

Phase 3 Project

Leah Pope

Flatiron Cohort: 81720-ONL-DS-FT

11/27/2020

ORGANIZATIONAL PROBLEM

Persona: Tanzania Water Authority

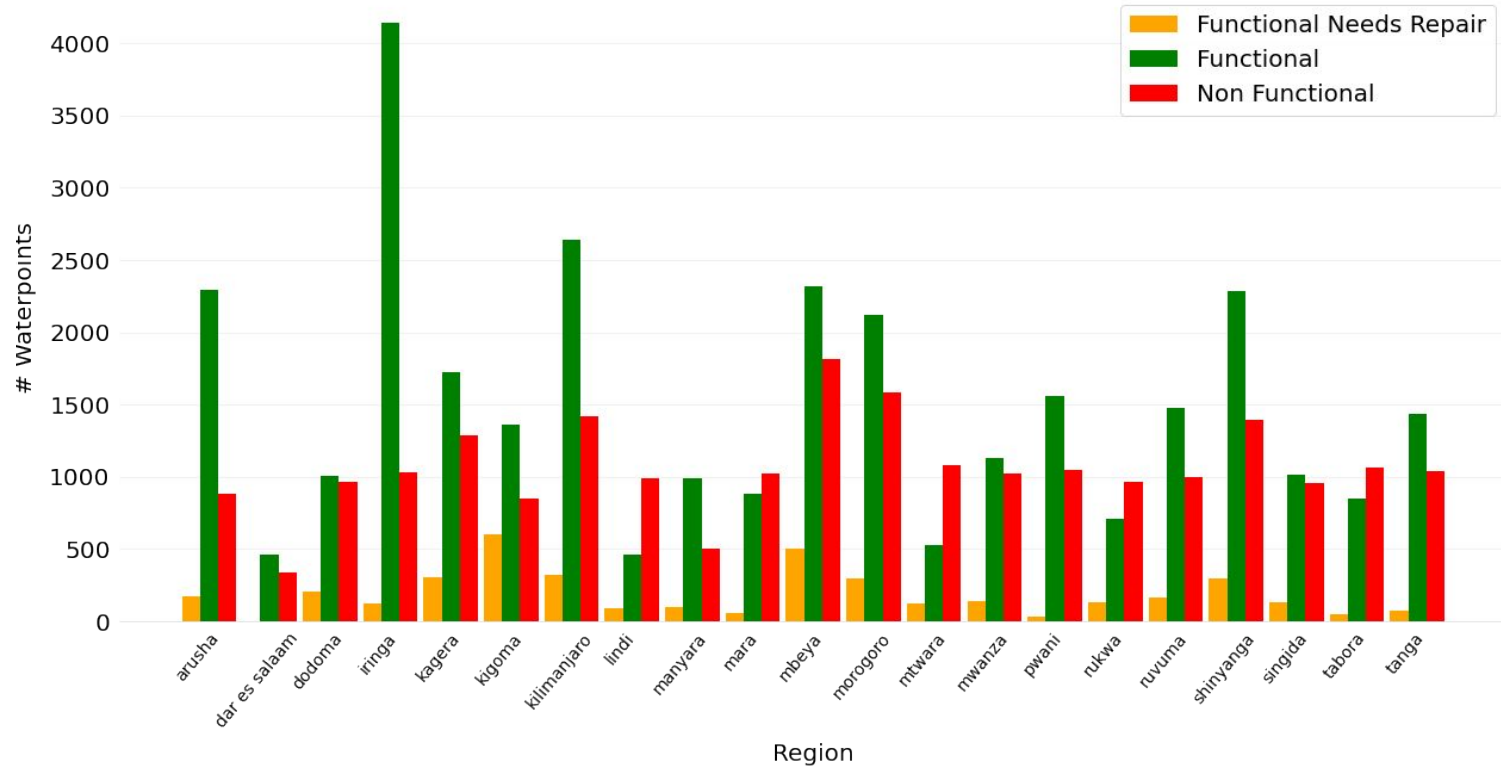
Needs: *"We want a dashboard reporting the status of waterpoints through Tanzania. We also want to know if we can predict waterpoint operational status from current data."*



