

PUMP IT UP: DATA MINING THE WATER TABLE

KAGGLE.COM/DATASETS/SUMEETSAWANT/PUMP-IT-UP-CHALLENGE-DRIVEN-DATA

STAKEHOLDERS

WELLS / WATER POINTS / WATER PUMPS

Central government

Regional & District governments

Local governments & Towns/Villages

Citizens/Villagers

DATASET

(59400, 40)

Water point funding organization

Water point installation entity & year built

Water point operating entity

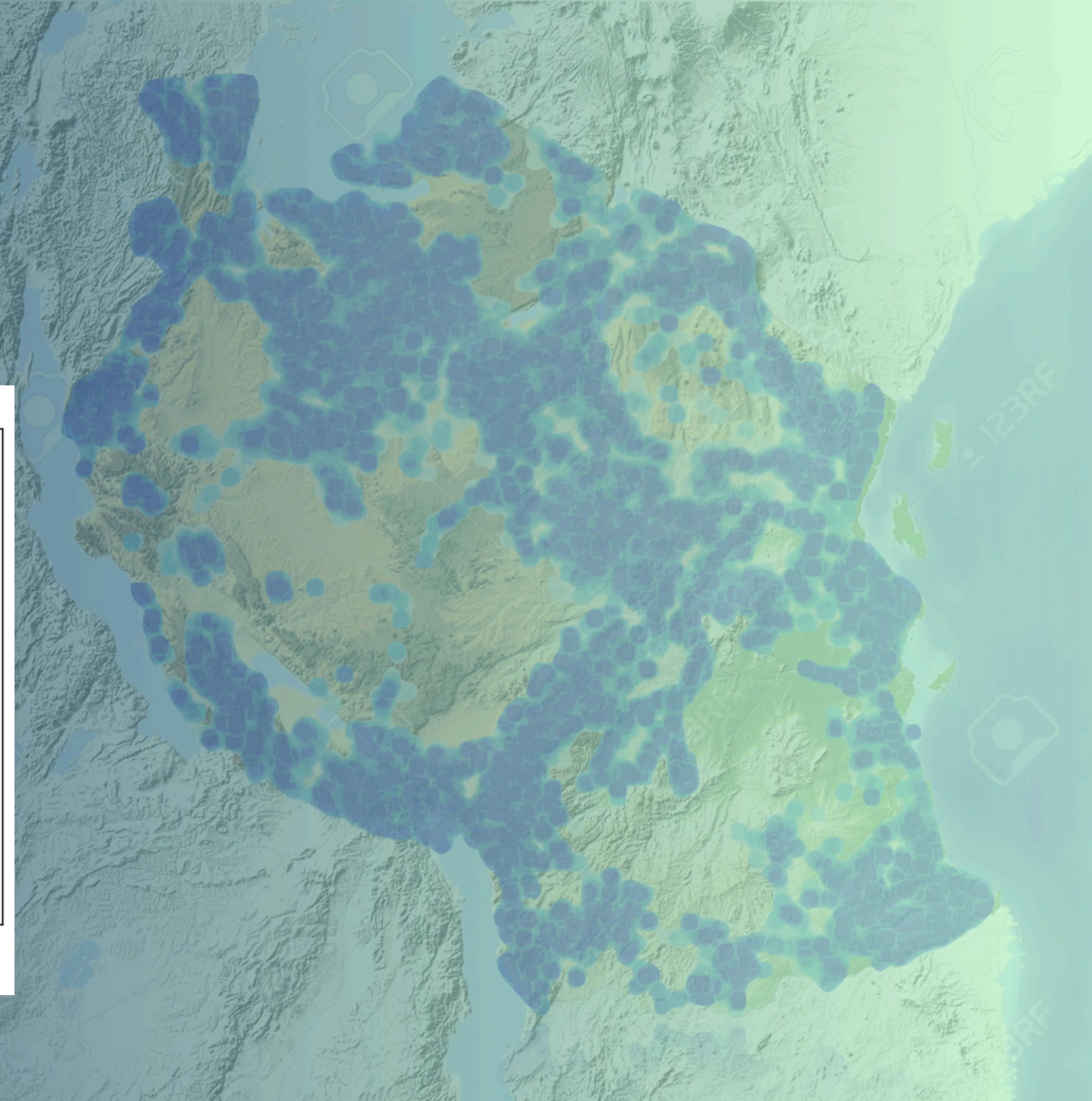
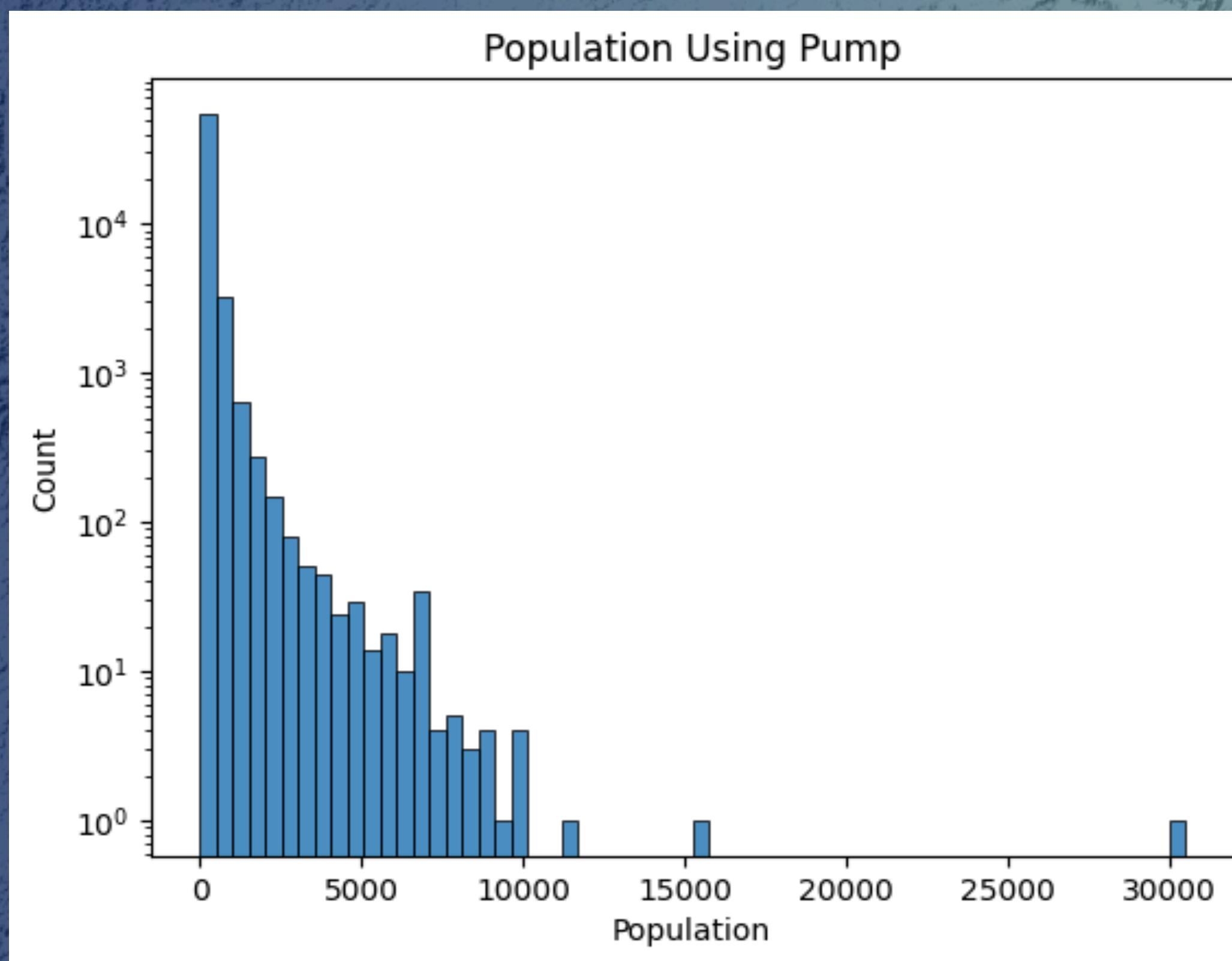
Water point location information

