Tanzania Waterpoints

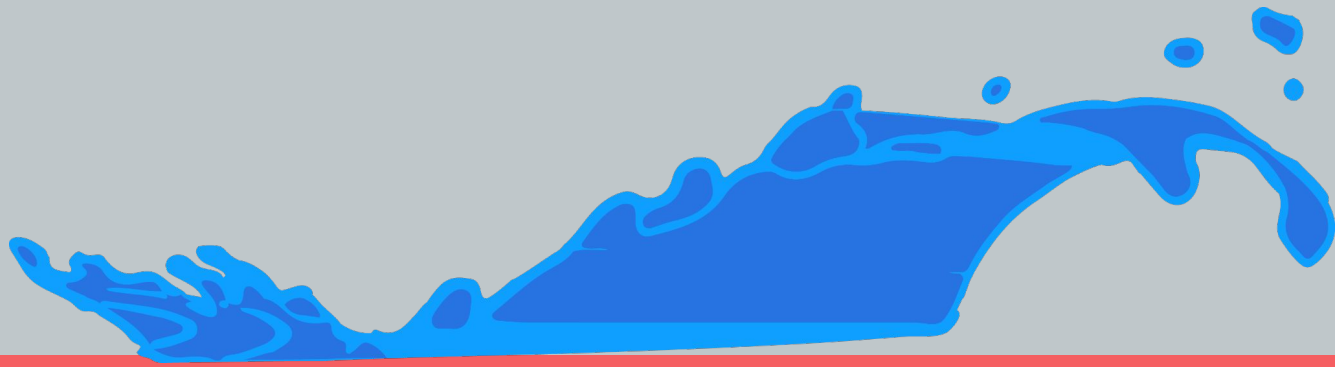
Data Analysis and Model

By: Mohammed Siddiqui
April 23, 2021

SUMMARY

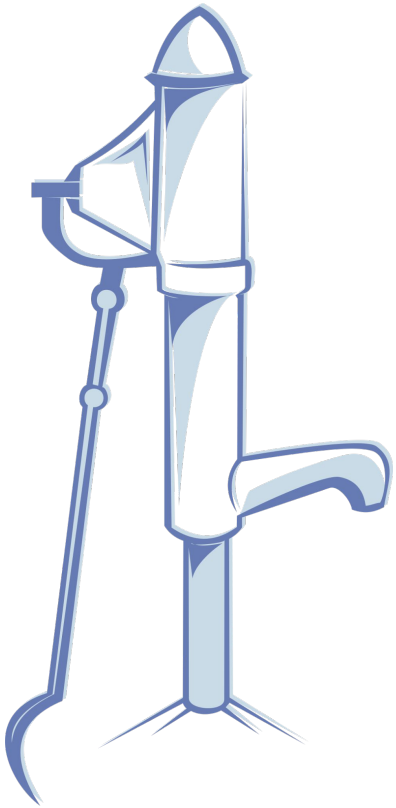
Descriptive analysis and modeling reveal which factors contribute most to repair needs:

- Water Quantity
- Payment
- Region and Local Government Authority



OUTLINE

- Business Problem
- Data and Methods
- Results
- Conclusions



BUSINESS PROBLEM

- WaterAid would like us to find areas needing the most help
- Create model to predict whether a waterpoint needs repair
- Identify factors which contribute to non-functionality

