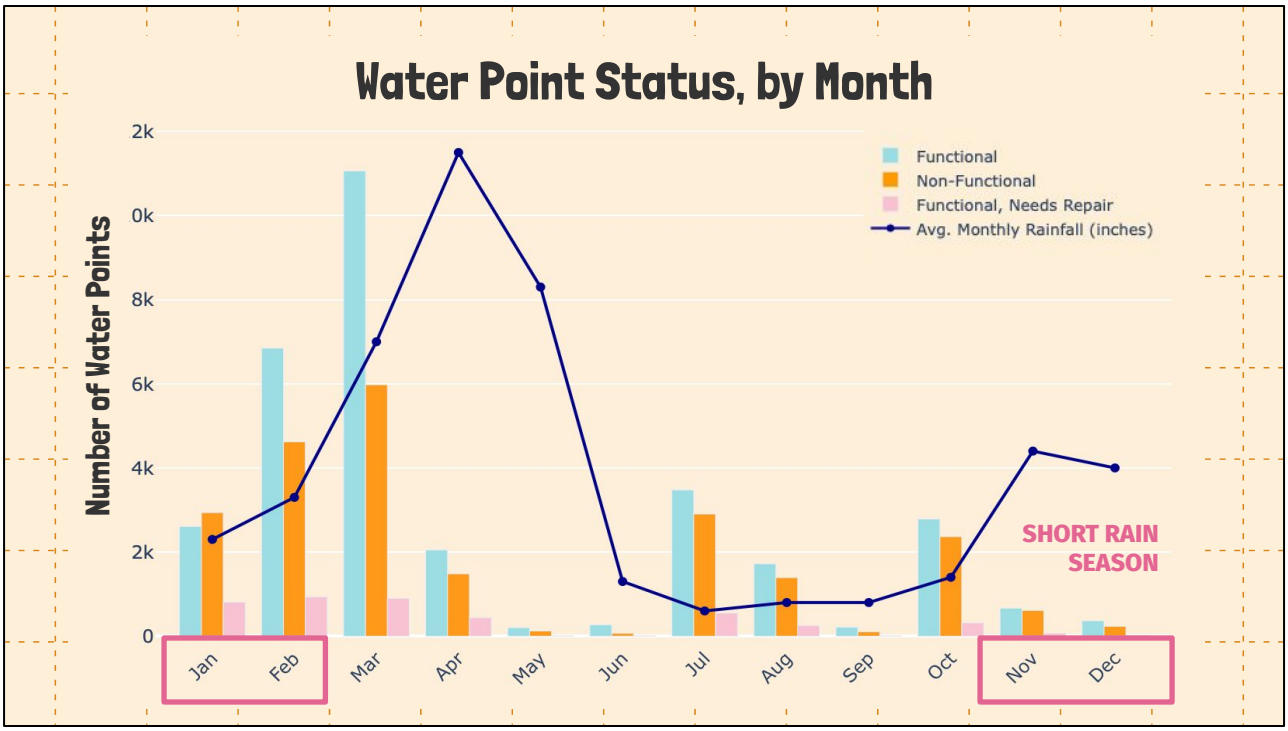
Tanzania average rainfall seasons:
Dry season – June to October (6 - 10)
Wet season – "Short Rains" November to May (11 - 2),
and "Long Rains" March to May (3 - 5)

We see a drop of in number of wells checked due to it being more difficult/costly to travel at this time. Flooding is common. There are many rainwater collection initiatives happening as well.