Water point type & water source

Water quality information

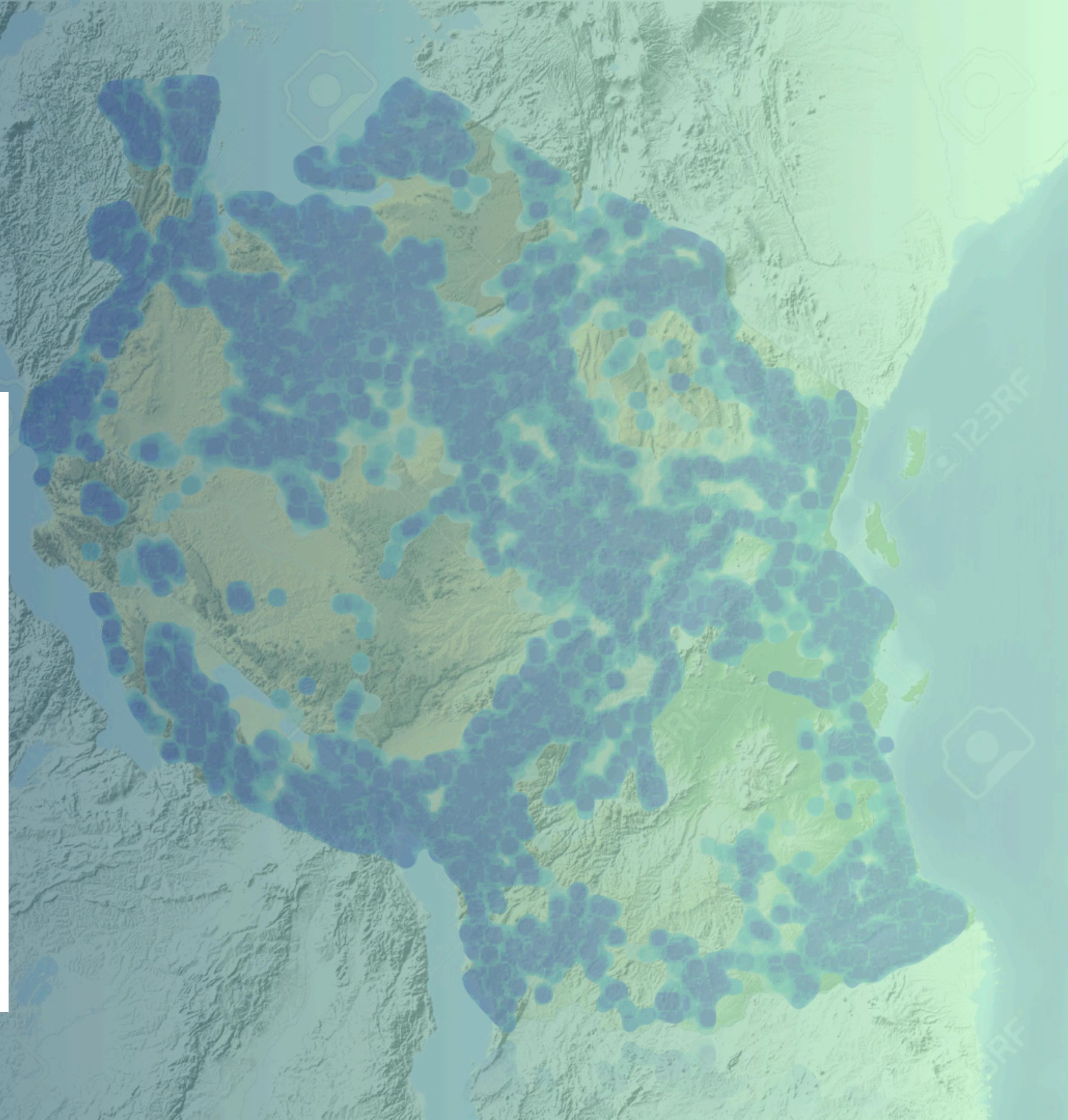
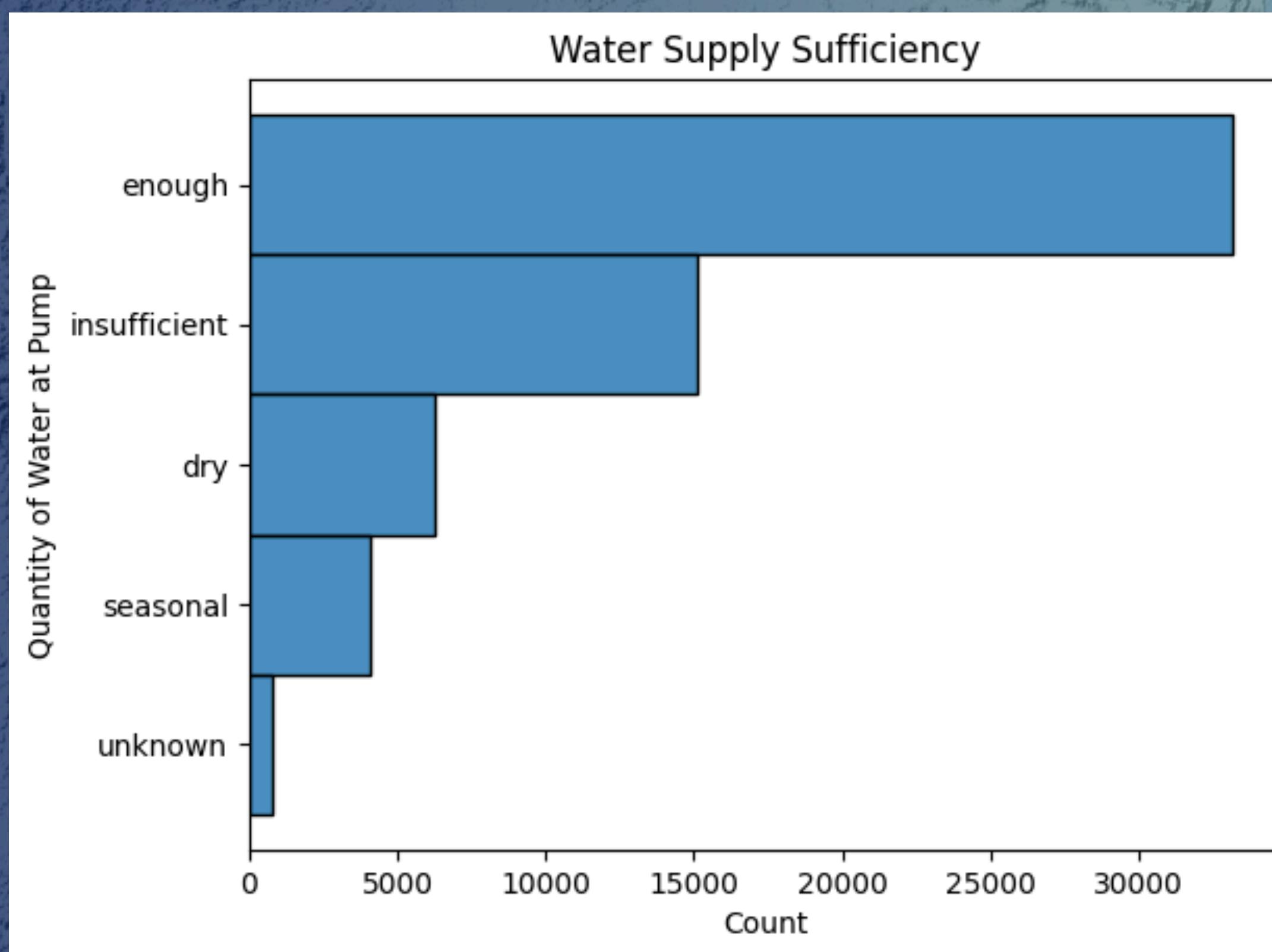
DATASET

