

TANZANIAN WATER PUMP PROJECT



MODULE 3 MACHINE LEARNING

Isabel Joseph and James Moulds

Challenge:

The Tanzanian Ministry of Water (TMOW) is having trouble keeping track of all the water pumps that are in operation around the country. They would like a way of predicting which wells are non-functioning or which require some maintenance. The aim of this project is to help the TMOW assign predictions to all its wells in the form of the following categories:

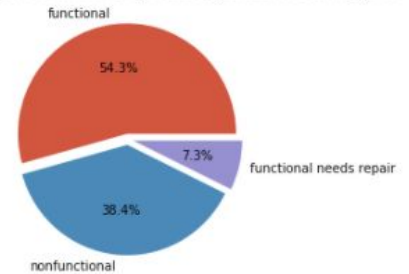
- The well is functional
- The well is functional but needs repair
- The well is non functional



Approach:

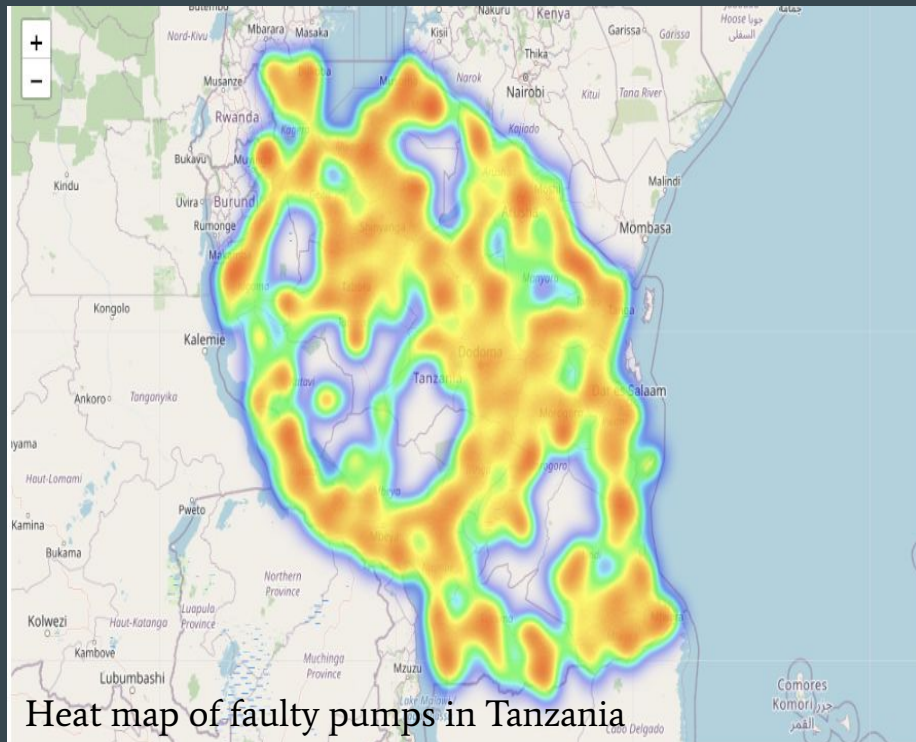
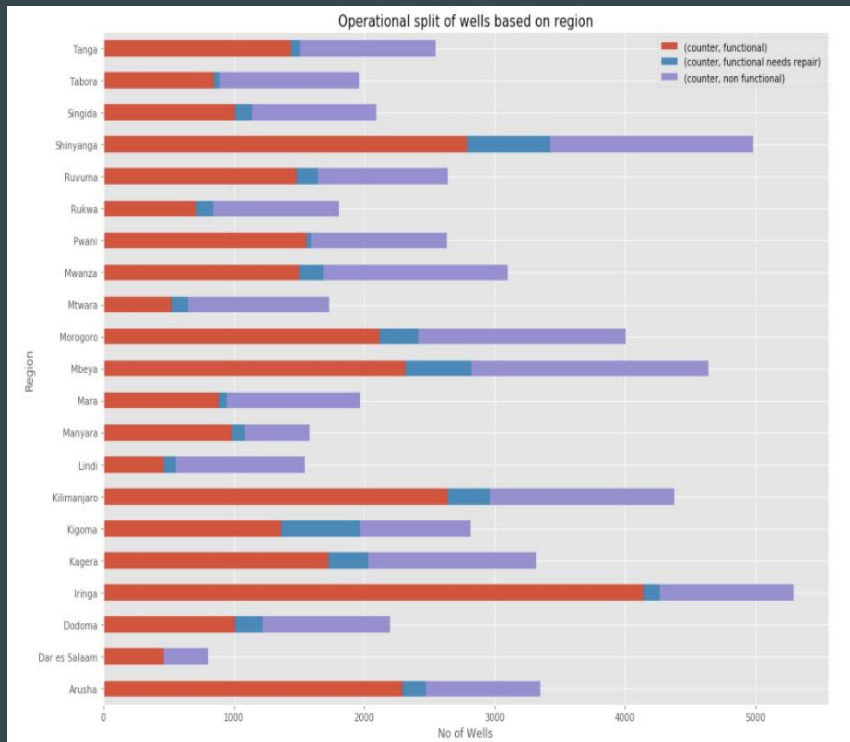
Due to the complex nature of a three class problem we have decided to split our pump categories into 'functional' and 'faulty' which combines the other two classes. This has the twin benefit of simplifying our modelling process but also removing the imbalance in our data set with the undervalued class of 'functional - needs repair' now being absorbed into a bigger set.

