FUNCTIONING WELLS IN TANZANIA

By: Lauren Esser

PROBLEM STATEMENT

Tanzania has an issue with water pumps failing. Within this project the goal was to predict which pumps are functional, which need some repairs, and which don’t work at all.

Image from drivendata.org

BUSINESS VALUE

- **CLEAN, POTABLE WATER FOR COMMUNITIES**
  - In sub-Saharan Africa hundreds or millions of people suffer from a lack of access to clean and safe water. (thewaterproject.org)
  - Globally 1 in 9 people still have no access to clean water.
- **REDUCE SICKNESS**
  - By providing clean water, citizens are less likely to drink unsafe water with microbes and pollution.
- **FIGHT HUNGER**
  - Access to water allows for less crop loss, hunger is reduced.
  - Schools can grow gardens that feed students and also teaches them valuable life lessons.
- **EDUCATION**
  - Children are freed from having to travel far to gather water and instead can attend class.
  - Girls under age 15 are twice as likely as boys to be the family member responsible for fetching water.
- **HELP FIGHT POVERTY**
  - Access to clean water can help communities build businesses

Data was taken from Taarifa and Tanzanian Ministry of Water

METHODS

The OSEMN Method was used in preparing the data and model creation. I will go over each step in the upcoming slides.

OBTAIN

- Uploaded data_values and data_labels csv.
- Observe unique groups and inspect information

SCRUB

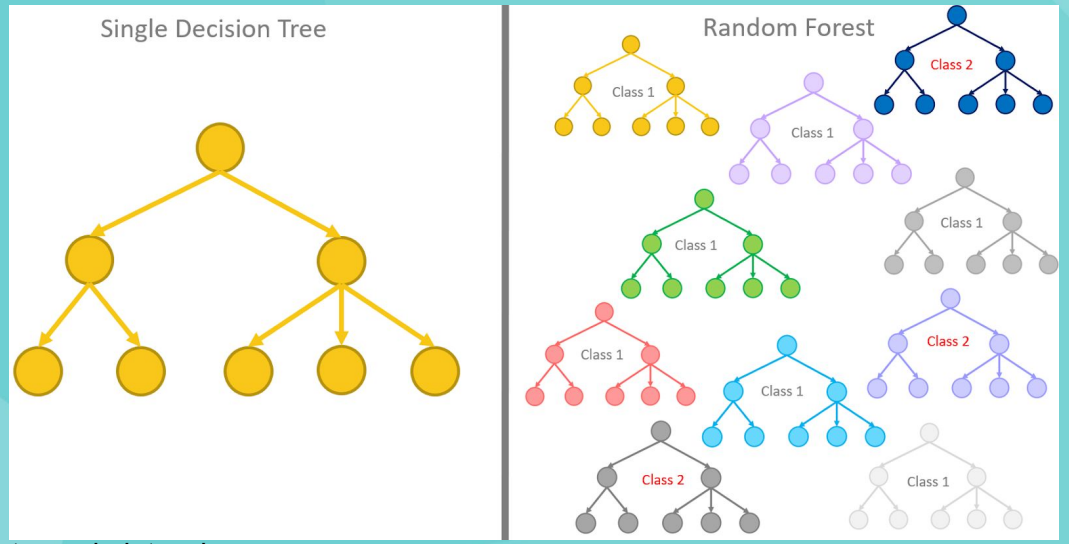
- Dropped repetitive columns
- Construction Year: Replaced 0 with 1959. Assumed well was built prior to recording in 1960.
- Public Meeting: replaced nulls with majority and said well was in a public place.
- Population: replaced 0 with median population.
- Subvillage & Scheme Management: replaced Nulls with “Unknown”
- Permit: changed 25.0 to False

EXPLORE

- Created Visuals
- Convert categorical data into indicator variables
- Dropped unneeded categorical columns
- LabelEncoder on my dependent group
- Train Test Split on the dataset

