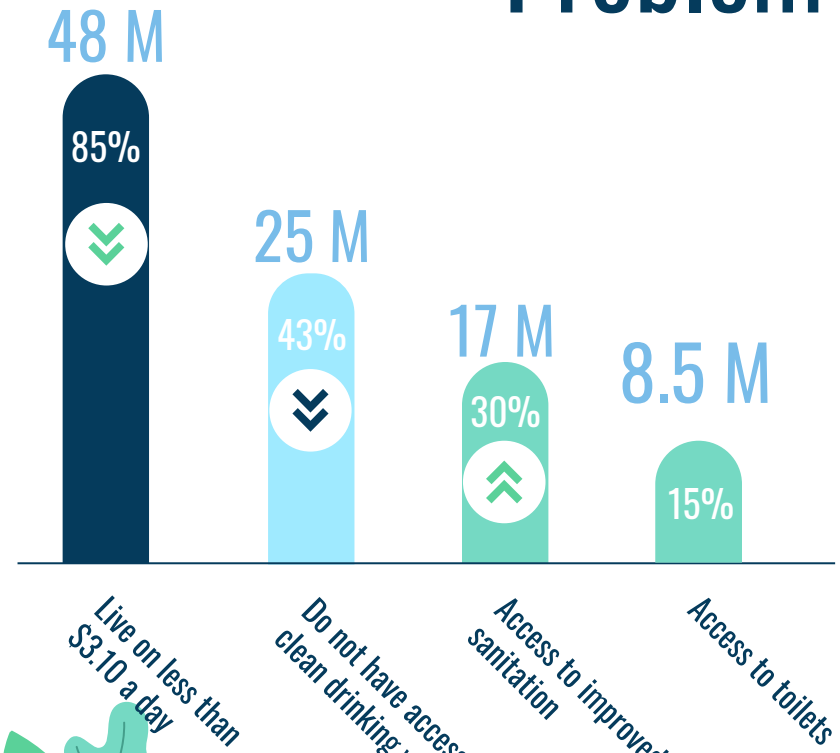
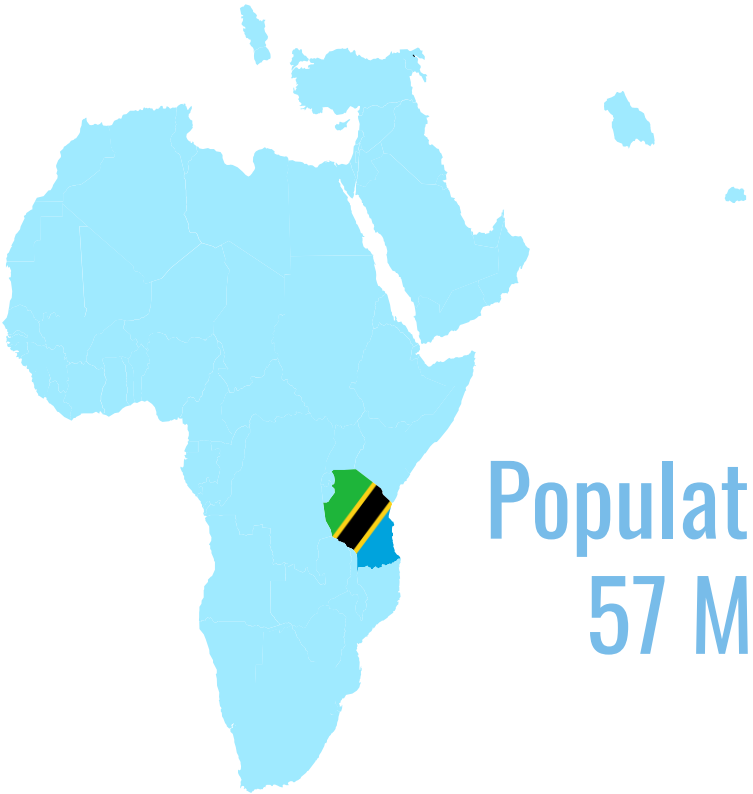
Pump It Up

Data Mining the Water Table

Presented By:
Morgan Jones

Problem Statement





Population
57 M

The United Republic of Tanzania Ministry of Water

Mission

To improve access to safe drinking water and sanitation services to all, and manage water resources so as to ensure national food security and sustainable industrial based economic development.

OUR GOALS

Analysis

Discover valuable insights to share with Ministry of Water to achieve their mission

Accuracy

Generate a model with > 0.80 accuracy in predicting functionality

Precision

Lowest value of false positives: model predicting a well is functional when it is not