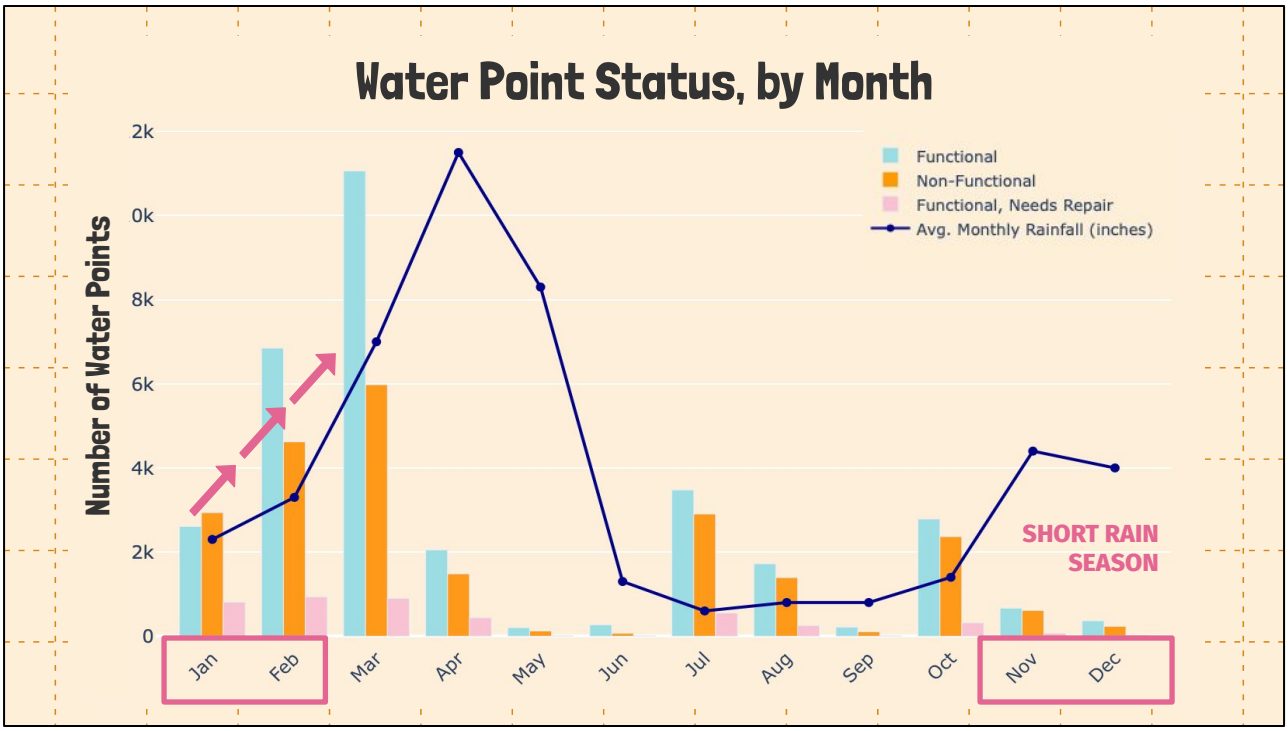
WINTER

Tanzania average rainfall seasons:
Dry season – June to October (6 - 10)
Wet season – "Short Rains" November to May (11 - 2),
and "Long Rains" March to May (3 - 5)