MODEL

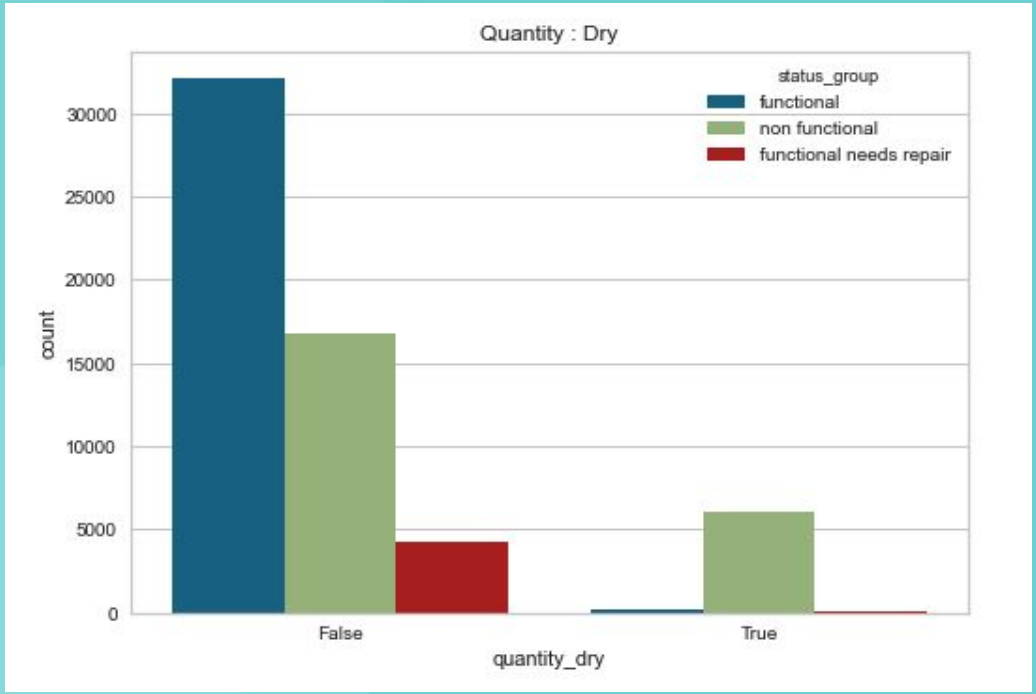
- Decision Tree Vanilla
- GridSearch with Decision Trees
- Random Forests Vanilla
- GridSearch with Random Forests