OUR OSEMiN PROCESS

Obtain

Import the data

Scrub

Manage datatypes, resolve missing/duplicate values

Explore

Find patterns among the relationships of variables in the dataset

Model

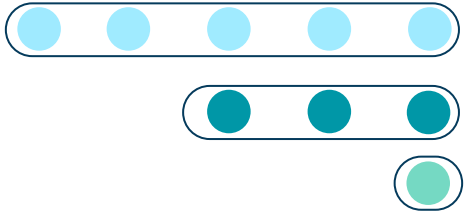
Create predictive models

iNterpret

Identify insights and visualize findings

OUR DATA

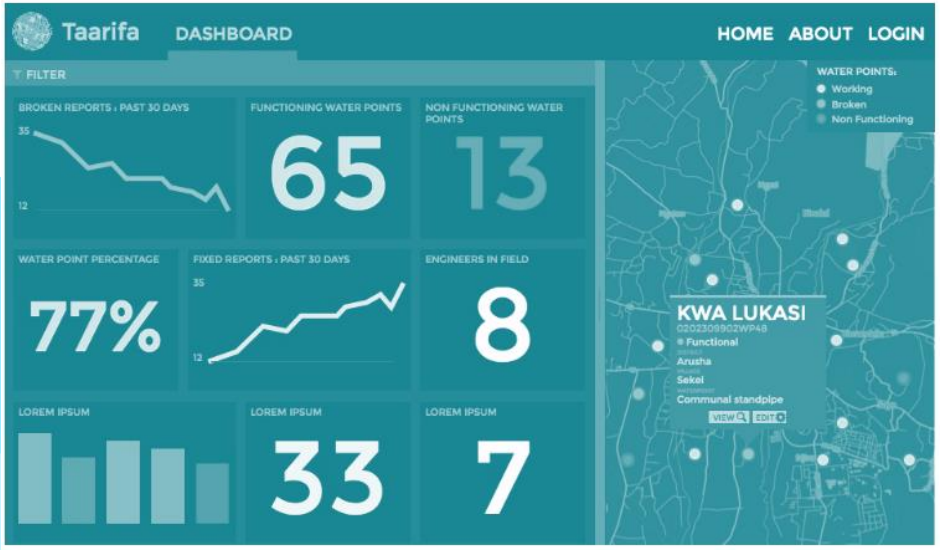
Target



Functional 54%

Non-Functional 38.4%

Functional needs Repair 7.2%



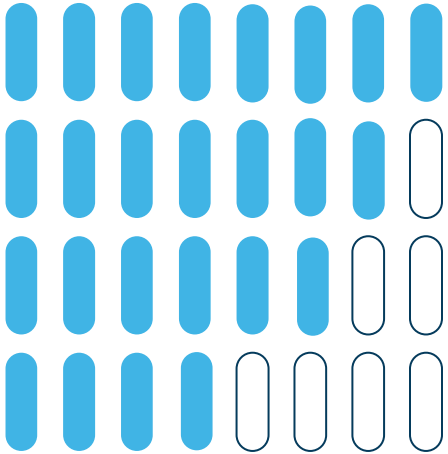
Key Features

COORDINATES

ELEVATION

