Water Pump Analysis

Business Problem

- The Tanzania Ministry of Water needs to predict which water pumps are functional, which need some repairs, and which don't work at all.

- An understanding of which factors cause a waterpoint to fail can improve maintenance operations and ensure that clean, potable water is available to Tanzanian communities.

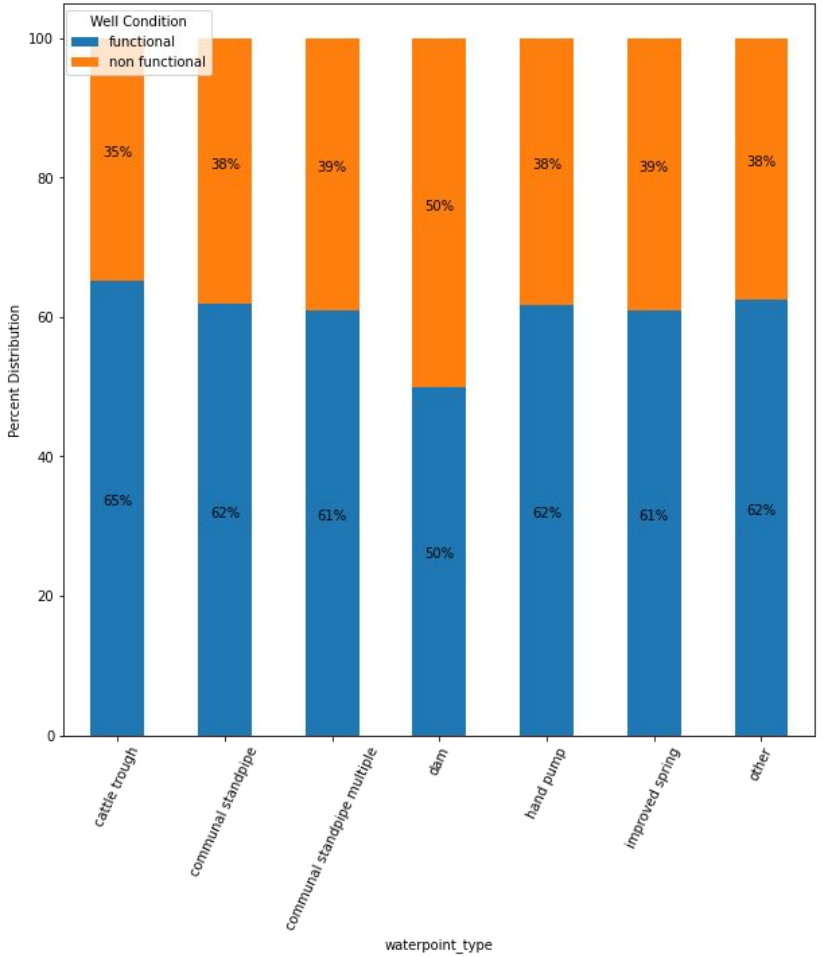
Analysis: What did we examine?

When examining what causes a waterpoint to fail we looked at:

- What kind of pump is operating
- When it was installed
- Where it was installed
- Who installed it
- How it is managed