towardsdatascience.com

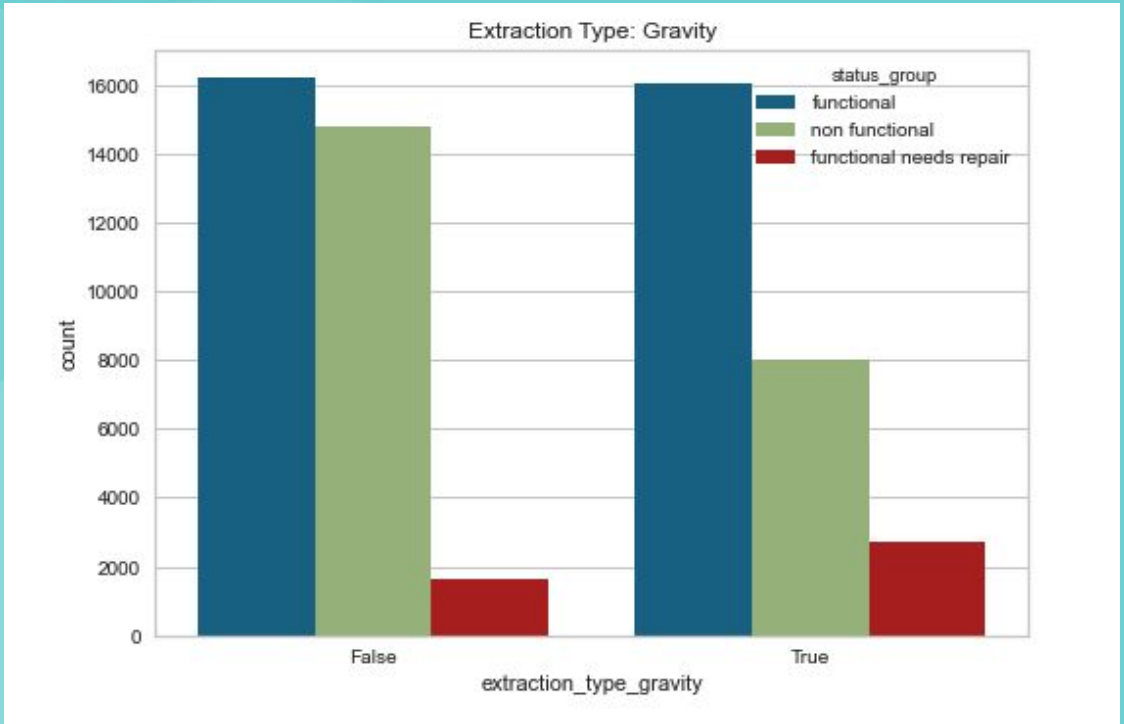
INTERPRETATIONS AND RECOMMENDATIONS

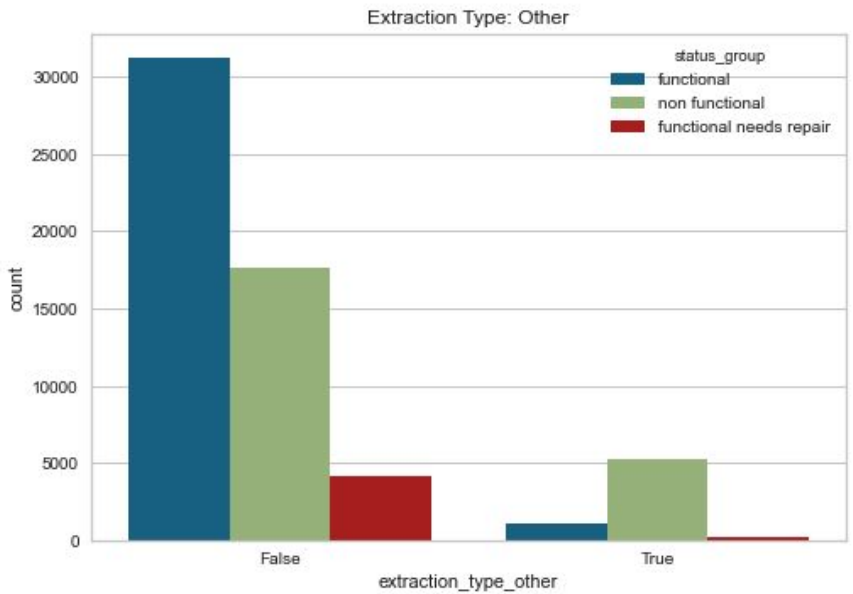
1. Keep track of the amount of available water in the wells not functioning.