CONSTRUCTION YEAR

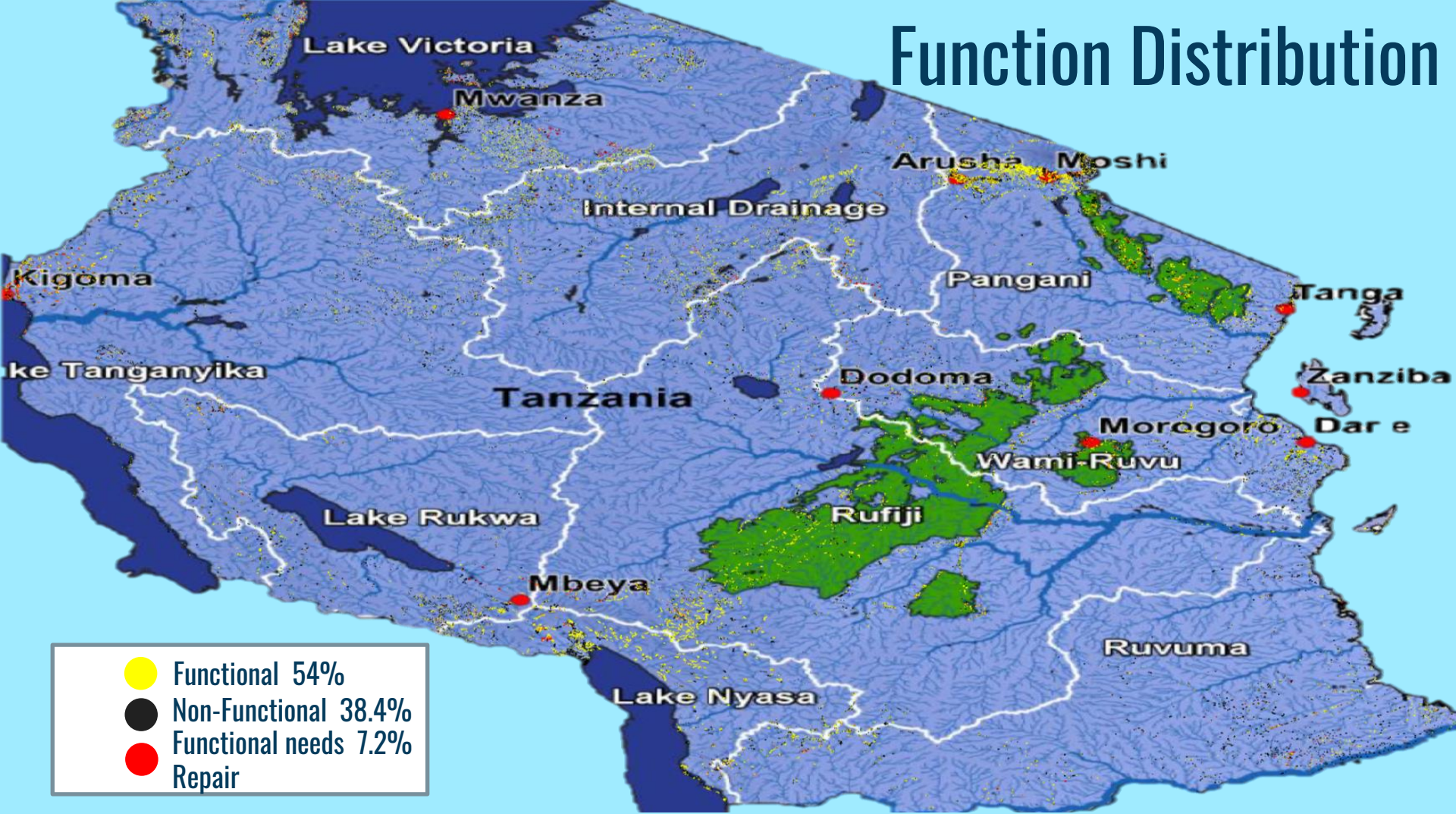
Amount of Water Available



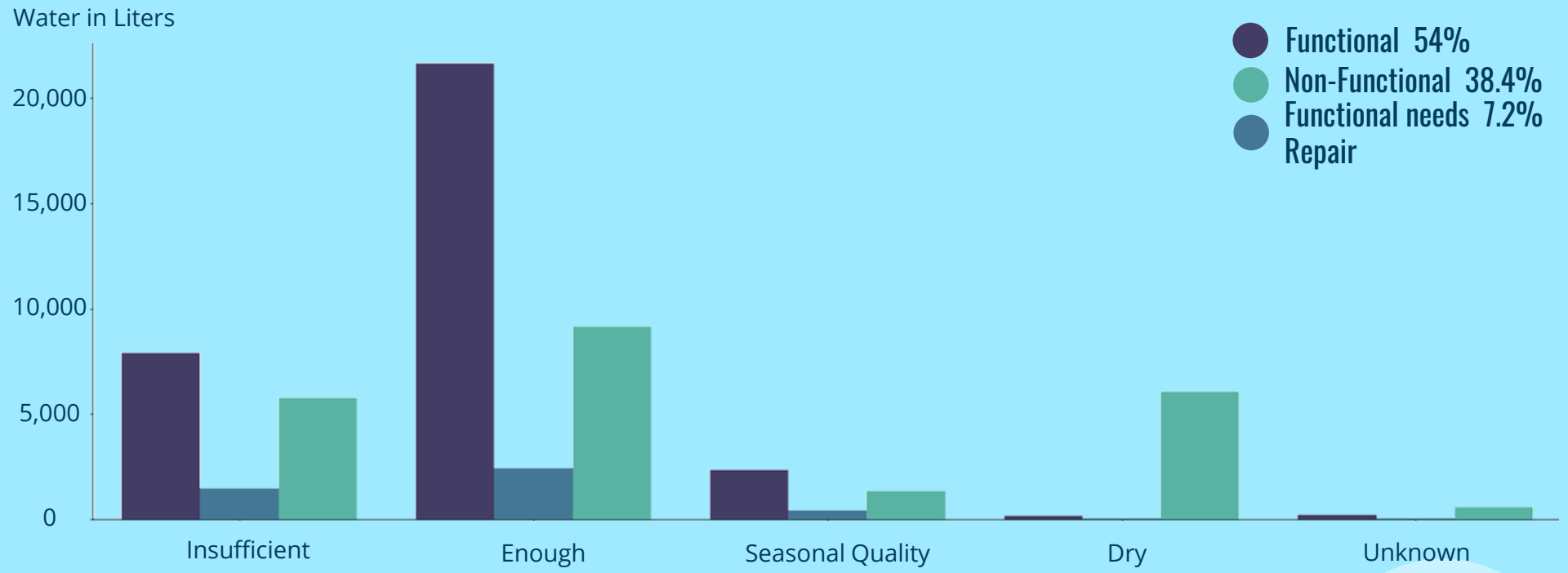
59400 Water Points

40 Descriptors

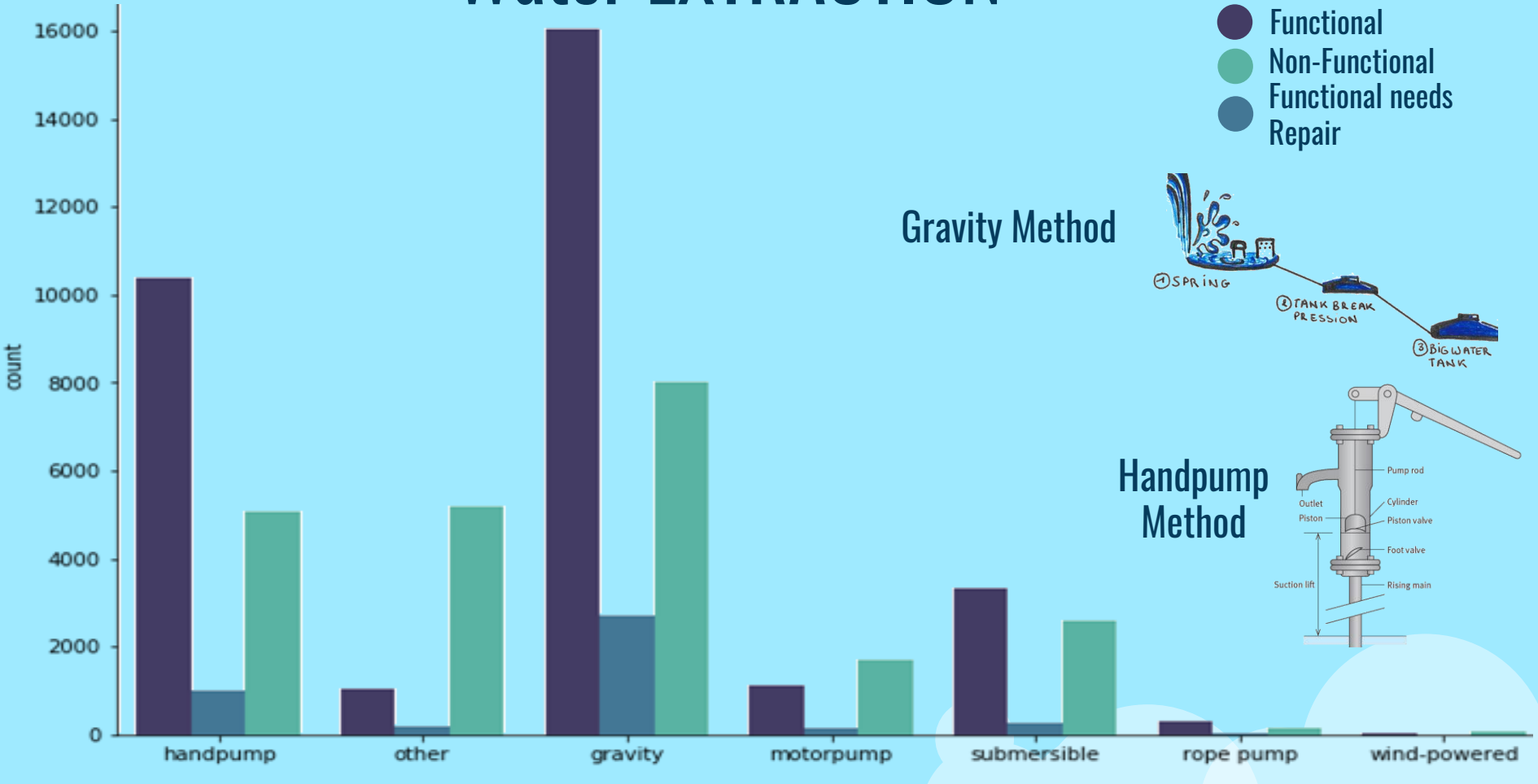
Function Distribution



Water QUANTITY



