Tanzania Water

wells

Arturo Delgado



Which water wells should we focus on the most?

My goal was to figure out which water wells to focus on based on region.



Data and Tools used.

Data Set details

Data set collected from DrivenData

- Total of of 41 features
- 59K observations
- Target variable consist of Functional, Non-Functional, Functional needs repair.
- 3000 na in public_meeting

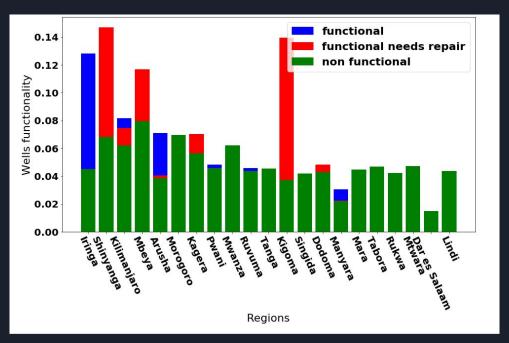
Tools

- Python
- Libraries
 - Sklearn
 - Pandas
 - o **numpy**

Regions and well Functionality

Are in percentages.

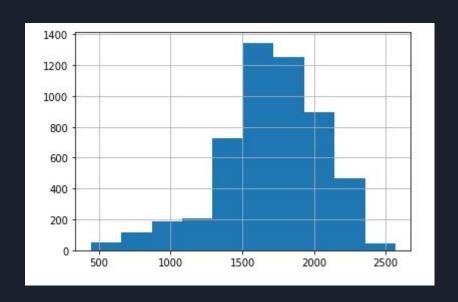
Certain regions contribute more into non functional wells, functional, functional needs repair.



Iringa Gps Height

X is Iringas wells

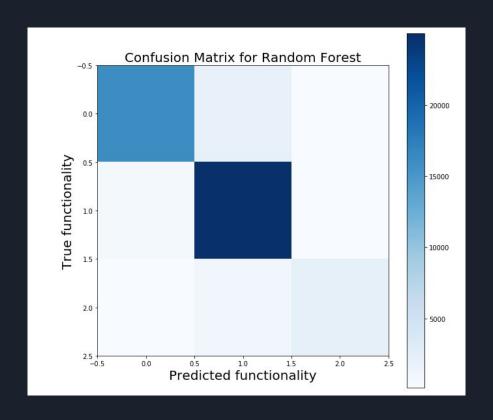
Y GPS Height



Random Forest Model

Random Forest top 5 Feature importance:

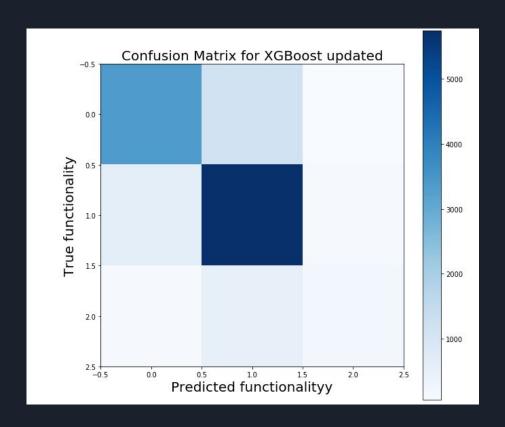
- population
- public_meeting
- Gps_height
- waterpoint_type_hand pump
- -quality_group_salty



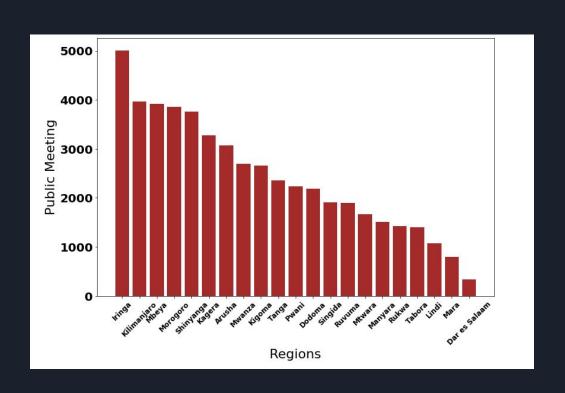
XGBoost Model

Top 5 Feature importances were:

- -waterpoint_type_hand pump
- -quality_group_salty
- population
- -public_meeting
- -gps_height



Regions and public Meetings



More to look into

- Use SMOTE to artificially create data for functional wells needs repair
- Continue to hyper parameter tune
- Explore gps height and each region
- Explore population in each Region
- Water Quality in each Region

THANK YOU

LinkedIn : Arturo Delgado

http://linkedin.com/in.arturo-delga do-ba94a1106