DATA & METHODS

Data

- Tanzania Waterpoint Mapping data from 2012
- Source: datadriven.org

Methods

- Machine Learning
- Classification

RESULTS

Prediction Accuracy

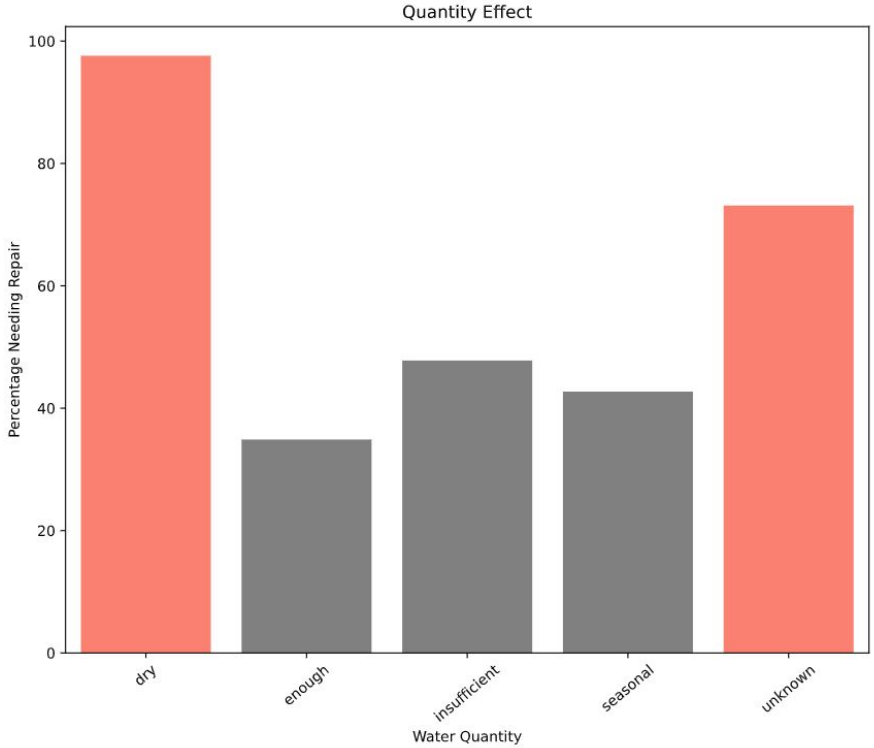
- Our model was able to correctly predict the status of a waterpoint 80.6% of the time.

Accuracy

80.6%

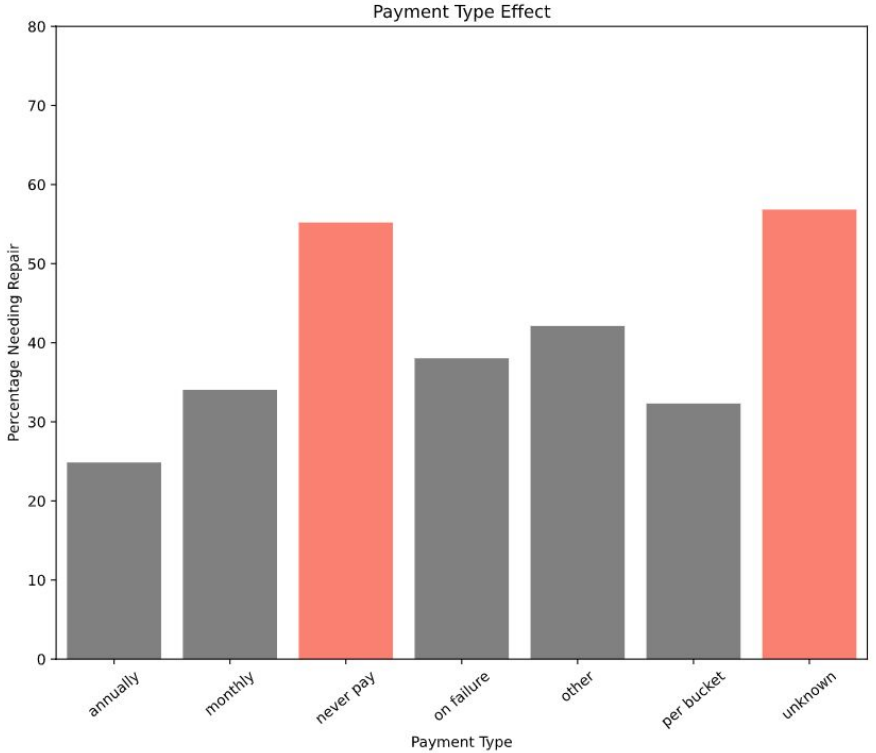
RESULTS

- Quantity of water plays a significant role.
- Waterpoints in ‘dry’ and ‘unknown’ areas need the most repair.
- Those areas should get priority