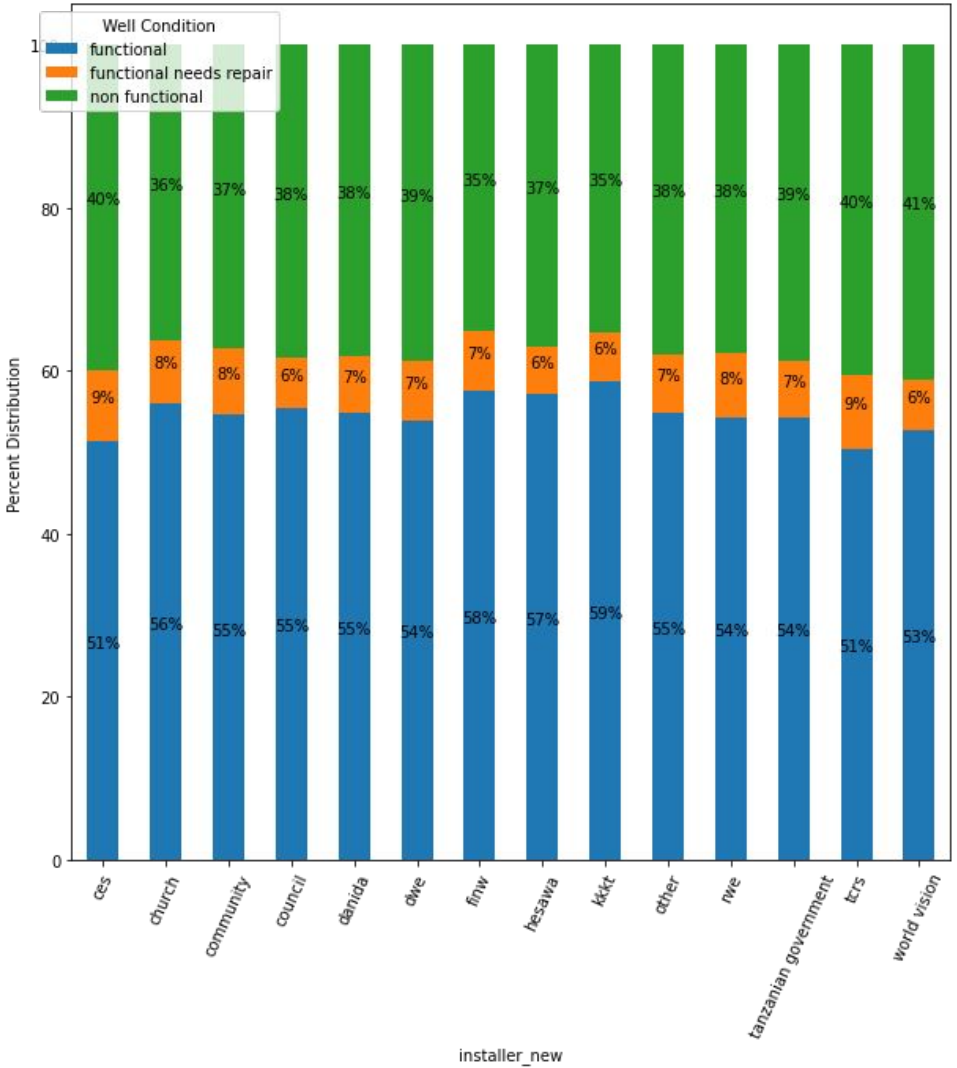
Waterpoint Type

- Dam pumps are much more likely to be non functional



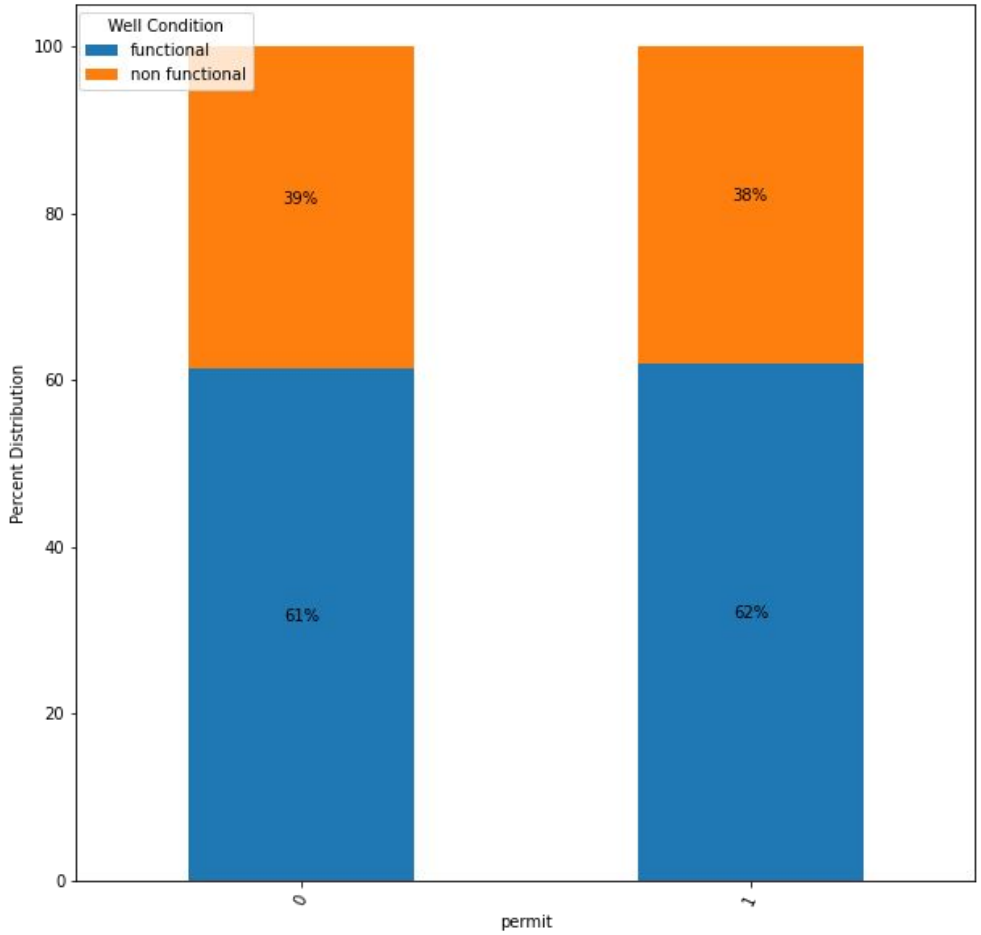
Installer New

- tcrs and ces both had the least fully functional water pumps



