

Tanzanian Waterpoints

Purpose

Goal: Use machine learning to help optimize time and resources for the waterpoints that need the most attention.

More specifically:

- Monitoring: What waterpoint features should be most monitored?
- Visiting Priorities: What waterpoints should be visited first?

Methods

1. Analyze waterpoints whose status we know.
 - What's the relationship between the various features and status group?
2. Build a model that can to categorize a waterpoint when we don't know the status
 - Three different algorithms were tested to build an optimal machine learning model
 - Models were optimized against false negatives for non functioning and needing repair waterpoints. This means making sure not to miss waterpoints that are in need.

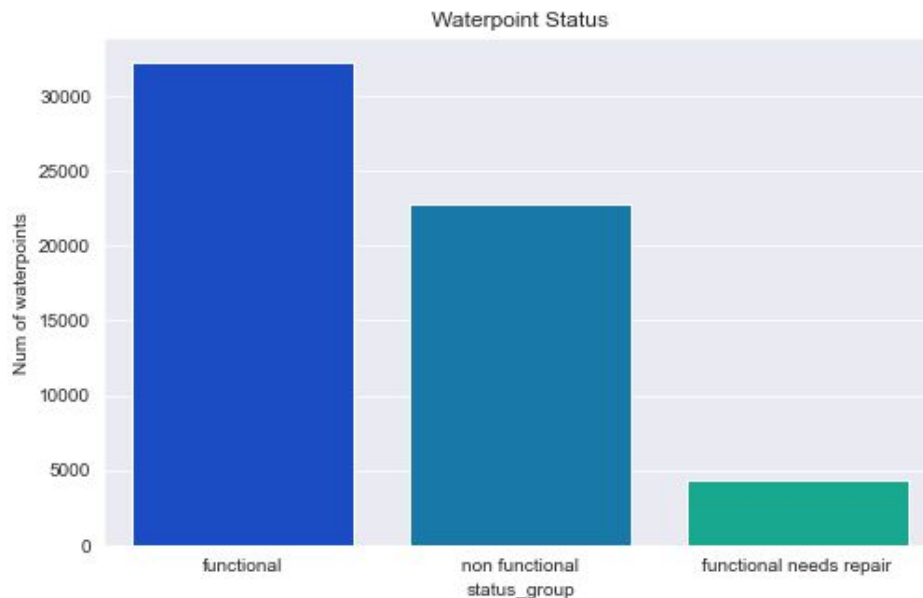
Data

Totals

- 40,000 waterpoints analyzed
- 14,850 waterpoints with unknown status

Each waterpoint

- 39 features per waterpoint
- 3 status groups / waterpoint



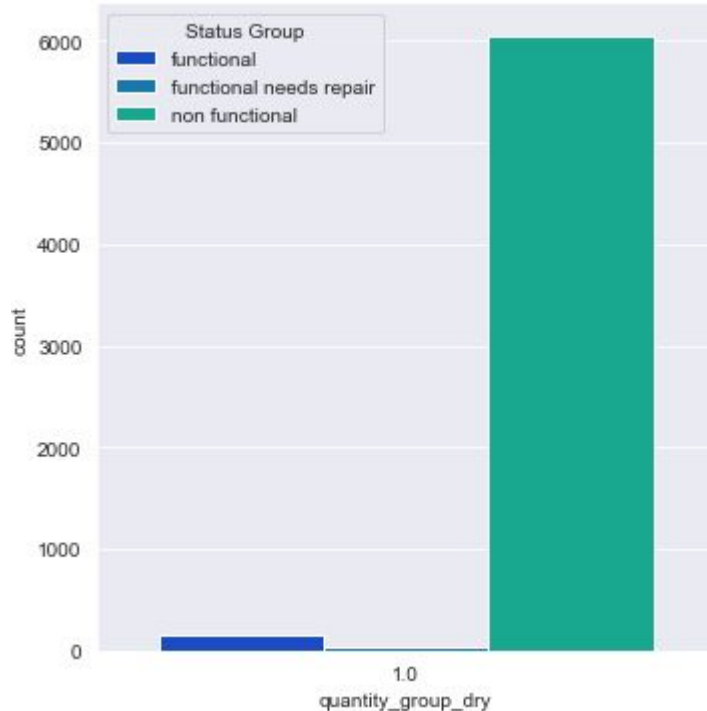
45% of waterpoints were in need of attention

- About 7% of waterpoints in the data were in need of repairs.
- 38% were non-functional.

