



Water Pump Functionality Prediction & Analysis

By: Vincent Welsh
September 21, 2021



Outline

- Business Problem & Understanding
- Data
- Results
 - GPS altitude is the biggest indicator of water pump functionality
 - Missing data regarding waterpoint source is linked to water pump failure
 - Missing data regarding waterpoint extraction method is linked to water pump failure
- Conclusions
- Further Work



Business Summary

- **Business Problem:**
 - The government of Tanzania is obligated to provide accessible clean water to its citizens. However, they are spending too much money for maintenance & repair on the water sources. By determining the leading causes in a faulty water pump the government can prevent mechanical errors. Preventing the faulty pump phenomena will save the government money. Additionally, this data may produce results that are valuable to any organization that builds water pumps.



Business Summary

- **Business Understanding:**
 - In order to save the government of Tanzania money I will determine which factors are most likely to cause water pump failure. To do this I will construct a machine learning model that will predict if a water source is faulty or not. Once I have confirmed my model is reliable, I will explain and show which variables are most correlated with faulty water sources. This will inform the government of Tanzania and other water pump manufacturers which factors cause a faulty pump.



Data

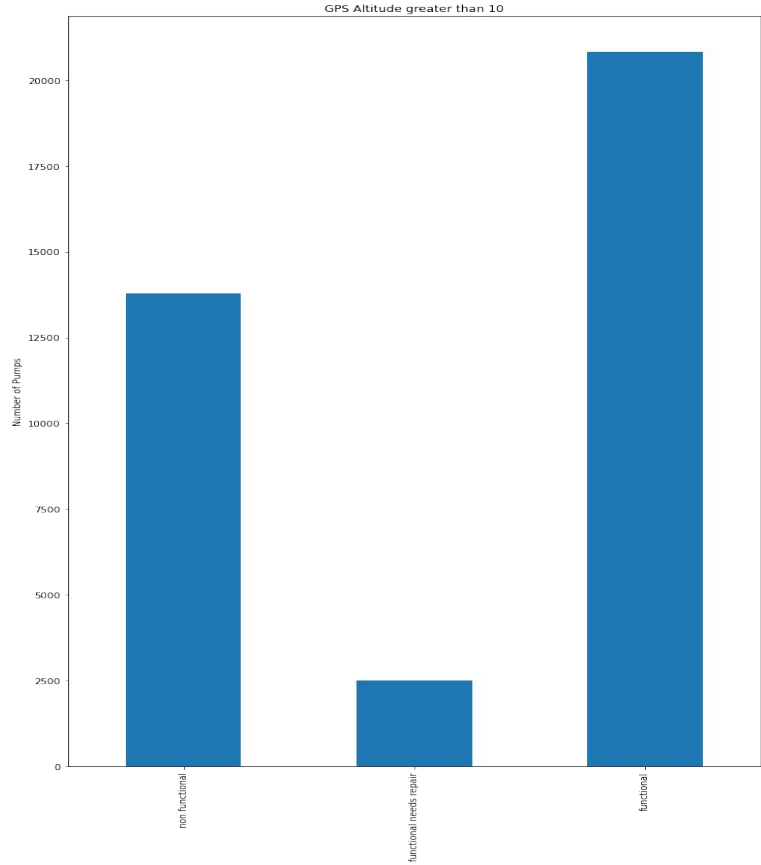
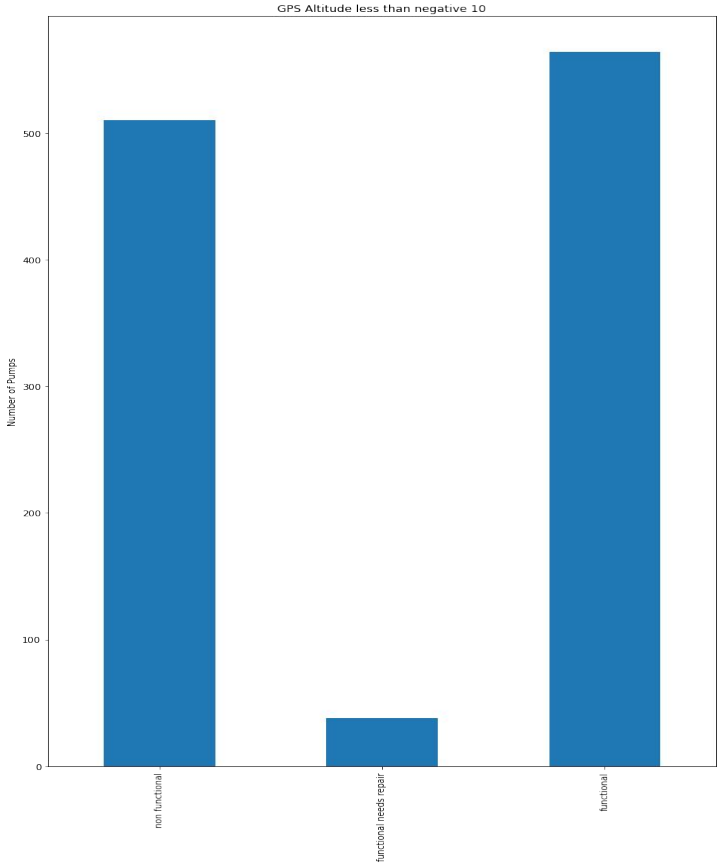
- The data from this competition comes from the Tanzania Ministry of Water. The data includes 3 classifications: Functional, Not-Functional, Functional-Needs-Repair. With the following predictors:
 - The date the water source was build
 - Geographic Location
 - Who funded the well
 - Altitude of the well
 - Organization that installed the well
 - Name of the waterpoint
 - Population around the well
 - If there was a public meeting (True/False)
 - Who operates the waterpoint
 - Permit (True/False)
 - Cost of the water
 - Water quality/quantity
 - Source of the water