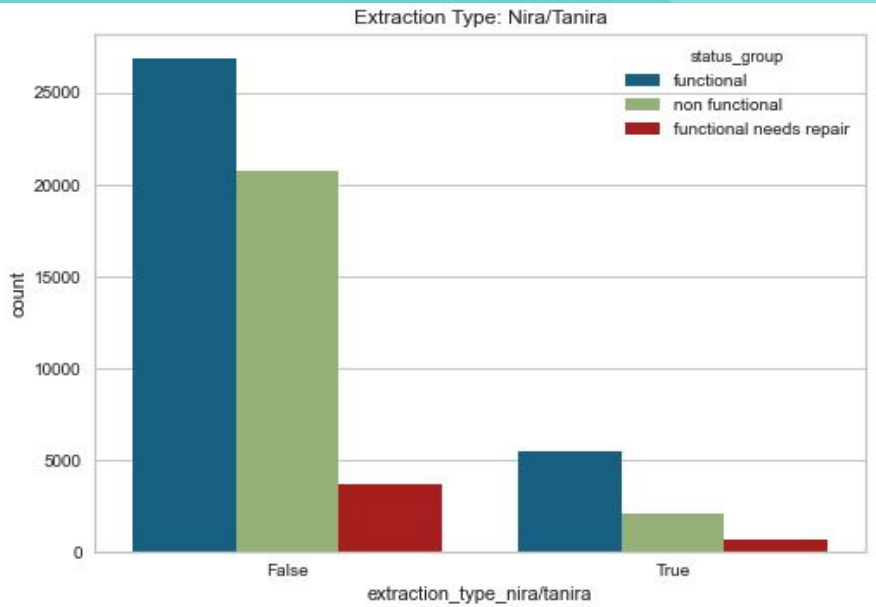


INTERPRETATIONS AND RECOMMENDATIONS 2

2. Look at how water is being extracted from wells. Focus groups are Gravity, Nira, and other.

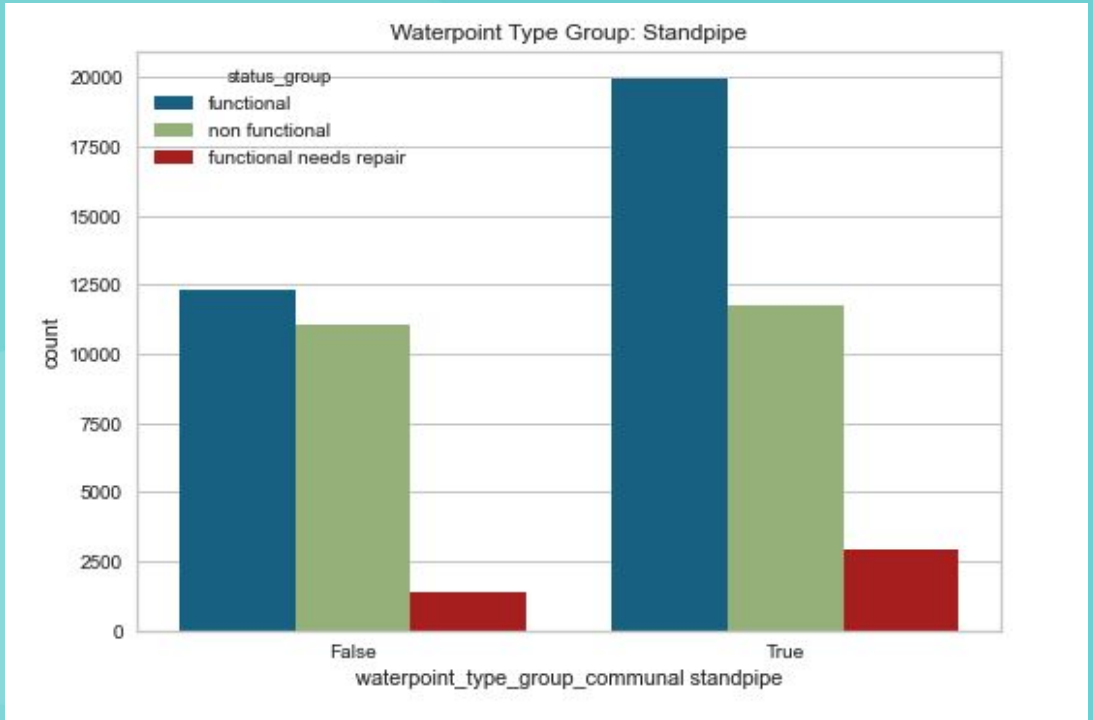


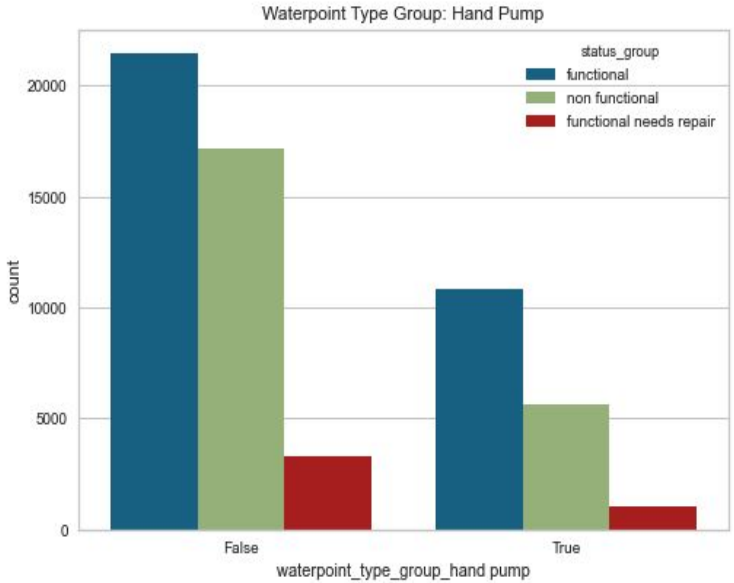

Extraction Type: Other
status_group
functional
non functional
functional needs repair
30000
25000
20000
15000
10000
5000
0
count
False
True
extraction_type_other

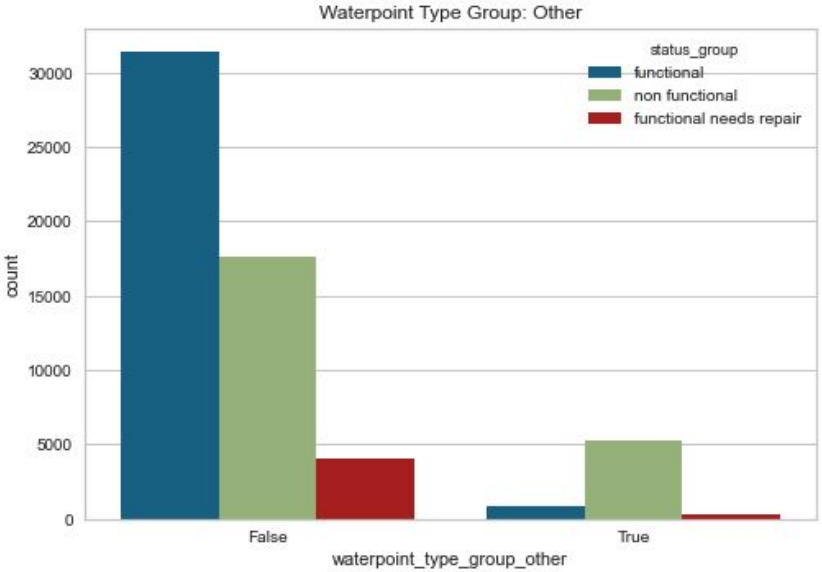

Extraction Type: Nira/Tanira
status_group
functional
non functional
functional needs repair
25000
20000
15000
10000
5000
0
count
False
True
extraction_type_nira/tanira