Most important features

1. Quantity of Water
2. Installer
3. Construction Year
4. Region

Dry Waterpoints / Status Group



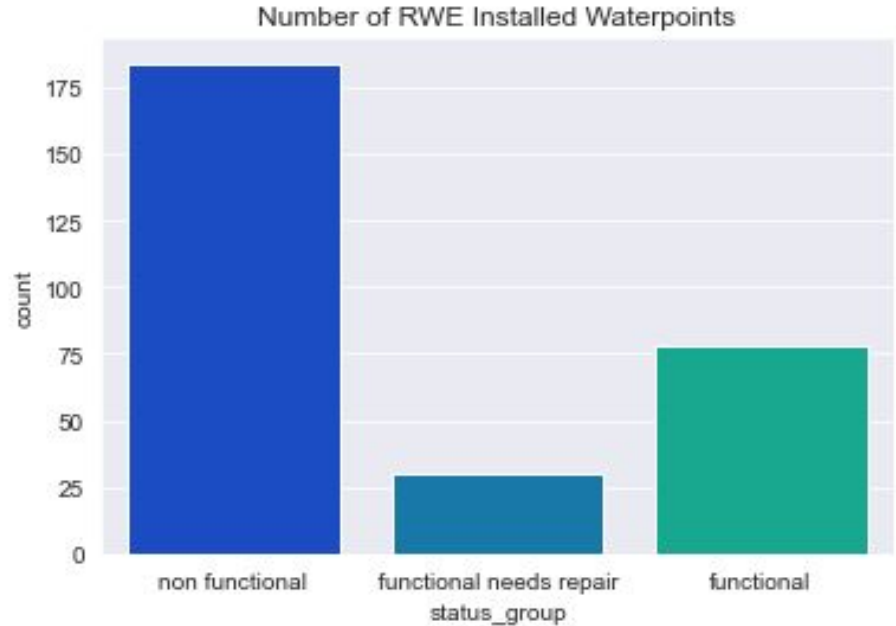
It is predicted that 97% of dry waterpoints are non-functional.

Recommendation: Monitor the quantity of water going forward. Prioritize visits to waterpoints that are considered dry first.

Installer

Waterpoints installed by these parties are predicted to have the highest incidence of issues:

- DWE
- Government
- RWE
- KKT
- DANIDA

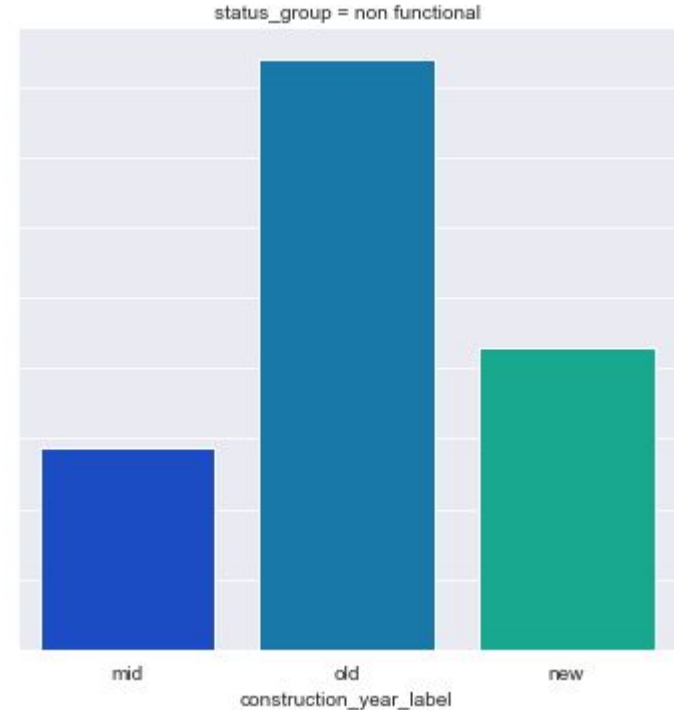


Recommendation: Visit waterpoints installed by these parties as a priority, especially those installed by “RWE” or “Government” as there seem to be both a high number and high proportion of in need waterpoints from these installers.

Construction Year

53% of non-functional waterpoints are predicted to be old (built during or before 1994).