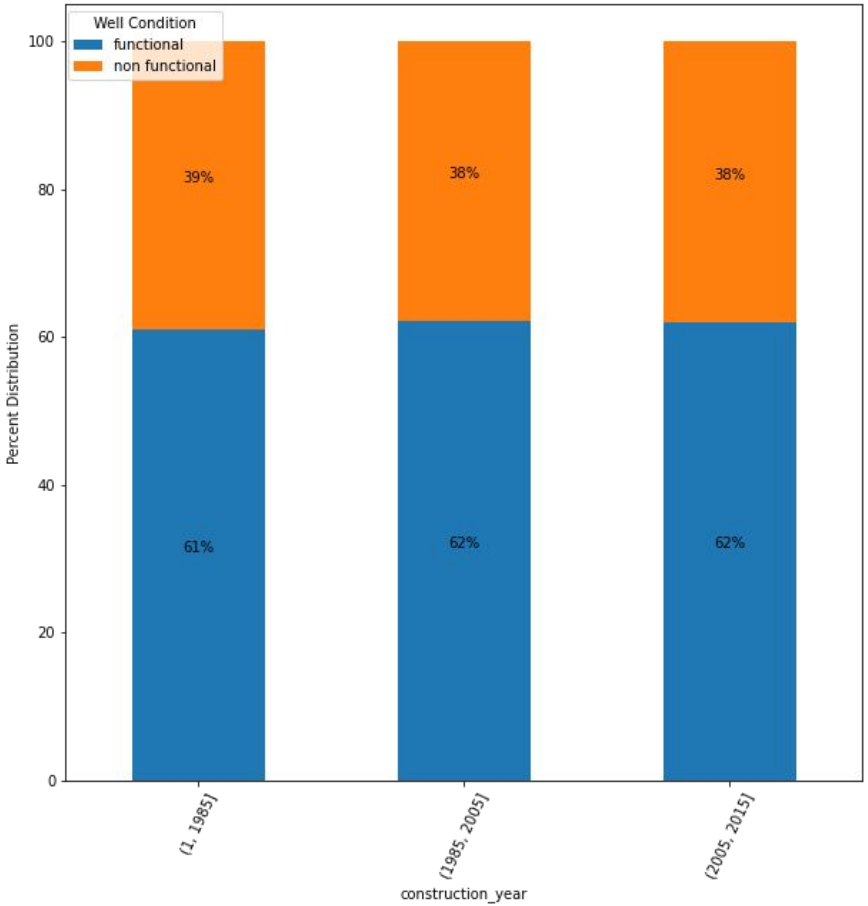
Permitted Wells

- Permitted pumps were more likely to be functional.



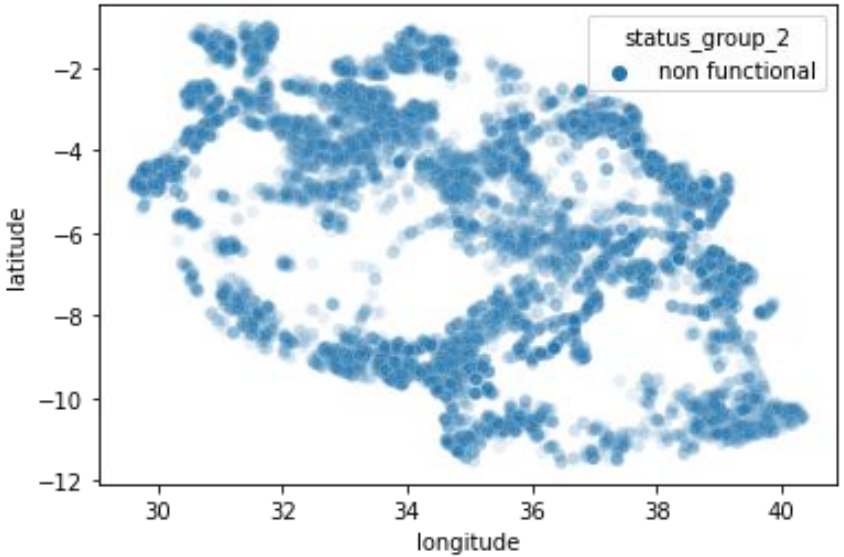
Construction Year

- Pumps older than 1985 are less likely to be functional.