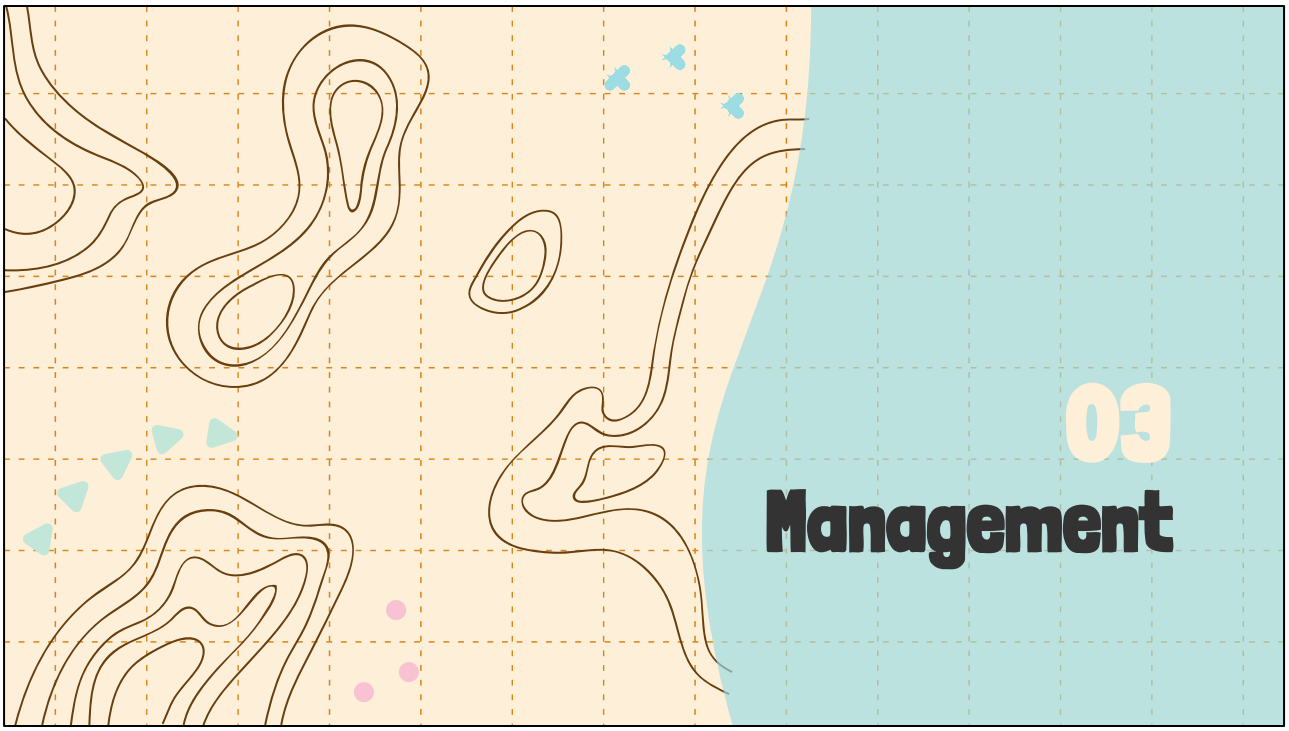


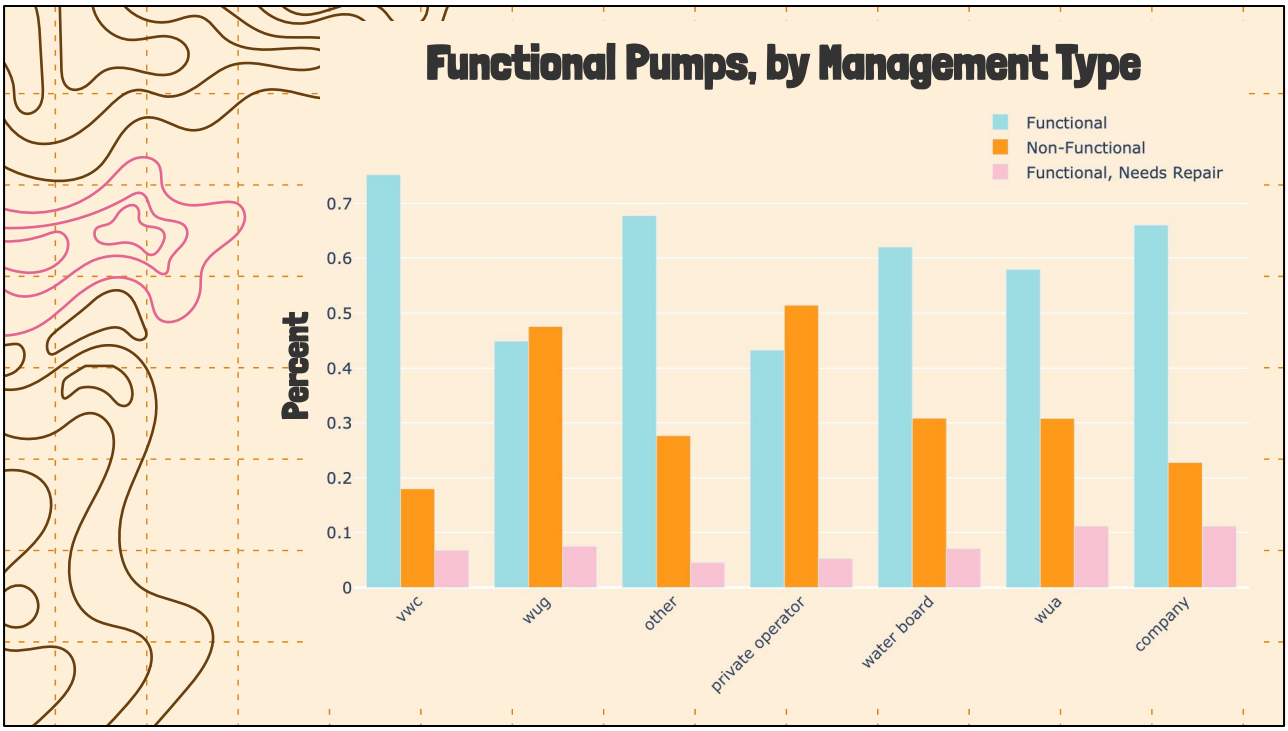
SPRING/SUMMER

Tanzania average rainfall seasons:
Dry season – June to October (6 - 10)
Wet season – "Short Rains" November to May (11 - 2),
and "Long Rains" March to May (3 - 5)



We see the greatest difference in functional/non after 3 months of “short rains” (culminating in February). The largest disparity is observed in March at the start of the heavy rains. Checking drops off here during peak rainy season. Possibilities: collecting rainwater as an alternate source, difficulty traveling during rainy conditions (flooding is common and heavy mud on dirt roads can be impassible). With heavy rains comes sanitation issues as rainwater can carry waste.


03
Management



Village Water Committee – Initiated as a formalised water point management structure in the 1991 Water Policy, these bodies continue to exist in all villages visited in Dodoma (although not in Singida) as a regulatory body with elected membership. They are effectively the default management framework across the three regions. The VWC is a cosignatory with the DWE to the village water fund.

Private Operator – An individual or group that is contracted by the VWC or other official body to undertake operation and some maintenance of the extraction system or single DP for an arranged fee. Tenderer is sometimes used as an alternative name; Wakala is the official term in Dodoma Rural and some other areas; Mpwapwa more commonly uses Mbia, or ‘caretakers’.

Water User Association – A legal entity autonomous from village government that oversees the management of an entire extraction system. The WUA has its own fund which functions as the village water fund. Officially, it will have authority over the Water User Groups (WUGs) using the water source.

Water User Group – These are the sub-village level management groups that are responsible for operation and maintenance of specific DPs within a WUA. In practice, sub-villages do not always create WUGs so management will revert to the WUA. Has its own bank account to fund small repairs and regular maintenance.

School – Refers to extensions of systems to schools. The water points are public but managed fully by the school and are often obligated to pay a contribution to the

Private – A privately funded extension to a water system. The owner may sell on the water and often must allow public access to any DP on the private system.