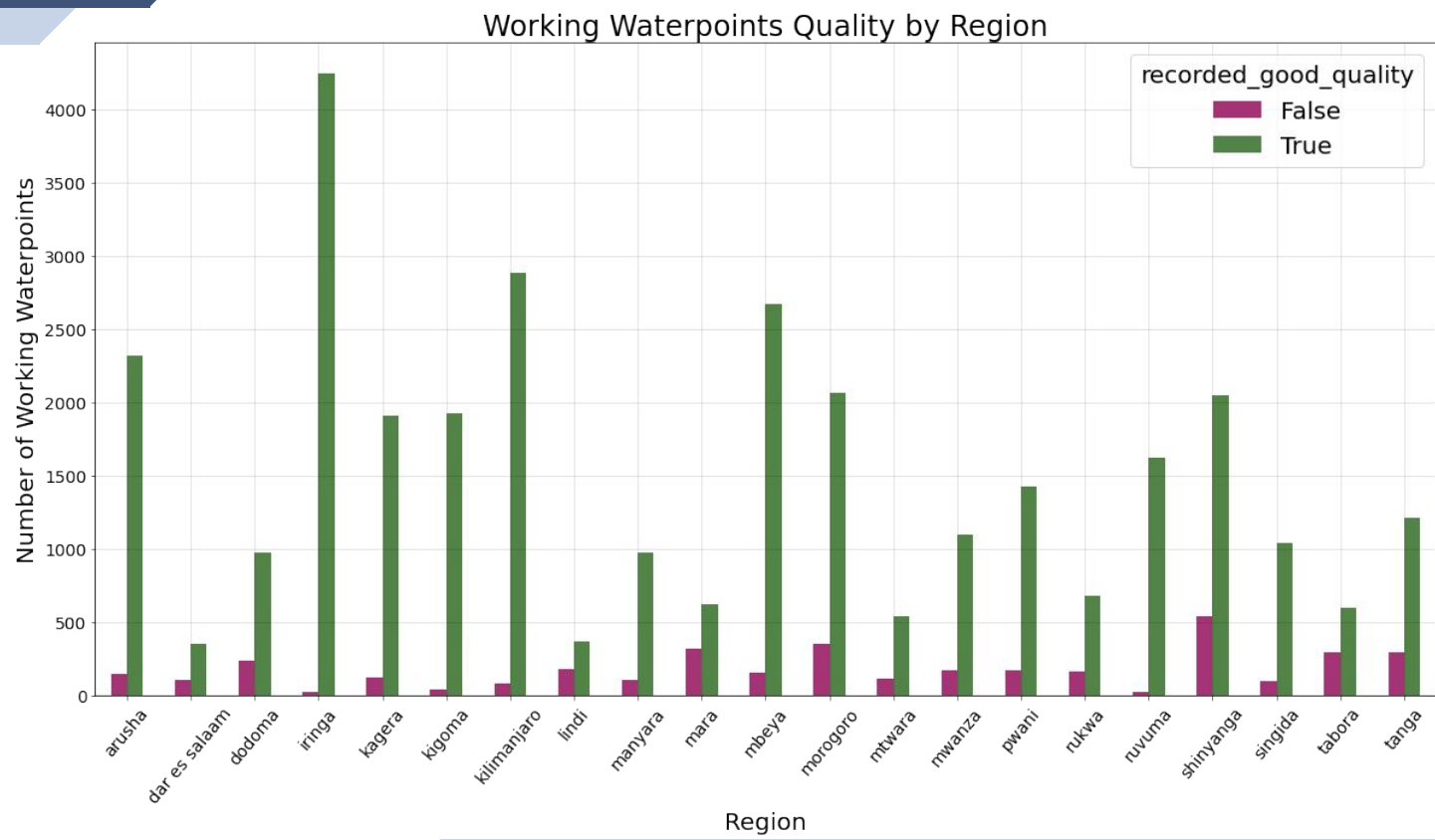
1

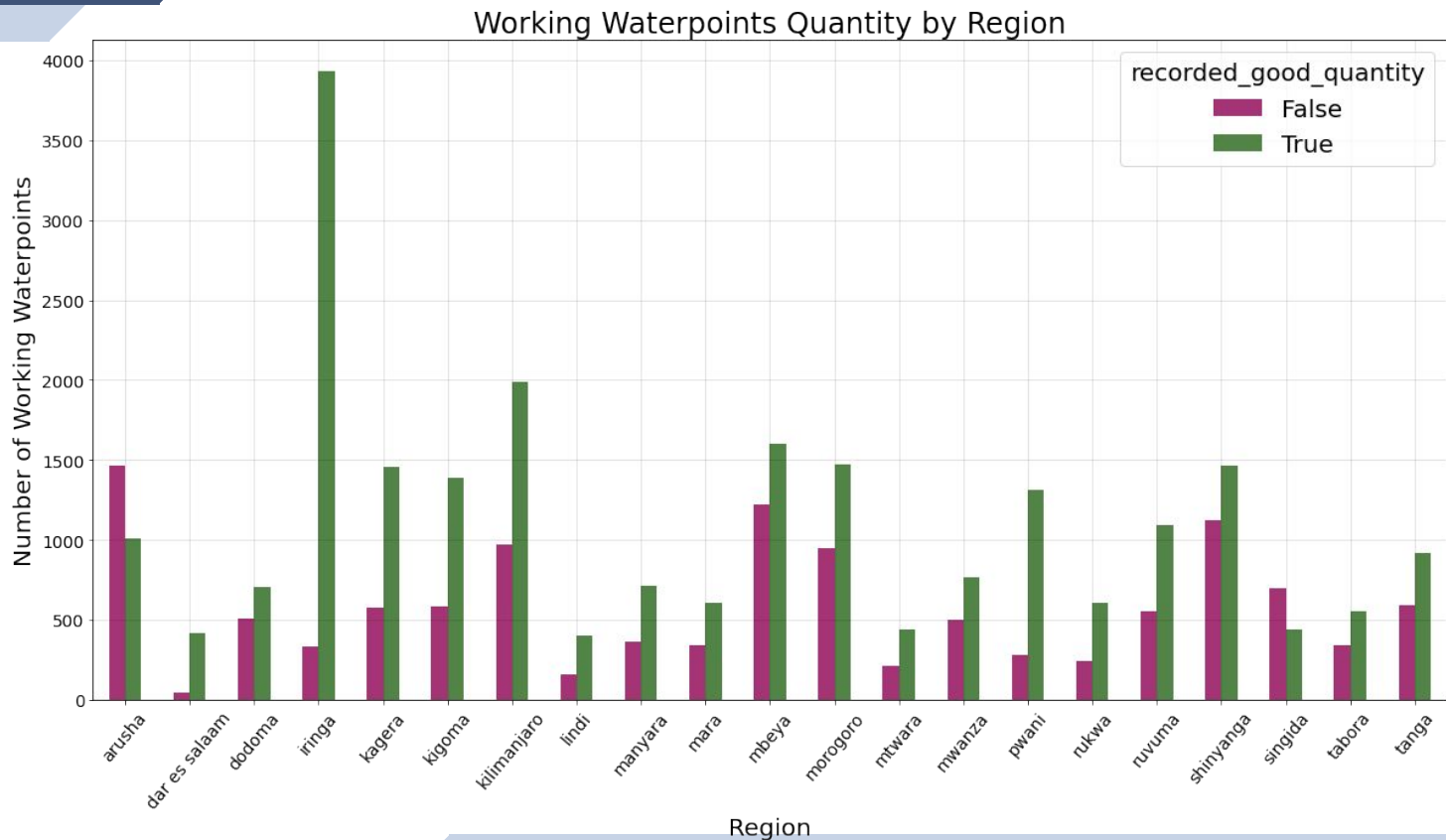
