Tanzania Water Wells

Emma Choate

Outline

- Business Problem
- Data
- Methods
- Results
- Conclusion

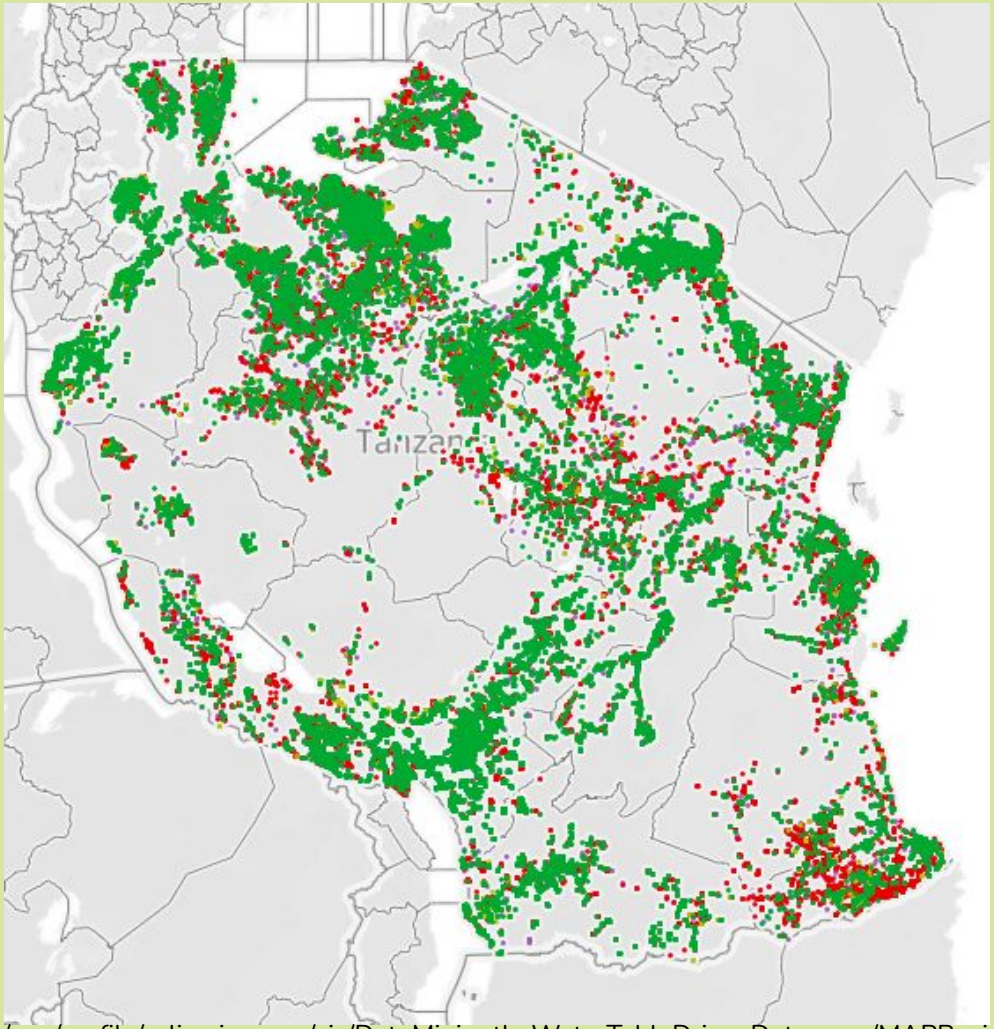
Business Problem

4 million people in Tanzania lack access to an improved source of safe water, and 30 million don't have access to improved sanitation. With the data that I was given, I was able to help this problem by predicting whether or not a well was functional, functional needs repair, or non functional.

Data

I used data from the DataDriven Competition on the water wells in Tanzania. The data gave information about the wells such as:

- 59,400 rows of data
- 41 columns
- Status of the well (target)
- Quality and quantity of the water
- Longitude and latitude of the well
- Population around the well
- etc.