Chart to show split of three target categories in training set



Geographic Indicators

Best performing region is Iringa with 4 out of 5 pumps fully functional, Arusha not far behind with 2 out of 3 wells functional
Worst performing region is Lindi with 3 out of 4 pumps in need of repair, Mtwara not far behind with 2 of 3 wells in need of repair
We advise that practices and resources from Iringa and Arusha are shared with the poorly performing regions



Temporal indicators

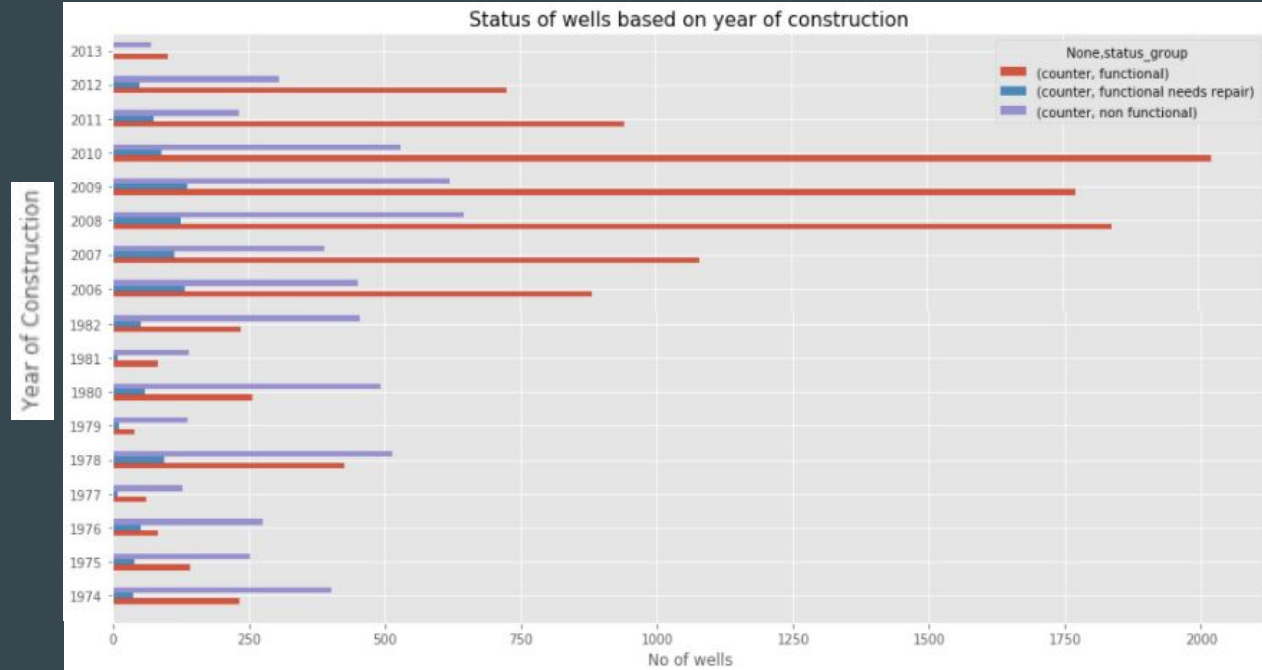
Age of the pumps is a clear indicator for the likelihood of a faulty pump.