Results: Feature Importance

1. GPS Height
2. Water point type
3. Extraction Type
4. Population
5. Amount of water available to waterpoint

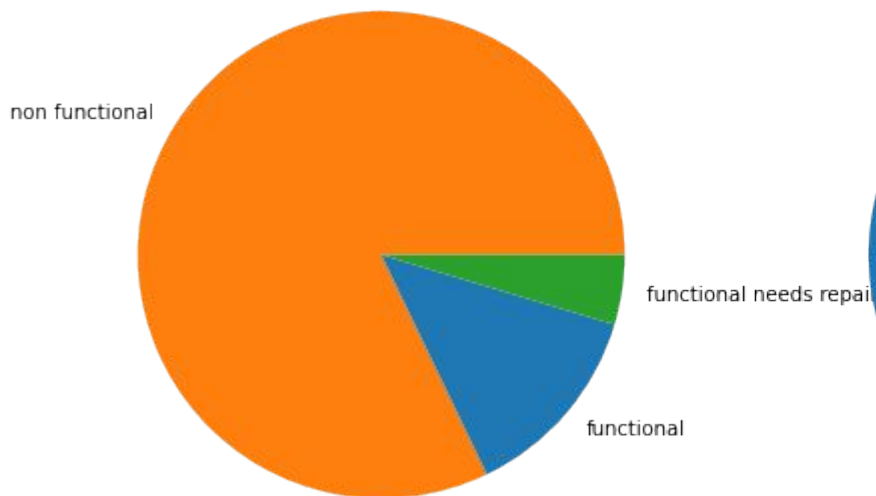
Results:GPS Height



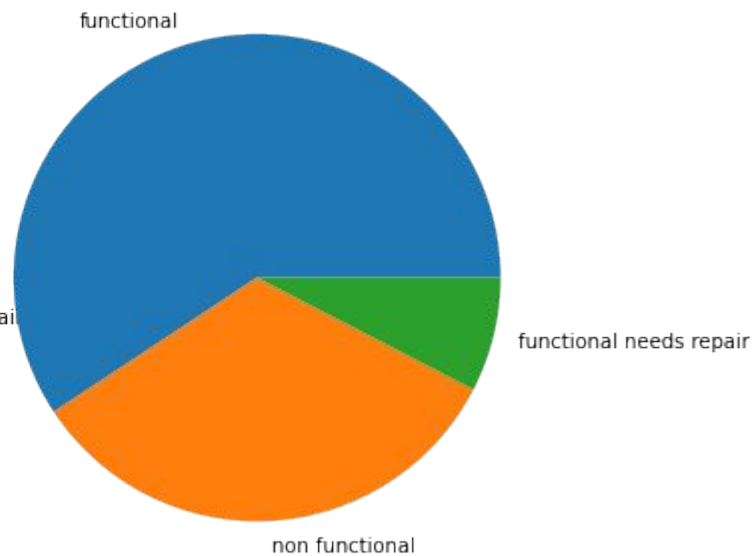
Results: Waterpoint Type



Target dist when waterpoint_type is unknown



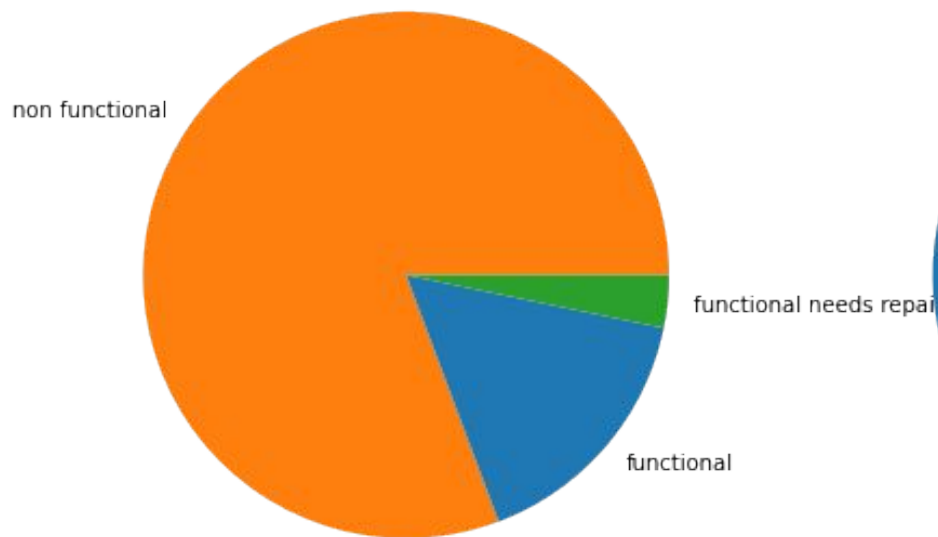
Target dist when waterpoint_type is known



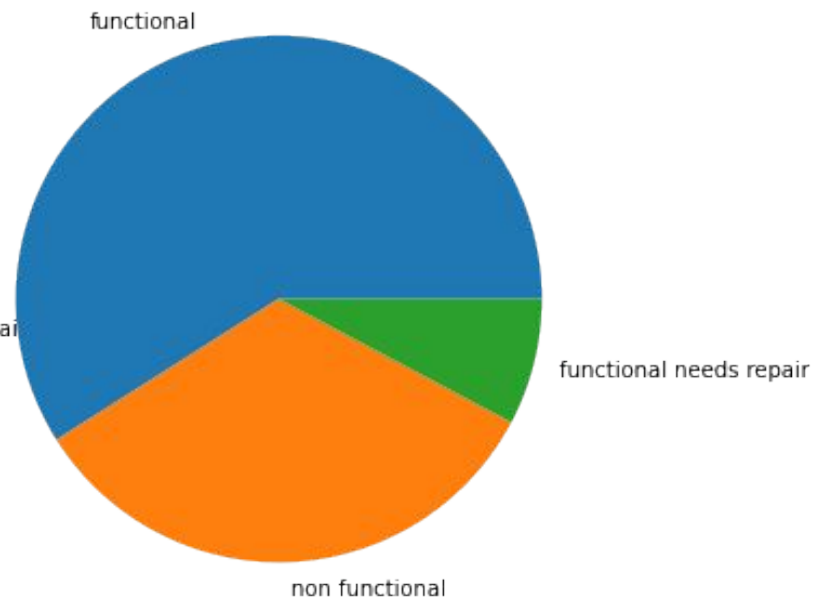
Results: Extraction Type



Target dist when extraction_type is unknown



Target dist when extraction_type is known





Conclusions

- Water points in lower altitude regions are having more problems
- When waterpoint source is unknown there are less functional water pumps
- Where the extraction method is unknown there are less function water pumps.



Further Work

- Provided more time with this data, I can derive the following insights:
 - How pump functionality differs region to region
 - How population can affect water-pump functionality
 - Determine how much water availability affect pump functionality



Vincent Welsh

Email: jvincentwelsh99@hotmail.com

Github: [@Eucalyptusss](https://github.com/Eucalyptusss)

LinkedIn: www.linkedin.com/in/vincent-404