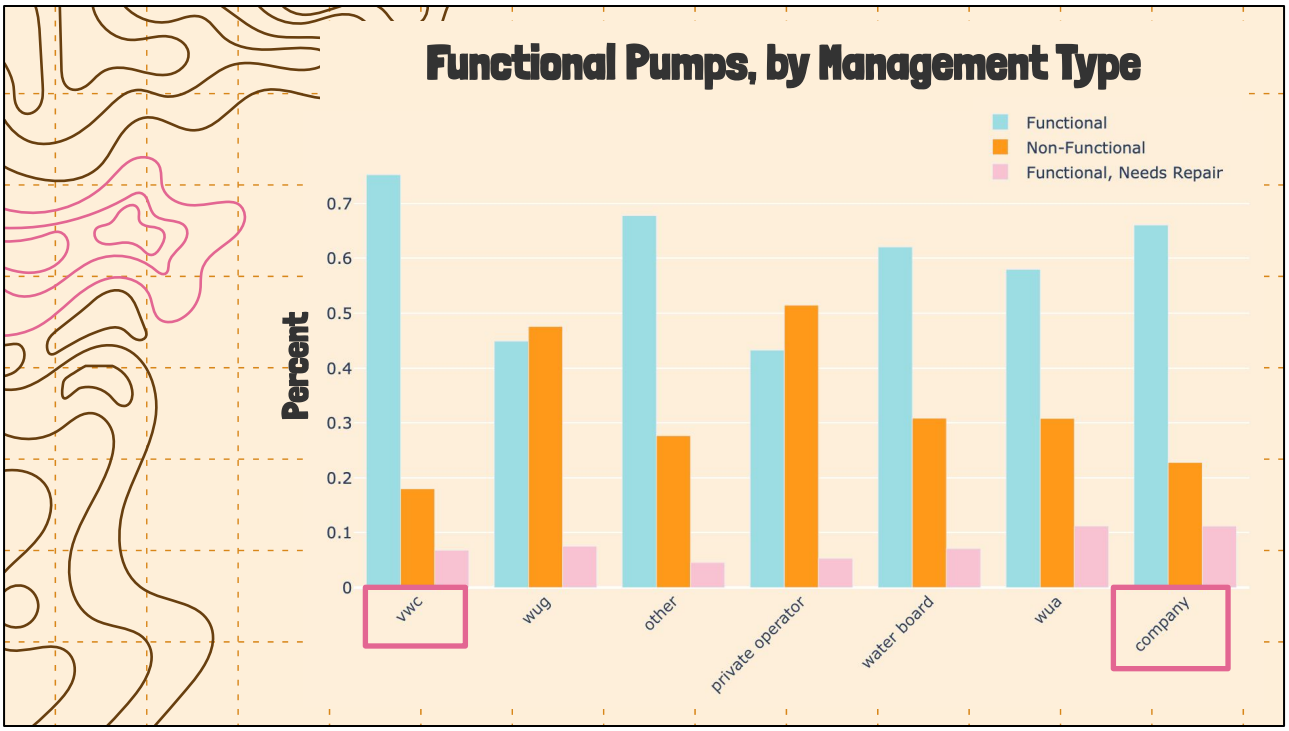
Company – A legal entity sometimes under contract.

Board/Trust – Also a legal entity, autonomous from village government that takes the place of VWCs.

Government – Only very few DPs of this type exist and they are almost exclusively in Iramba district, Singida. While there is no first hand evidence to explain this system it most probably describes a take-over by village government in the absence of a VWC or other management type.

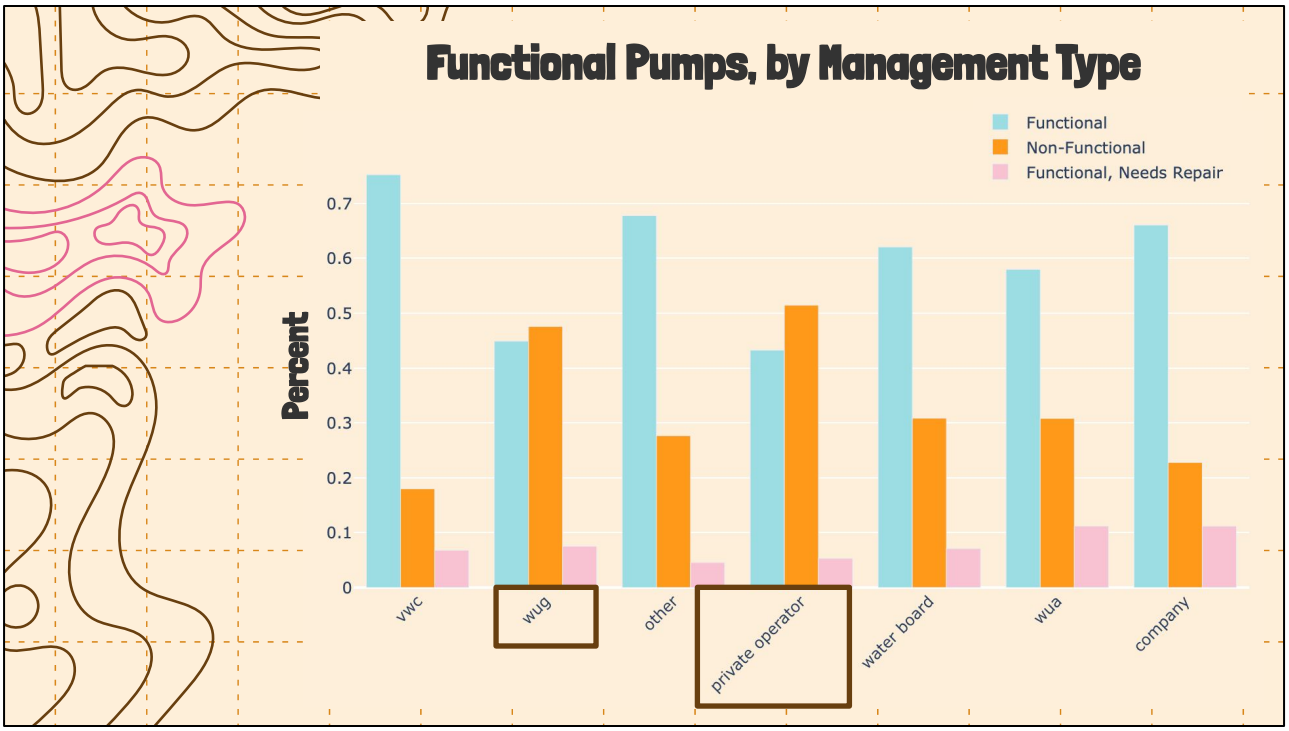
SOURCE: PRIVATE OPERATION IN THE RURAL WATER SUPPLY IN CENTRAL TANZANIA:

QUICK FIXES AND SLOW TRANSITIONS, Sam Moon August 2006 via WaterAid Tanzania



BEST:
Village Water Committee – Initiated as a formalised water point management structure in the 1991 Water Policy, these bodies continue to exist in all villages visited in Dodoma (although not in Singida) as a regulatory body with elected membership. They are effectively the default management framework across the three regions. The VWC is a cosignatory with the DWE to the village water fund.

Company – A legal entity sometimes under contract.

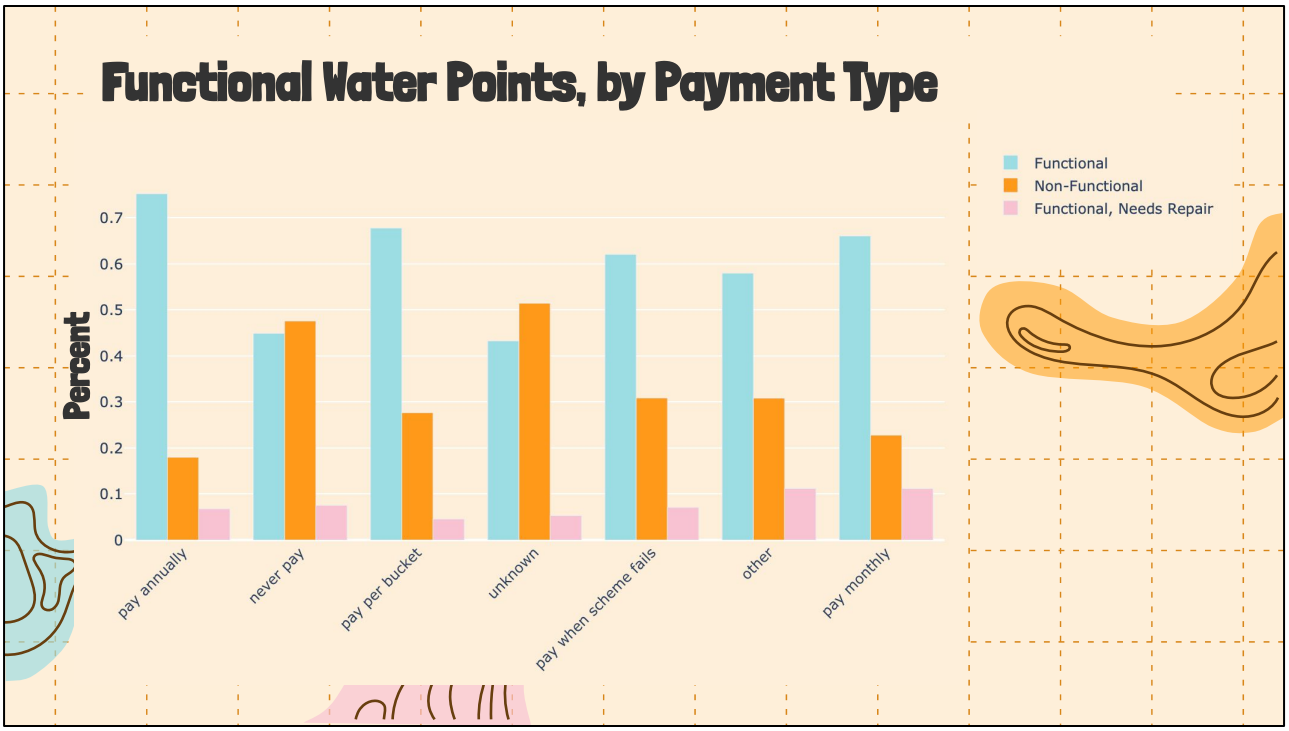


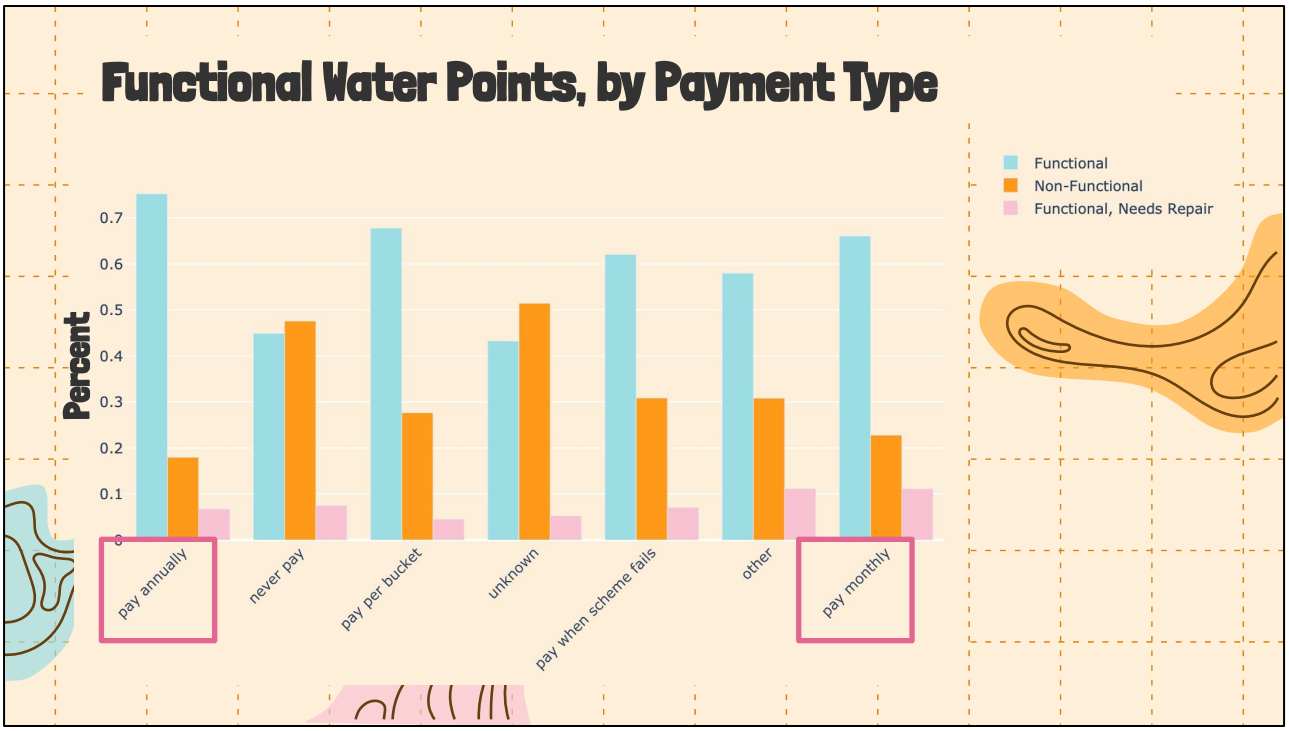
WORST:

Private operator and WUG (Water User Group) have a near even split in functional and non. The VWC (Village Water Committee) and Company (legal entity, under contract) and other lead with ratio of functional wells. Relatively even percent of wells need repairs.