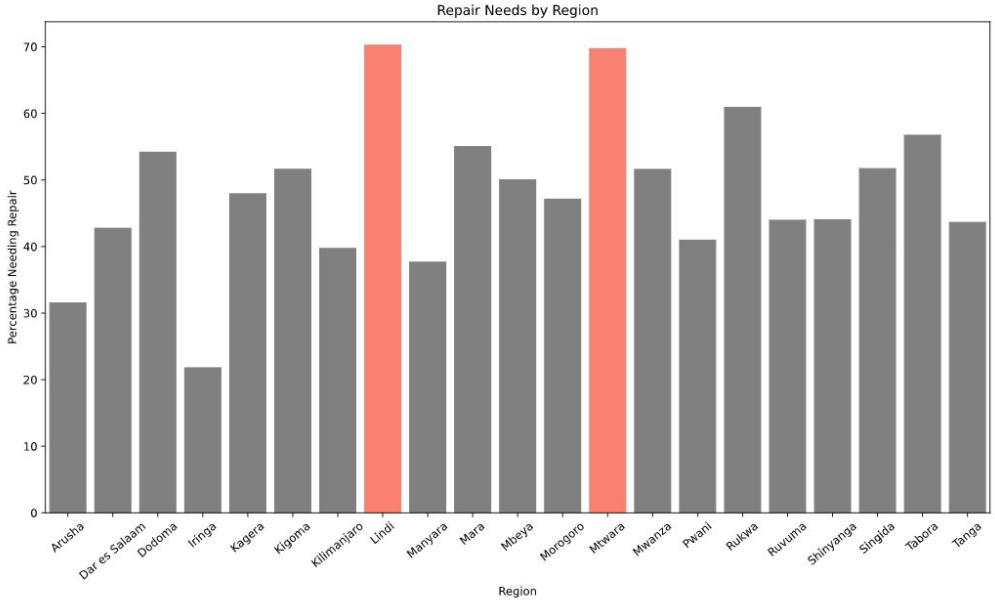
RESULTS

- Waterpoints that receive regular payments tend to stay more functional.
- Higher repair needs for ‘never pay’ and ‘unknown’



RESULTS

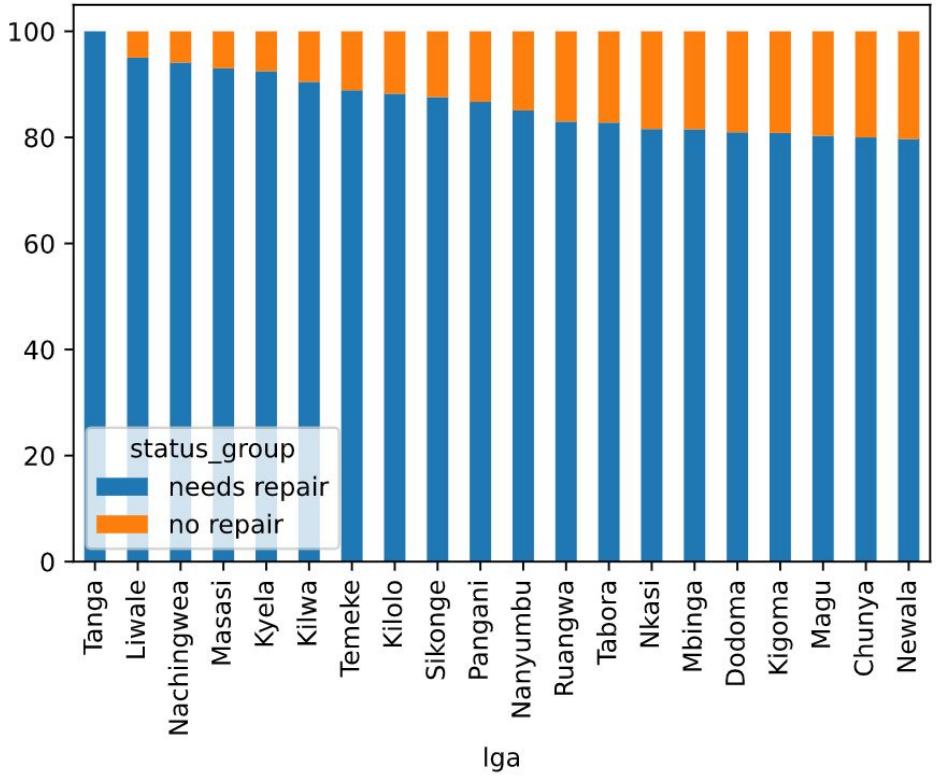
- There are two regions of particular interest here.
- Lindi and Mtwara have the highest needs.
- Both are located in the southeast right next to each other.



MAP FOR REFERENCE

RESULTS

- Within areas of not enough water, these LGAs are most likely to need repair.
- These 20 have over 80% of their waterpoints in need of repair.



CONCLUSION

FACTORS AFFECTING THE NEED FOR REPAIR

- Quantity of Water
- Payment
- Region and LGA

NEXT STEPS

- Deal with missing data on latitude, longitude and elevation
- Differentiate between rural and urban waterpoints
- Obtain data on maintenance

Thank You

Email: siddiqui@mohammed@gmail.com

GitHub: @saifword