INTERPRETATIONS AND RECOMMENDATIONS 3

3. Take a look at waterpoint type focusing on Communal Standpipe, Handpumps, and “Other”.

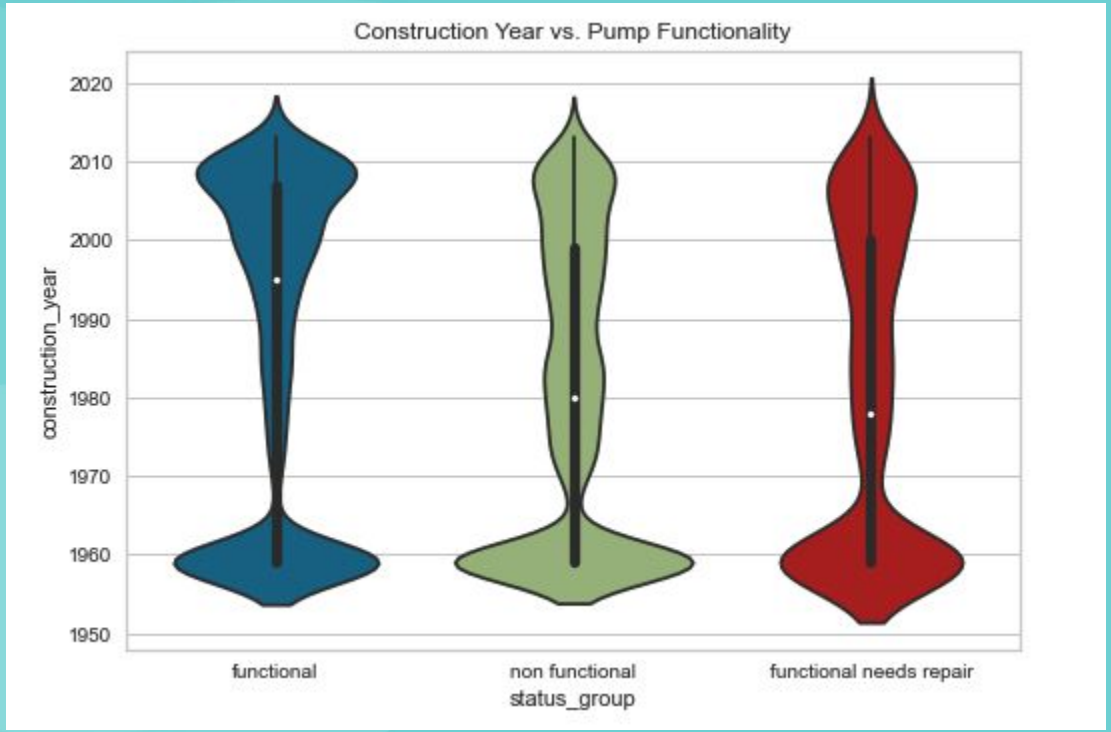



Waterpoint Type Group: Hand Pump
status_group
functional
non functional
functional needs repair
20000
15000
10000
5000
0
count
False
True
waterpoint_type_group_hand pump


Waterpoint Type Group: Other
status_group
functional
non functional
functional needs repair
30000
25000
20000
15000
10000
5000
0
count
False
True
waterpoint_type_group_other

INTERPRETATIONS AND RECOMMENDATIONS 4

4. Look at pumps that are marked as construction before 1965 or have no construction date available.



RECOMMENDATIONS SUMMARY

1. Ensure water pumps have access to a water source.
2. Check wells that use extraction type of Gravity, Nira, or are labeled as “Other”.
3. Check wells that use waterpoint type of Handpump, Standpipe, or are labeled as “Other”.
4. Inspect pumps that are built before 1965 or have no construction year listed.

NEXT STEPS

1. What are the “other” types of well extraction?
2. Does who manufactured the pump play a role in functionality?
3. Does number of daily uses impact functionality?
4. Does distance from the ocean or a large body of water play a role?
5. Does the amount of pollution within the community play a role on functionality?

THANK YOU

Farther Questions?
Contact me at CLEsser02@gmail.com or
Lauren Esser on LinkedIn