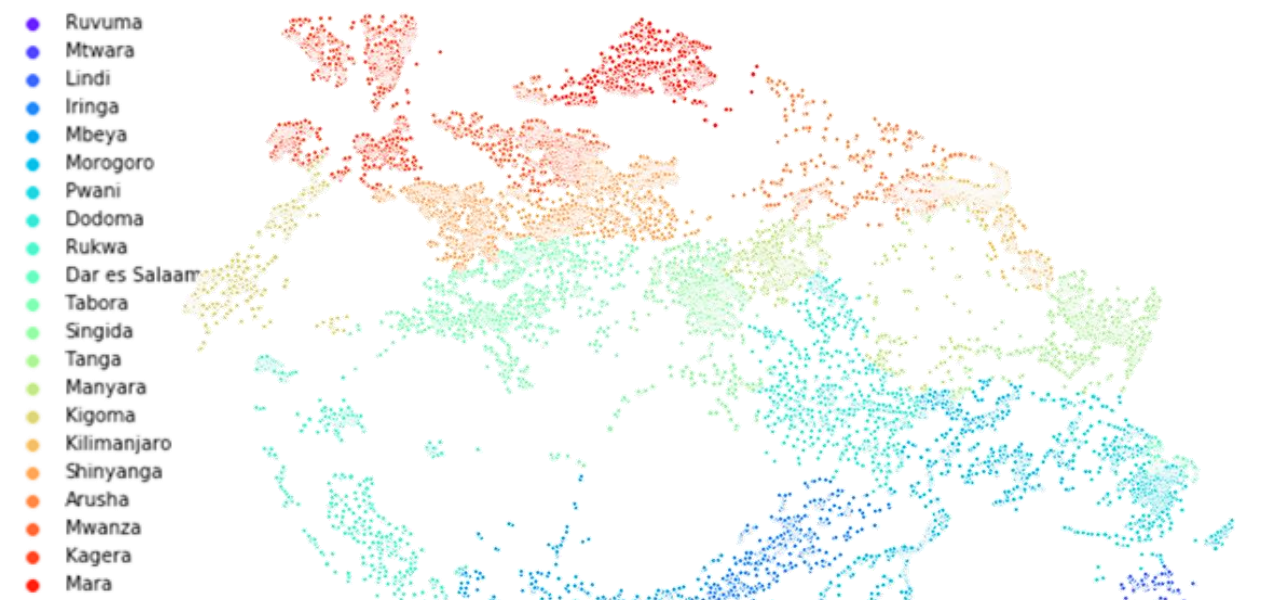
Water EXTRACTION

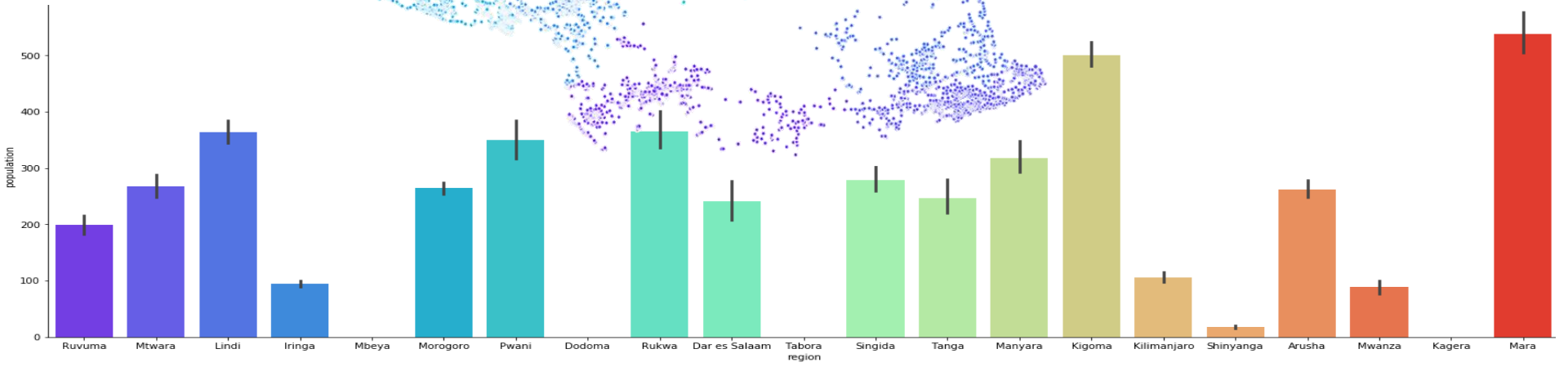


Gravity Method

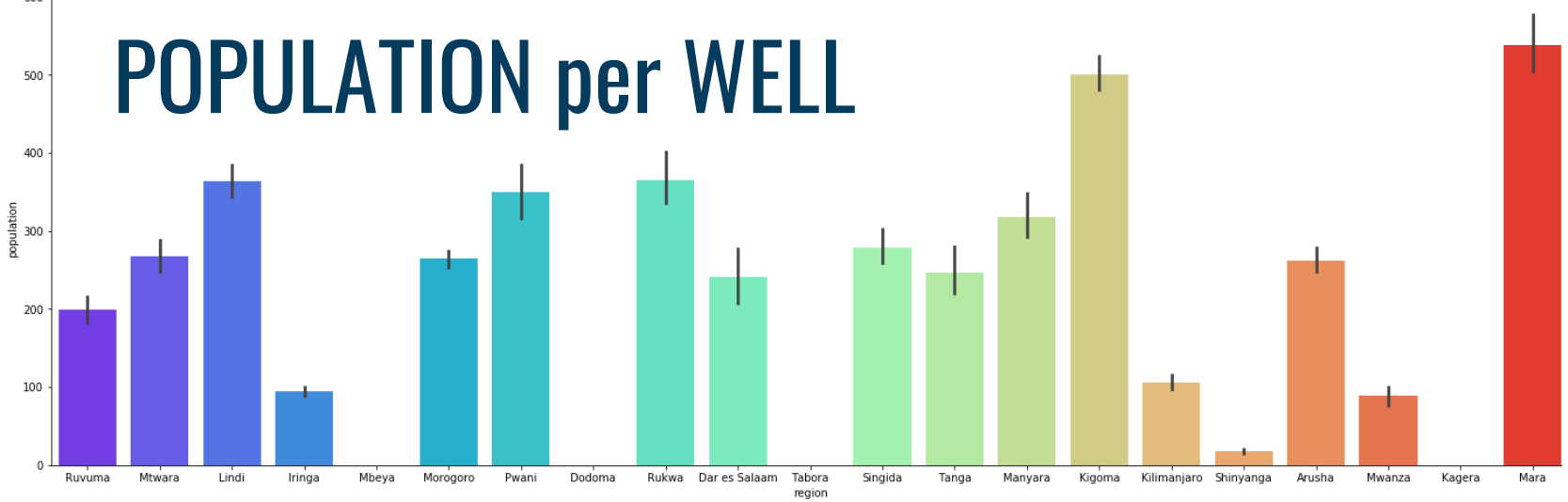
Handpump Method

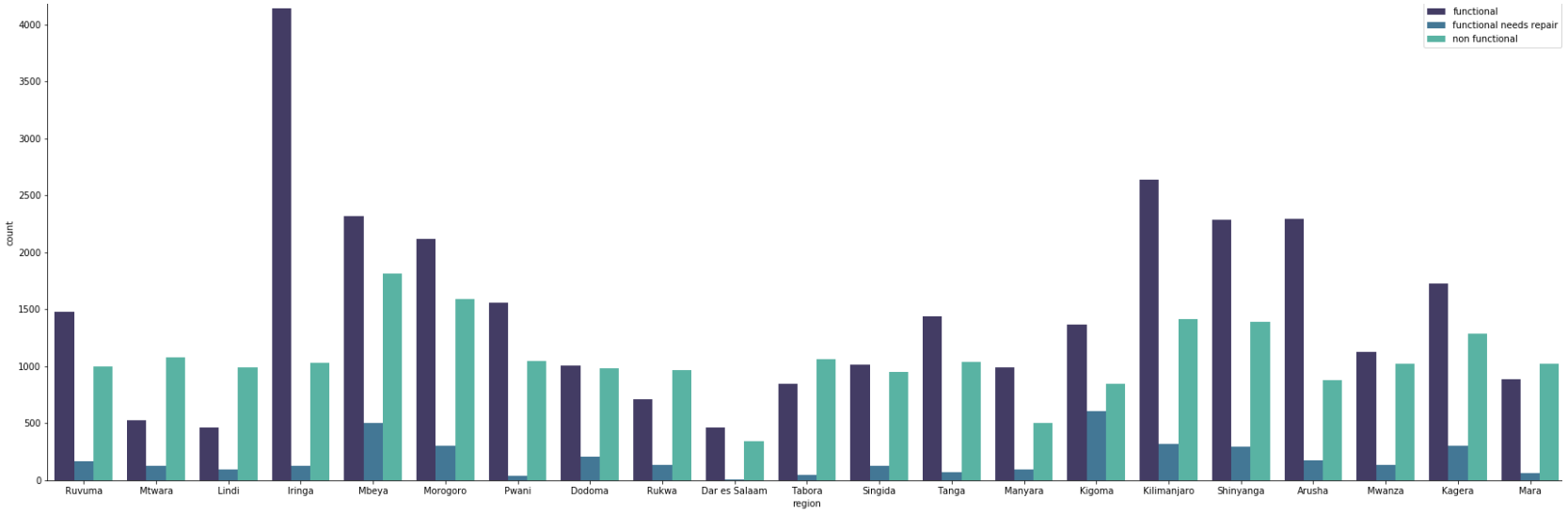
POPULATION per WELL Distribution by REGION