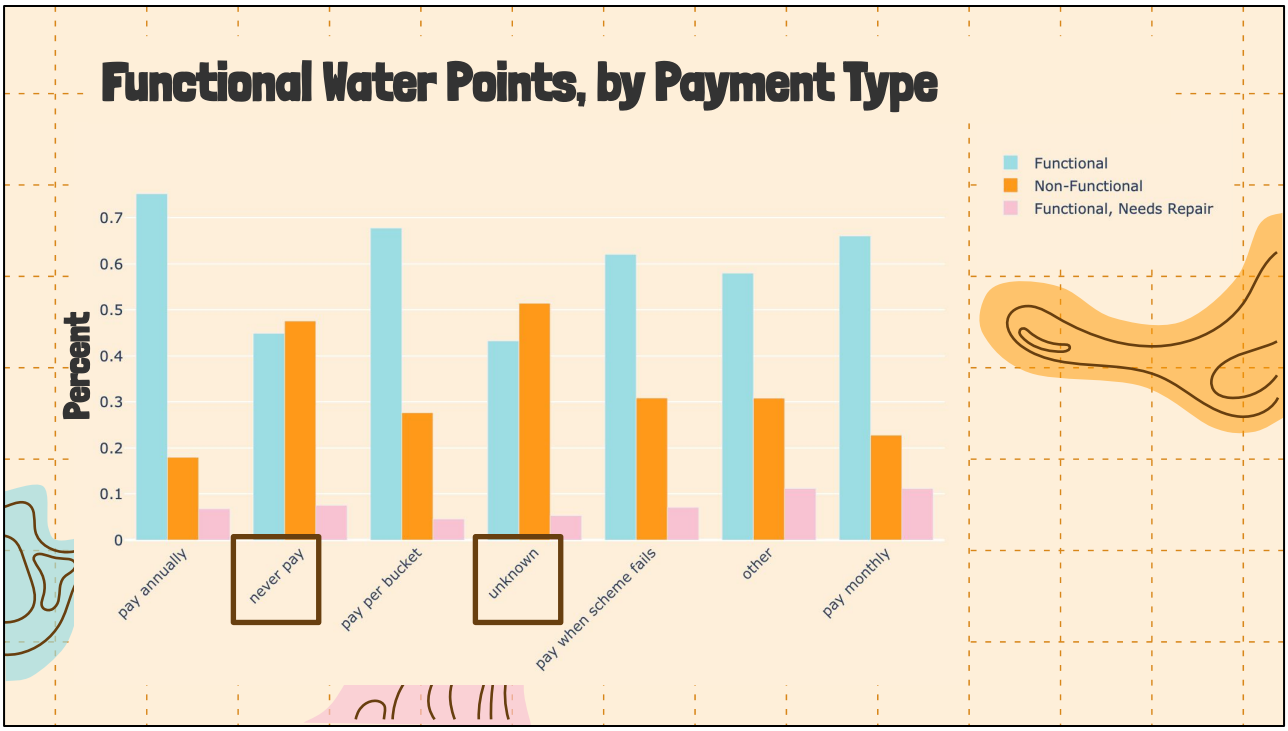
04

Payment Type


Functional Water Points, by Payment Type
Functional
Non-Functional
Functional, Needs Repair
Percent
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
pay annually
never pay
pay per bucket
unknown
pay when scheme fails
other
pay monthly



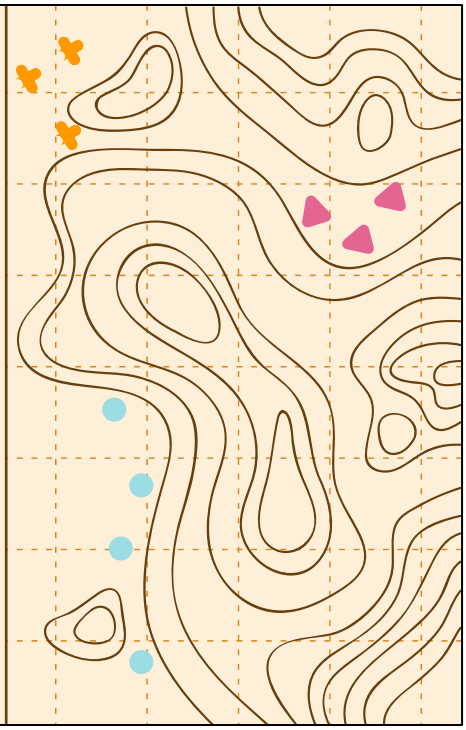
BEST: payment of any kind or “other” payment arrangement (which includes charity schemes such as tokening) all have almost double function well rate to non-functioning. The fees likely allow them to perform regular maintenance, and keeping their wells functioning is required to keep customers.



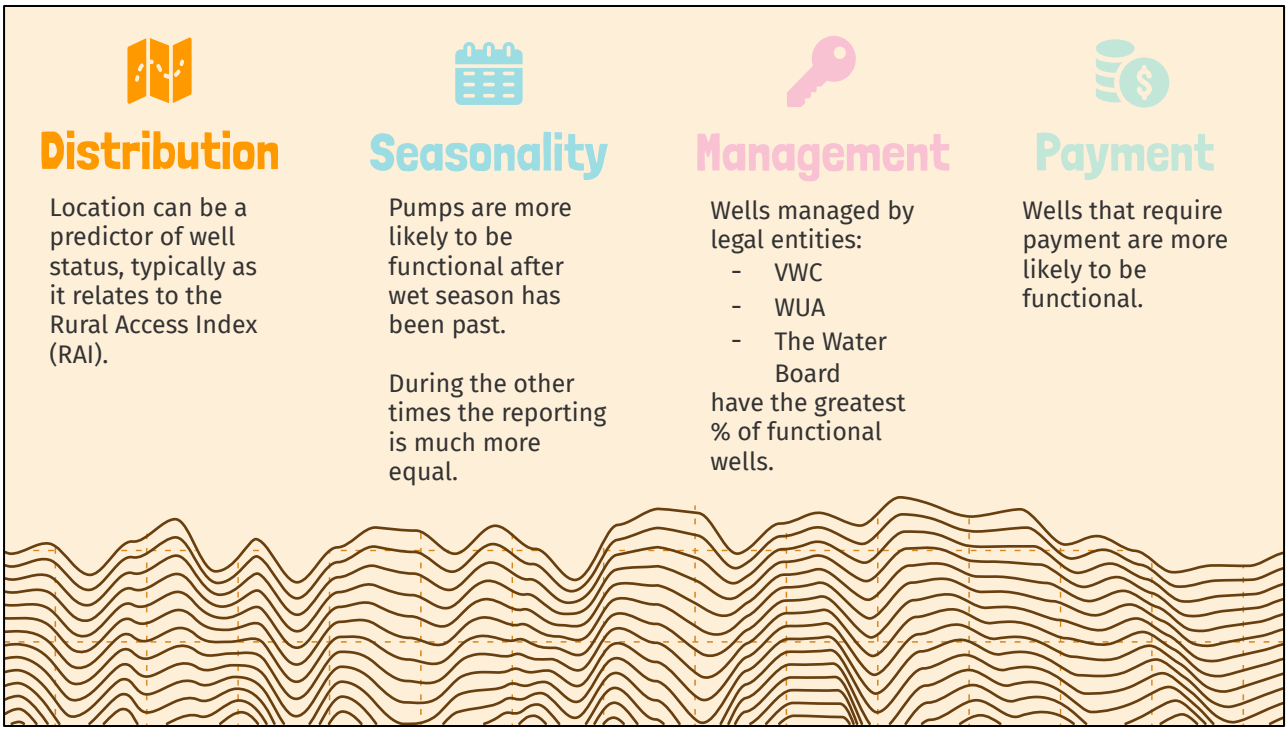
WORST: Unfortunately (but expectedly) the paid and ‘unknown’ (which likely has no pay scheme, if it is not reported) have more non-functional wells than functioning. Could be due to no resources for maintenance or repairs, or be too rural.