EXPLORATORY DATA ANALYSIS

Waterpoint Operational Status by Region



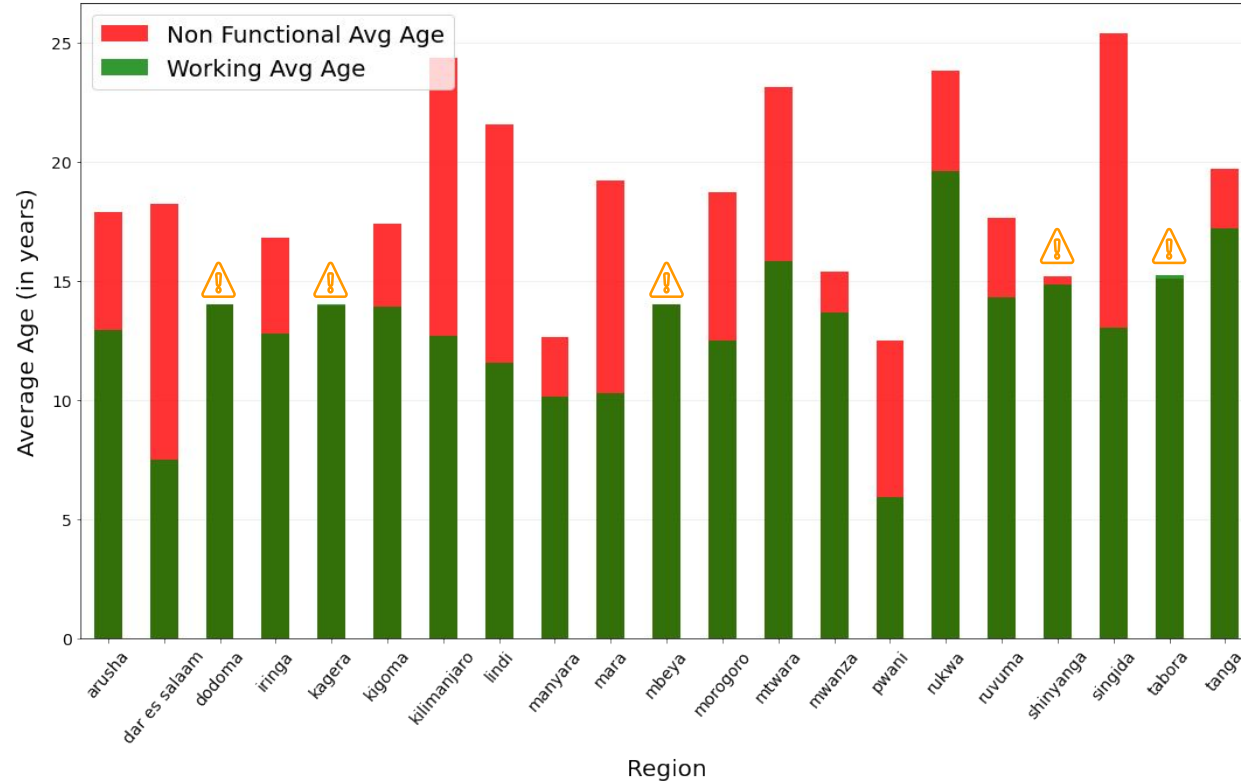
57,565 waterpoints - 61.3% Operational, 38.7% Not Operational