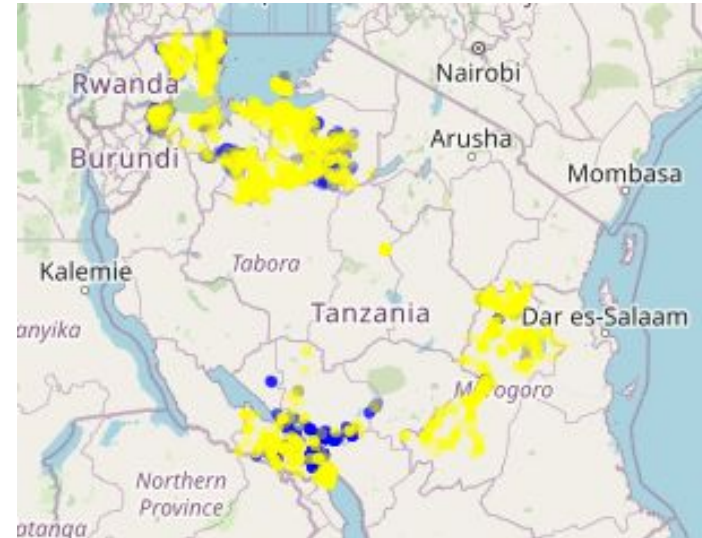
Recommendation: Prioritize visits to older waterpoints, and continue monitoring those built before 1994 going forward.



Region

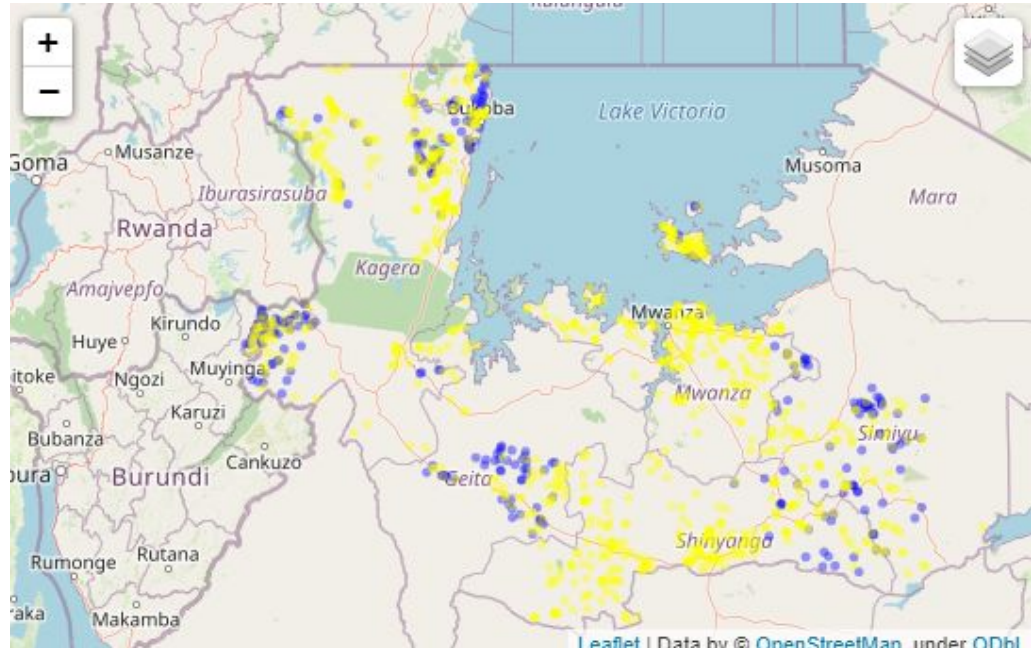
These regions are predicted to have the highest concentrations of non functional or needing repair waterpoints:

- Shinyanga
- Morogoro
- Mwanza
- Kagera
- Mbeya



Region

Recommendation: Monitor waterpoints in those regions, but visit the regions around Lake Victoria first: Shinyanga, Mwanza, and Kagera.



Recommendations Recap

Monitor these features of waterpoints going forward:

- Quantity of water
- Installer
- Construction year
- Region

Visit the waterpoints in Shinyanga, Mwanza, and Kagera around Lake Victoria first. If further criteria is needed, look for waterpoints considered dry, and installed by RWE or Government.

Next Steps

- Why do waterpoints around Lake Victoria seem to have more issues?
- What makes waterpoints installed by RWE more prone to being non functional or needing repair?

Next Steps

- Set up alerts for waterpoints qualify as 'dry.'
- For data scientists: more algorithms. Only 3 were used in this project to build the machine learning model, others may result in more accurate predictions.