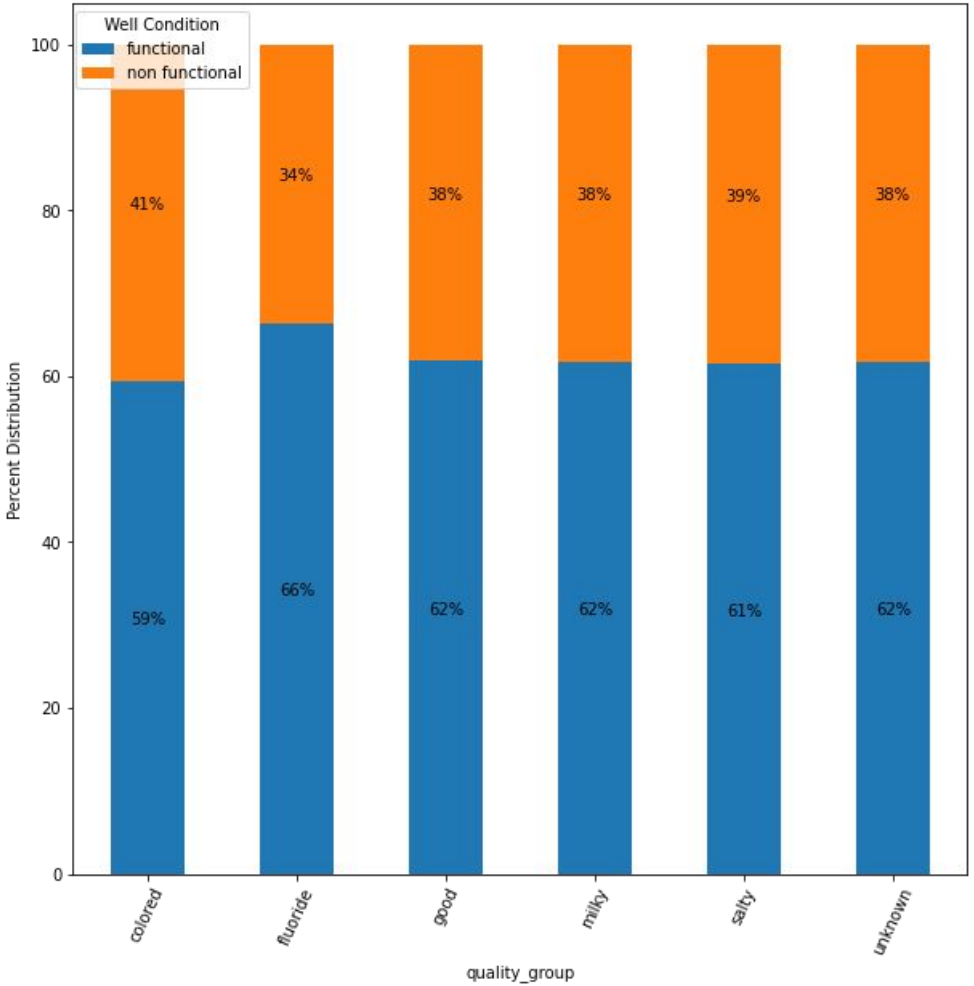


Latitude and Longitude

- latitude and longitude of the waterpoint were the strongest indicator of pump functionality
- Further investigation of specific incidence by region should be investigated



- Water quality which is colored or salty is more likely to lead to non functional pumps



Conclusions

- Installing pumps in low water quality areas leads to more broken pumps
- Permitted pumps were more likely to be functional.
- Pumps older than 1985 have poor functionality

Recommendations

- Expand permitting system. Waterpoints which were permitted were more likely to be functional.
- Areas which already have low water qualities need more maintenance of their pumps we should target these areas for immediate maintenance.
- Add a data feature for best guess of when pump lost functionality
    - This can be used to better predict lifetime of pumps
- Review all pumps older than 1985 as these are more likely to be broken

Future Work

- Exploration of other numerical features such as gps height and population
  - when we examined a boxplot of these features they appeared not to affect the target.
  - However we may have removed these prematurely and should examine the effect of these variables on the model once log normalized.
- Look more closely at pump functionality by geographic region.
  - The latitude and longitude of the waterpoint were the strongest indicator of pump functionality
  - Since this is by far our best feature further investigation is warranted to identify which specific regions are struggling the most with water supply and why.

Questions

Thank you for your time!