Age Comparison

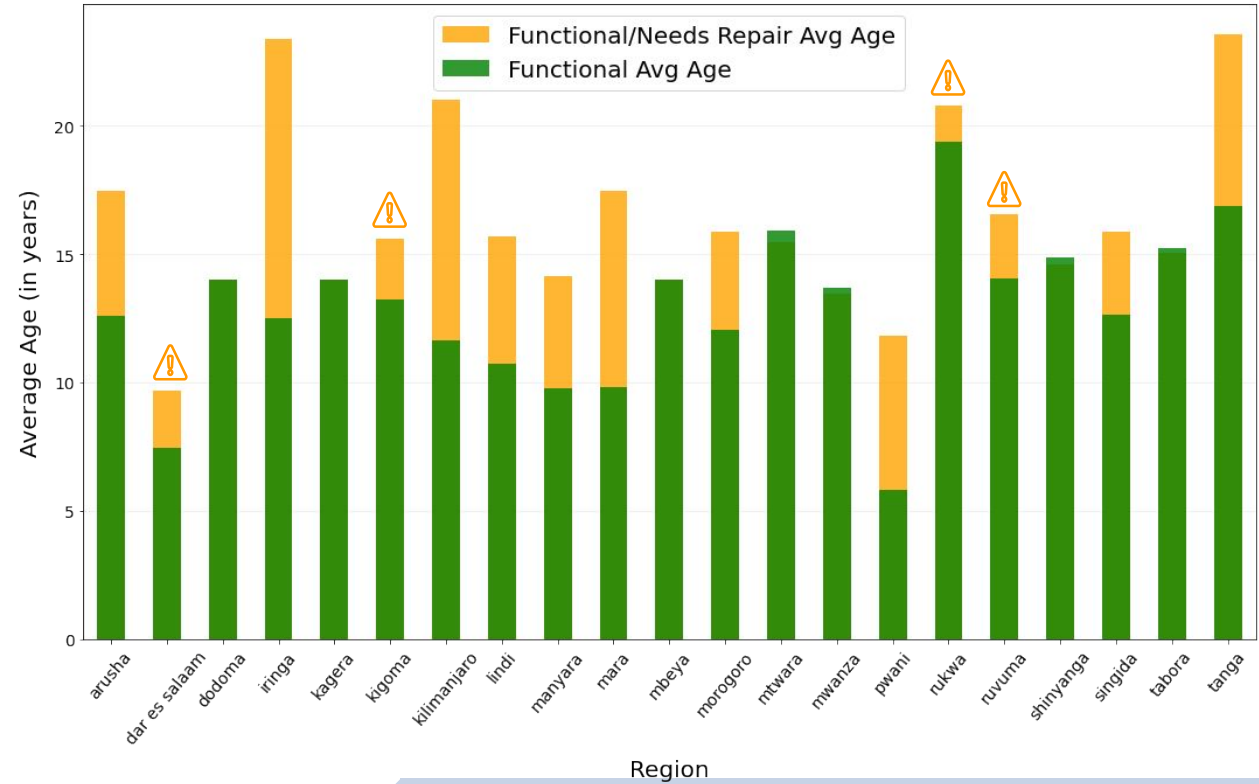
Comparing Average Age of Non Functional Waterpoints to Working Waterpoints by Region



Older Waterpoints are more likely to be Non Functional

Age Comparison

Comparing Average Age of Functional/Needs Repair Waterpoints to Functional Waterpoints by Region



Waterpoints in need of Repair tend to be older than waterpoints with no repair needs

2

CLASSIFICATION

Classification Modeling

- **3** Total Algorithms explored
 - ▷ Random Forest, Gradient Boosting, XGBoost
- **15** Features Used (out of 44)
- **Accuracy** for Model Evaluation Metric
- Model Face Off!
 - ▷ Tuned Gradient Boosting vs Tuned XGBoost
 - ▷ Overall Accuracy ~79%. Which is "better"?
 - ▷ **Tuned Gradient Boosting**





THANKS!

Questions? You can find me at

leah@metisconsultingllc.com

<https://www.linkedin.com/in/leahspope/>

<https://github.com/lspope/dsc-phase-3-project-online>

APPENDIX

FUTURE WORK

Water Availability

- How many people use waterpoints as primary water source? By region, district, ward, etc.
- For lower Quality/Quantity waterpoints, what is the size of the population served. Where are these waterpoints located (region, district, ward)
- What do we know about the payment types for waterpoints? Breakdown by country, region water basin.

Working Waterpoints Analysis

- Who are the major installers?
- Who are the major management groups?
- Who are the major scheme managers?
- What are the major water sources?

CREDITS

Special thanks go to:

- Flatiron Cohort Members and Instructor
- Feature Engineering for ML Course
- DrivenData (Challenge hosts)
- Friends and Family



PRESENTATION CREDITS

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by [SlidesCarnival](#)
- Photographs by [Pexels](#)



THANKS!

Leah Pope

Full Time Flatiron Data Science Student

leah@metisconsultingllc.com

<https://www.linkedin.com/in/leahspope/>

<https://github.com/lspope/>