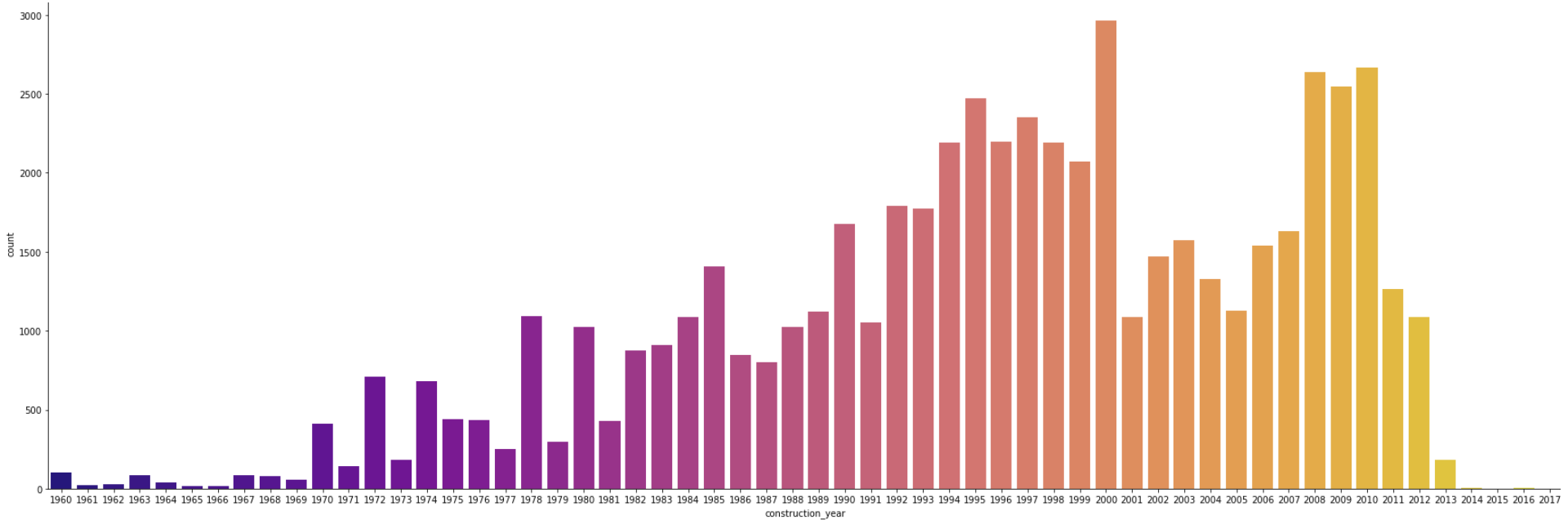


POPULATION per WELL

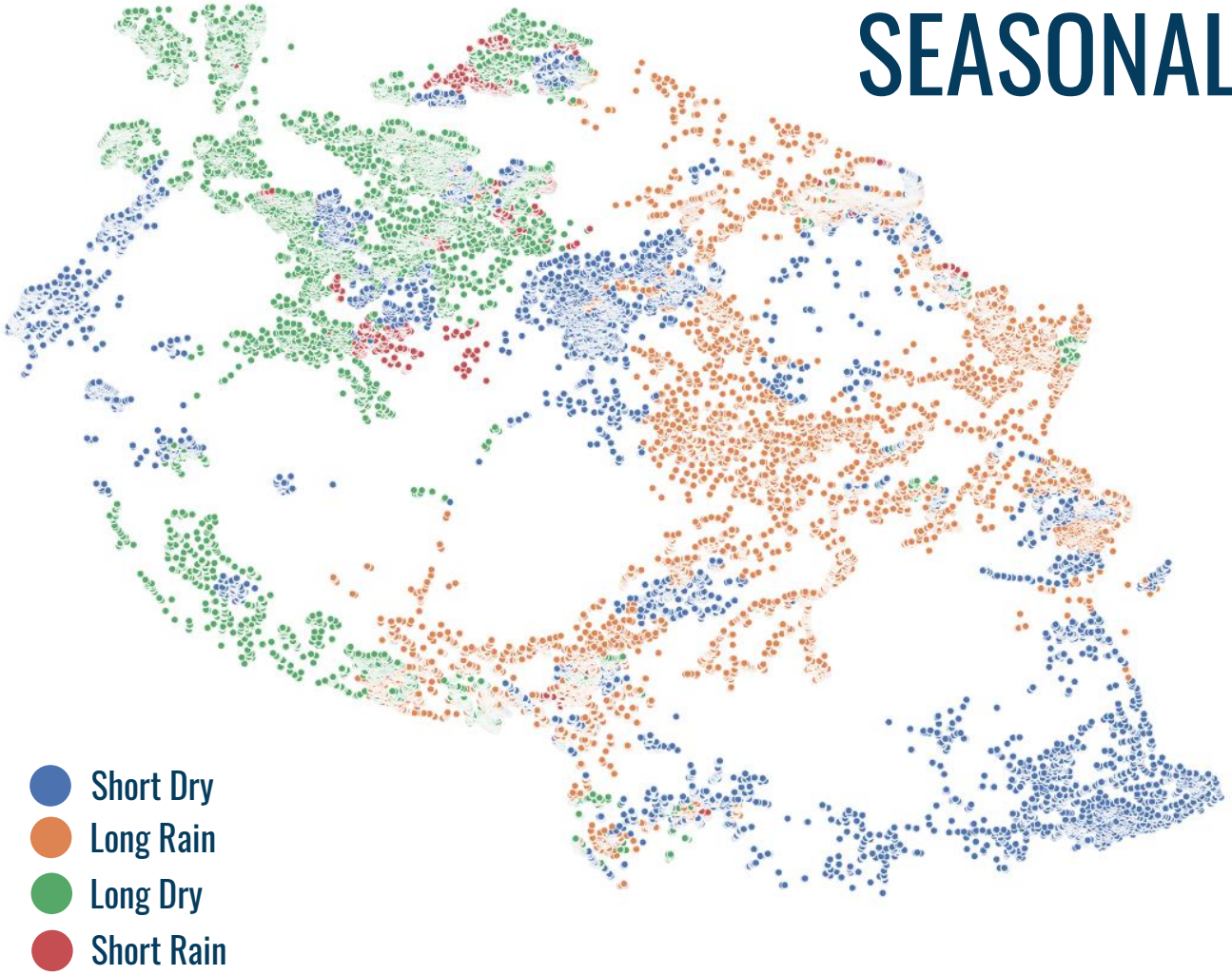


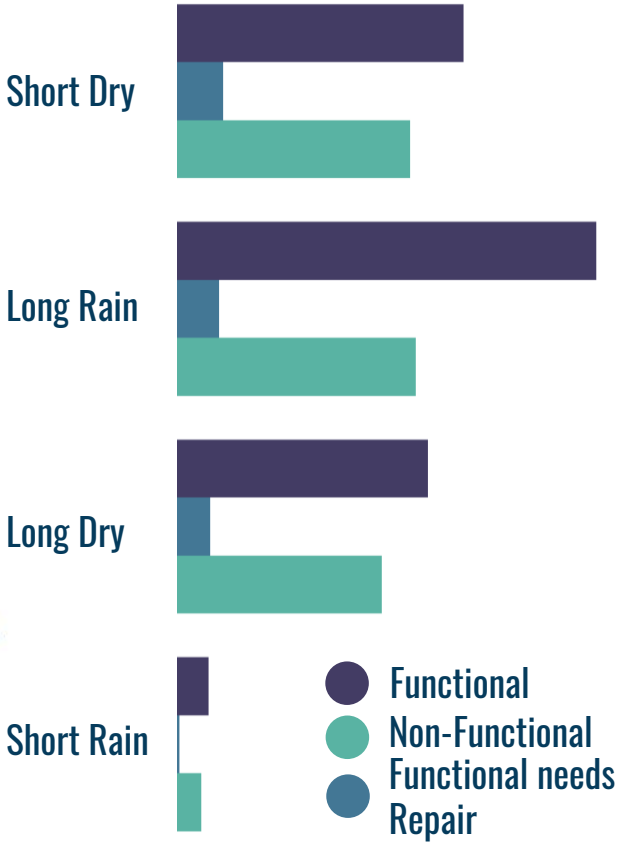


CONSTRUCTION YEAR distribution

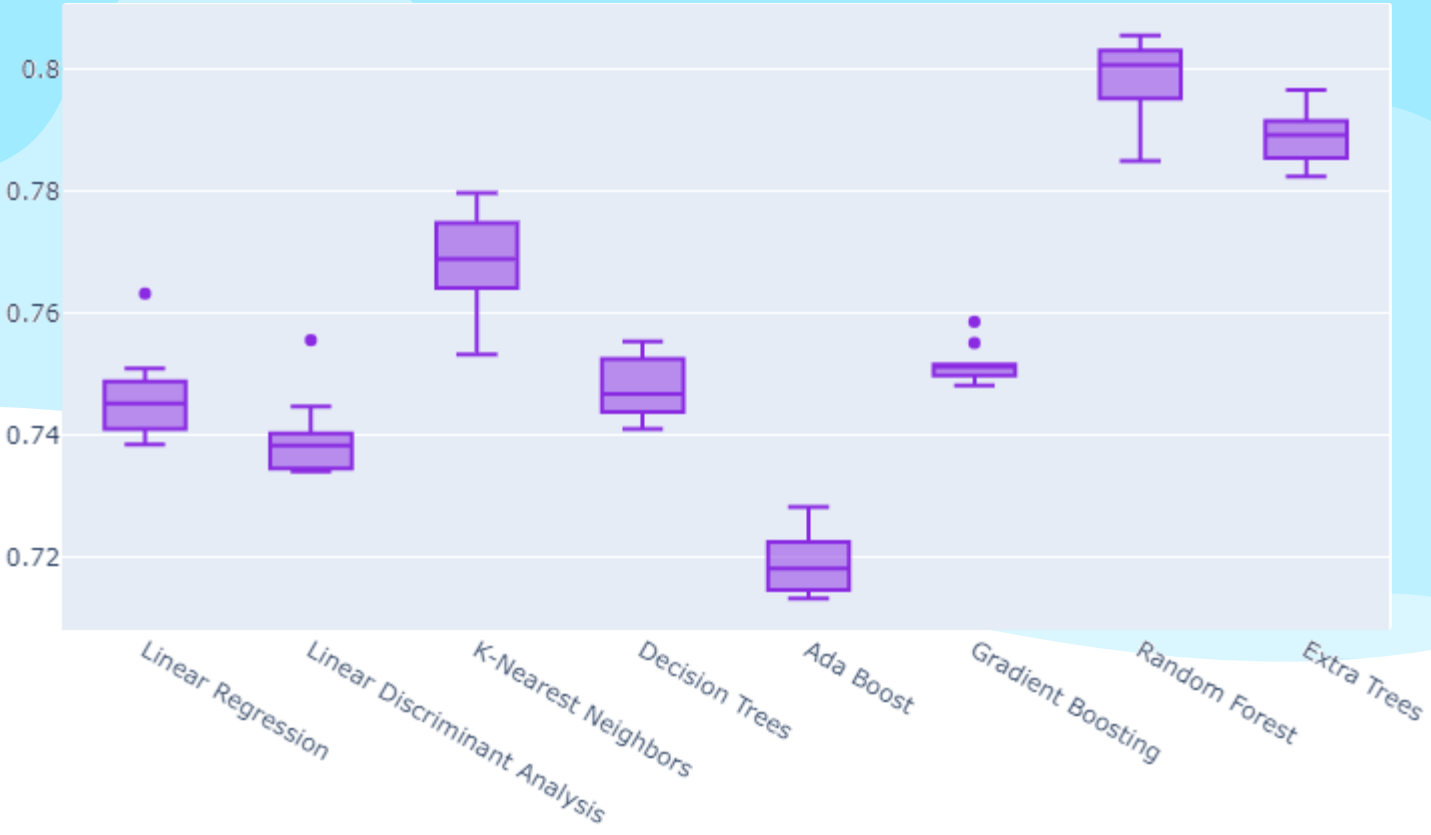


SEASONALITY distribution





OUR MODEL ACCURACY



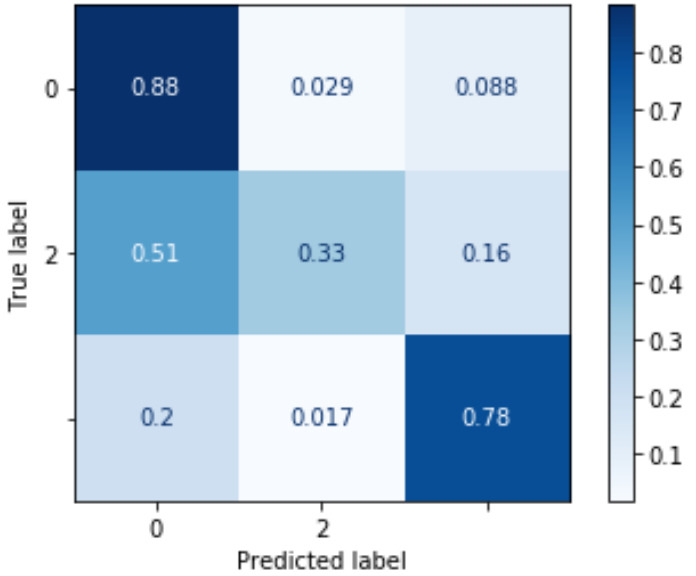
OUR RANDOM FOREST CLASSIFIER PERFORMANCE

ACCURACY

The best performing model was our Random Forest classifier which accurately predicts functionality 80 of 100 predictions

FALSE POSITIVES

On average for every 100 non-functioning waterpoints, our model would predict 20 as functioning.



OUR RECOMMENDATIONS

Gravity and Handpump wells are the simplest, and most common well type. Build teams experienced in repair and maintenance of these well types.

Maintenance Specialities