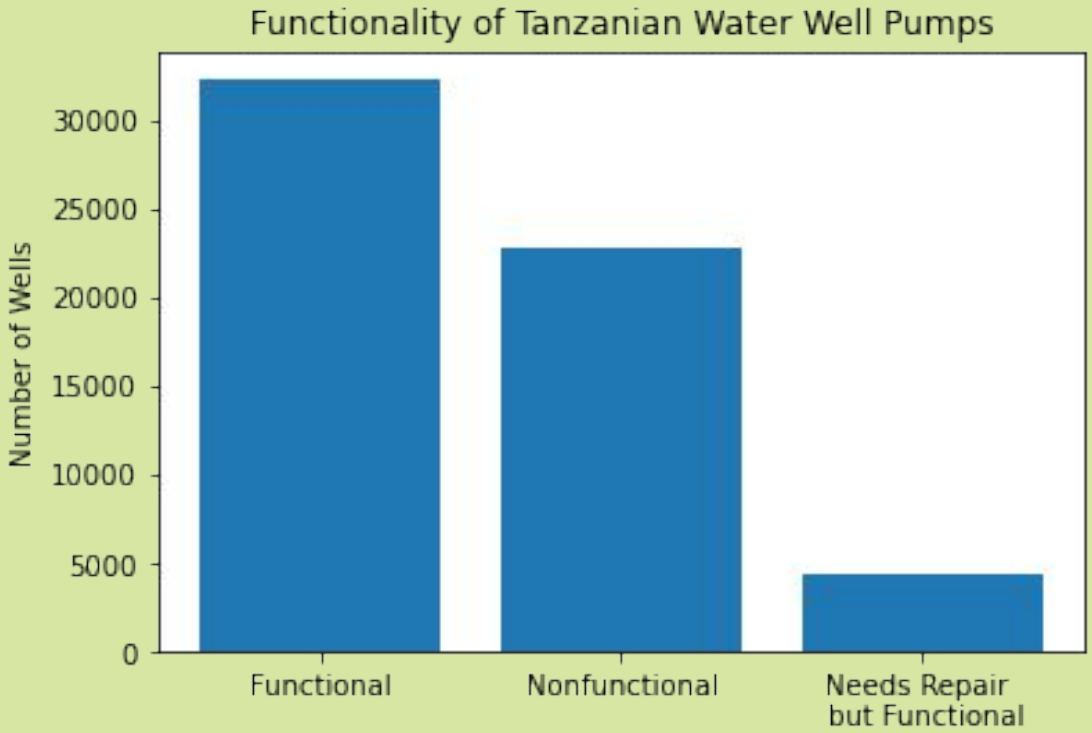
https://public.tableau.com/app/profile/calin.uioreanu/viz/DataMiningtheWaterTableDrivenData_com/MAPBasin

Methods

- Imported necessary packages
- Exploratory data analysis
- Cleaned data
- Created a baseline model
- Created multiple complex models using different modeling techniques to make the best prediction

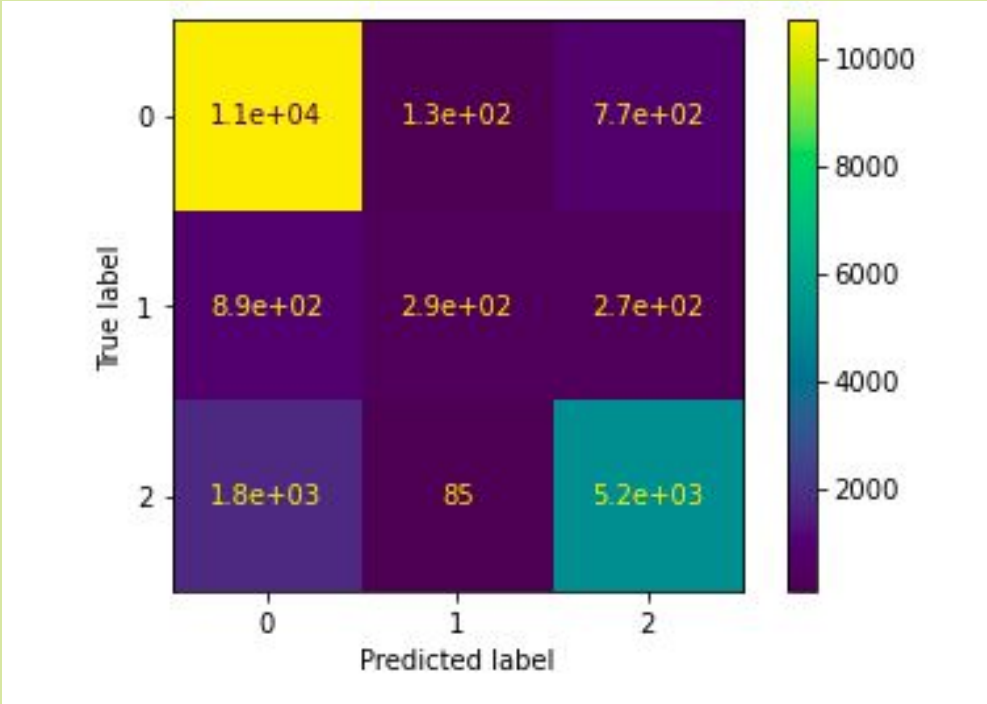
Baseline Model

- Features Used: region, source type, quantity, quality, longitude, latitude
- **Training Accuracy: 0.689**
- Interpretable yet not very predictive



Complex Model

- Used all features in the sorted training set
- **Validation Accuracy: 0.785**
- Predictable yet also relatively interpretable



Conclusions & Future Directions

- A model, with 78% accuracy, that can predict the functionality of a well
- My model can be used in other countries with similar problems
- Given more time and a stronger computer, I would continue to improve my models so they are not so overfit

Thank You!

Email: emmachoate11@gmail.com

GitHub: @EmmaChoate