Water Pump Functionality Prediction & Analysis

By: Vincent Welsh
September 21, 2021

Outline

- Business Problem & Understanding
- Data
- Results
  - Pumps at higher altitudes are less functional
  - Certain rural ward's need aid to help with waterpoint source establishment and record keeping
  - When water pump sources go dry water pump failure follows
- 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

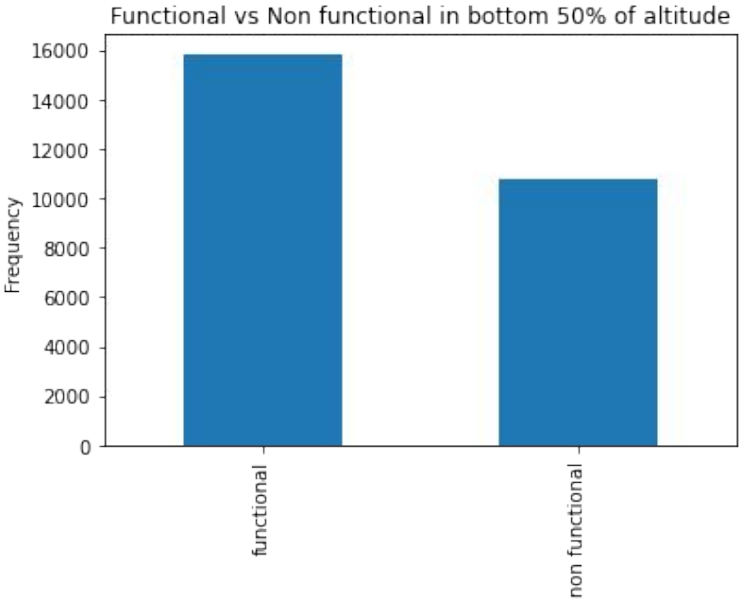
Results: Feature Importance

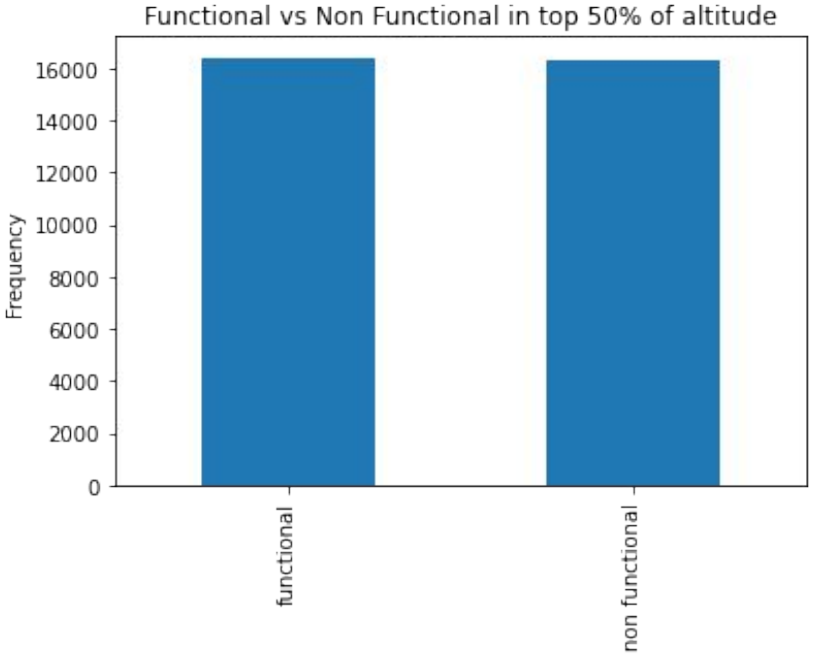
Our model correctly predicts pumps that need repair 84.3% of the time and correctly predicts 72.4% of all pumps needed repair. From this model I know the most predictive features are:

1. Quantity of water
2. Water point type
3. Region
4. LGA
5. Extraction Type

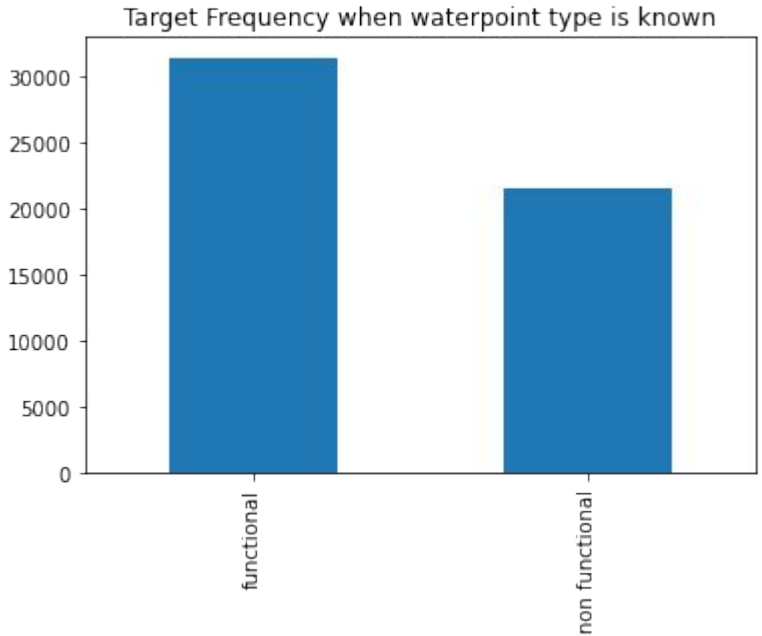
Results:GPS Height

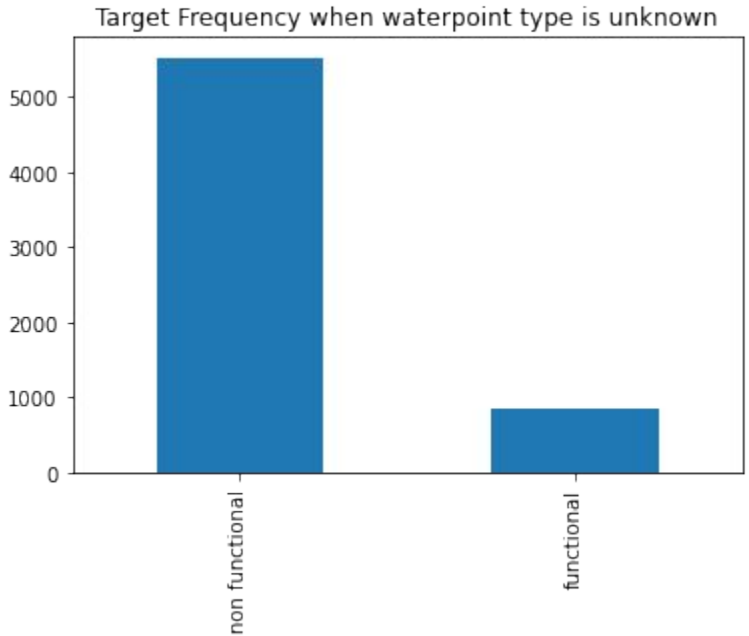
Target Frequency separated by mean altitude of 668 m:





Results: Waterpoint Type in Rural Wards (Part 1)

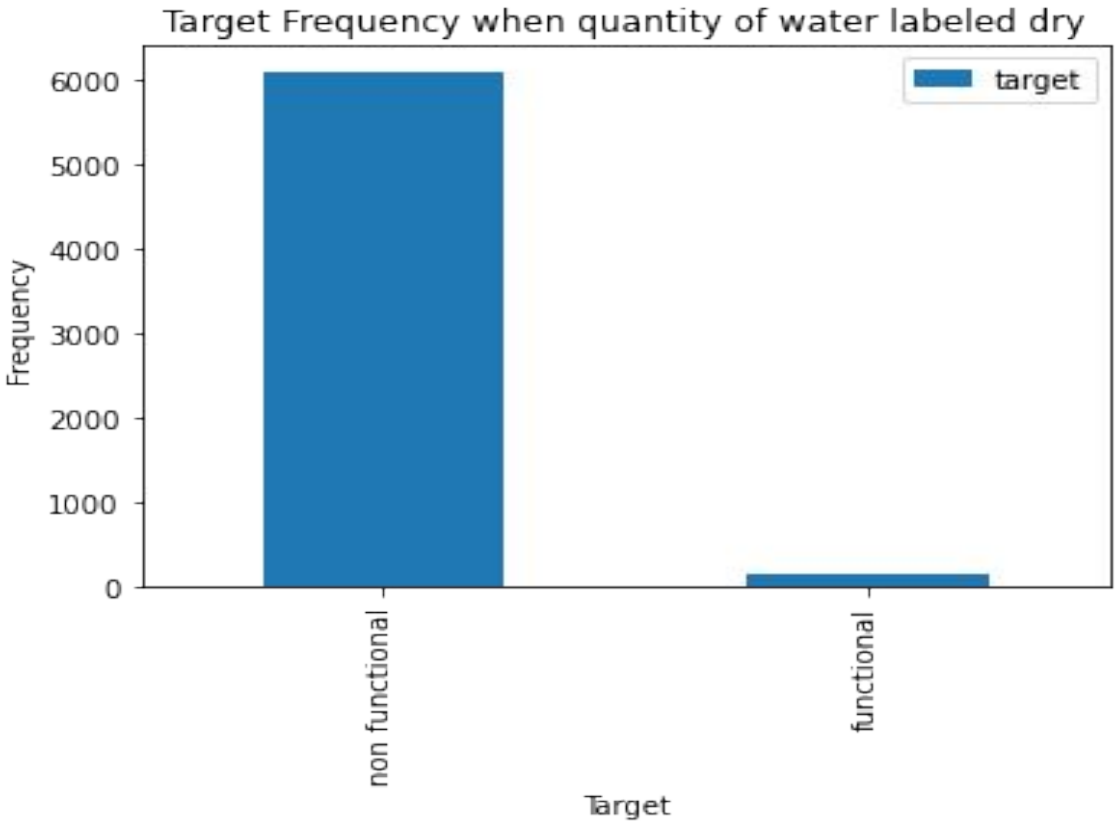




Results: Waterpoint Type in Rural Wards (Part 2)

- **Marangu Mashariki** (23,734 population)
    - Known waterpoint types: 8
    - Unknown waterpoint types: 43
- **Namajani** (13,739 population)
    - Known waterpoint types: 13
    - Unknown waterpoint types: 51
- **Pangani Mashariki** (54,025 population)
    - Known waterpoint types: 13
    - Unknown waterpoint types: 51

Results: Quantity of Water



Conclusions

- Build future water pumps above the altitude of 650 meters when possible
- Send aid to construct and record waterpoint sources to rural wards, specifically:
  - Marangu Mashariki
  - Namajani
  - Pangani Mashariki
- Build future pumps so that their water source is abundant i.e. cannot within reason dry out.

Further Work

- Provided more time with this data, I can derive the following insights:
  - How much does population affect demand for water and how does this demand affect pump functionality?
  - What is the ideal extraction method?

Vincent Welsh

Email: jvincentwelsh99@hotmail.com
Github: @Eucalyptusss
LinkedIn: www.linkedin.com/in/vincent-404