69% of pumps **installed** in the 8 year period between **2006 and 2013** are reported as **functional**.

66% of pumps installed in the 8 year period between **1974 and 1982** are reported as **faulty**.

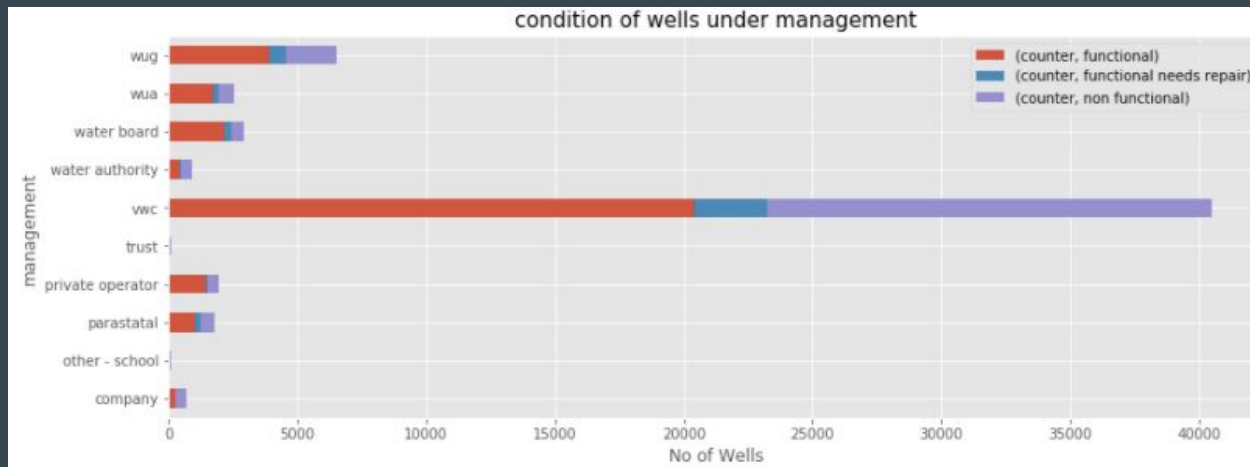
Broad brush advice would be to inspect and overhaul old stock installed from this period.

More complete data would increase accuracy of this advice.



Is poor management a factor?

Background research and anecdotal evidence suggested that poor management and inefficient maintenance regimes were largely blamed for the prevalence of faulty pumps which currently sits somewhere around 40% of all pumps in Tanzania. It seems from this image that one company type - vwc (village water committee) is responsible for the maintenance of a very large proportion of the nation's pumps. Potentially offering greater training and centralised support could improve the situation in these case.



Our predictors performance

Take Away 1 Of all the faulty pumps in the set, we correctly identified 71% of them!!

Take Away 2 Of all the pumps that we said were faulty only 60% were actually faulty which might mean a large amount of unnecessary engineer visits.

Because of the large relative cost of false negatives in this case, and the fact that the nature of this project is likely to have a greater tolerance for false positives than false negatives, our threshold is set low.

To put another way our client is likely to value reduction in the amount of faulty wells that we miss over the amount of functioning wells that we incorrectly predict are faulty.

This situation could be improved if our model was able to separate the negative and positive distributions more effectively meaning the trade off between FP and FN could be minimised.