Water Point Classifier

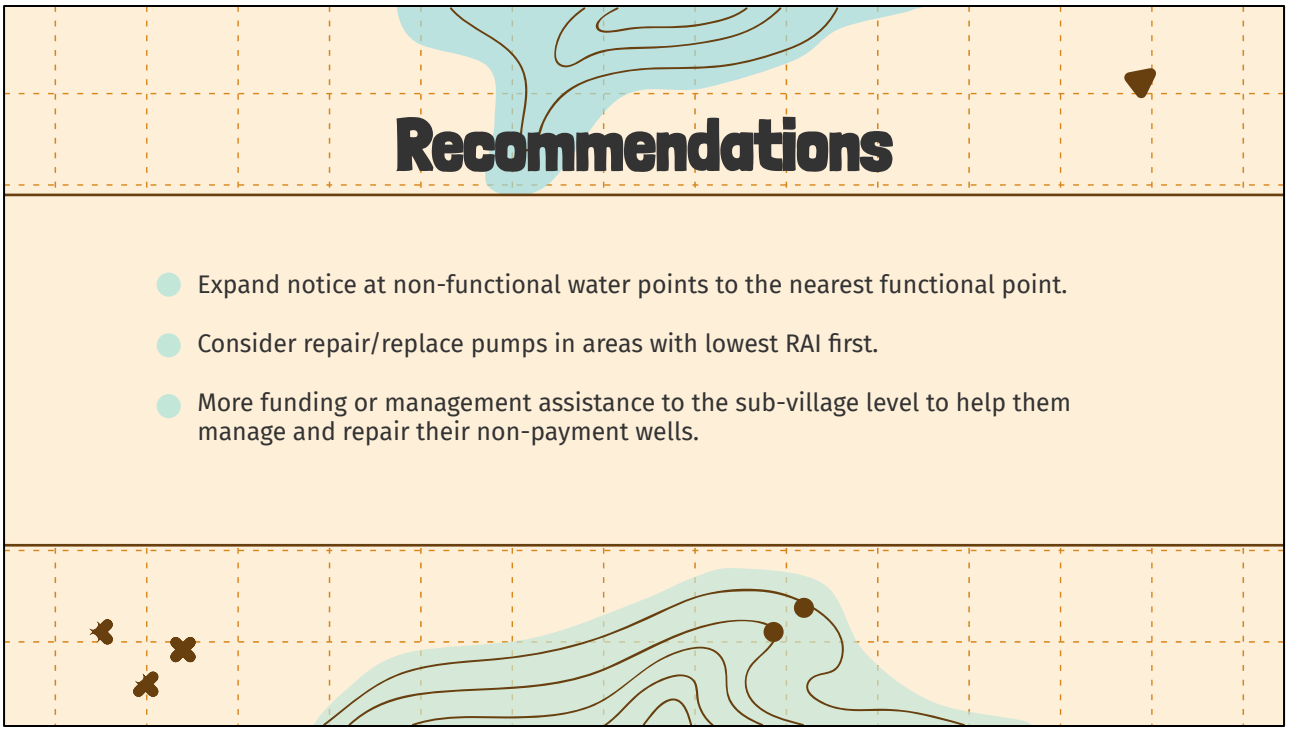
- Model predicts water point status with **77.9% accuracy**
- Model is well fit to unseen data
- Important features:
  - 'quantity' : 'dry'
  - 'water_point_type' : 'other'
  - 'extraction_type' : 'other'
  - 'management_type'
  - 'region'




Summary of
Findings



Summary of findings/Conclusions



Since the distribution of funct. And non-funct. Water points were somewhat evenly dispersed, before pumps can be repaired: expand notice at non-functional water points to the nearest functional point.

There are certain management schemes that are failing, more oversight or assistance to help recover these schemes, or move on to proven schemes.

Future Work

- Research quality and quantity of water per water point
- Investigate a further breakdown of failing management entities
- Explore installer/funder effect on status

Thank you!

- Slide deck from SlidesGo
- Competition hosted by DRIVENDATA
- Find my full analysis on GitHub: @anna-dang

Anna D'Angela | Detroit, MI

Appendix: Model Performance

Model:

```
XGBClassifier(learning_rate=0.2, max_depth=6, min_child_weight=2,
              objective='multi:softprob', subsample=0.7)
```

Train Accuracy: `"Cross-Val Score (k=5):'0.7762794612794612"`

Test Accuracy: Your score for this submission is: 0.7797