Population per well is important when keeping wear on wells low. Install more wells in those areas with higher populations to increase well longevity.

Population Use Analysis

There is a seasonal pattern to report creation. Locating repair teams in the North West regions during the long dry season, and in the East during the long rain could shorten repair times.

Seasonal Awareness

OUR FUTURE WORK

MORE DATA

Collecting more data, and increasing usage of TAARIFA would improve predictability of models

INCREASE COMPLEXITY

Neural Networks and Quantum Based Classifiers could improve predictive performances for the ministry to get clean water accessible to more citizens

COMPUTER VISION

Investment in Drones with Computer Vision models designed to find visual patterns of functionality could greatly increase predictive performance

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**.

Please keep this slide for attribution.

Thanks!

Do you have any questions?

APPENDIX

Code:

- https://www.dataquest.io/blog/settingwithcopywarning/
- https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/
- https://scikit-learn.org/stable/modules/compose.html
- https://www.kaggle.com/pouryaayria/a-complete-ml-pipeline-tutorial-acu-86
- https://www.kdnuggets.com/2017/03/simple-xgboost-tutorial-iris-dataset.html
- https://towardsdatascience.com/fine-tuning-a-classifier-in-scikit-learn-66e048c21e65Landing page template ecology landscape
- People saving the planet ecology concept

Tanzania:

- https://water.org/our-impact/tanzania/
- https://www.maji.go.tz/pages/mission-statement
- https://www.maji.go.tz/uploads/publications/sw1580706373-2nd%20%20Call%20for%20Abstract%20&%20Concept%20Note_2020%20Maji%20Week%20Scientific%20Conference.pdf
- https://www.iaea.org/sites/default/files/documents/tc/tanzania.pdf