VISUAL ONE



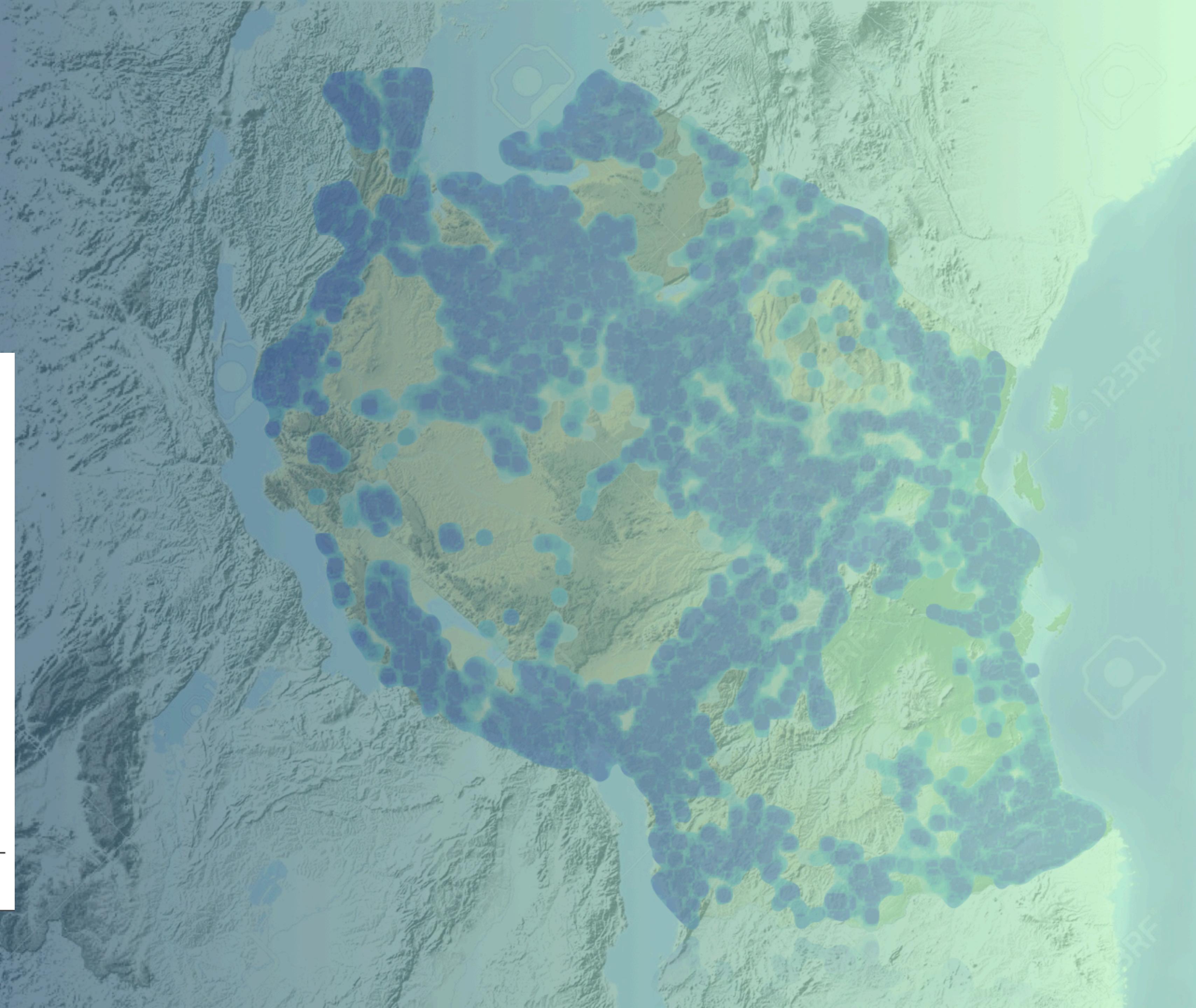
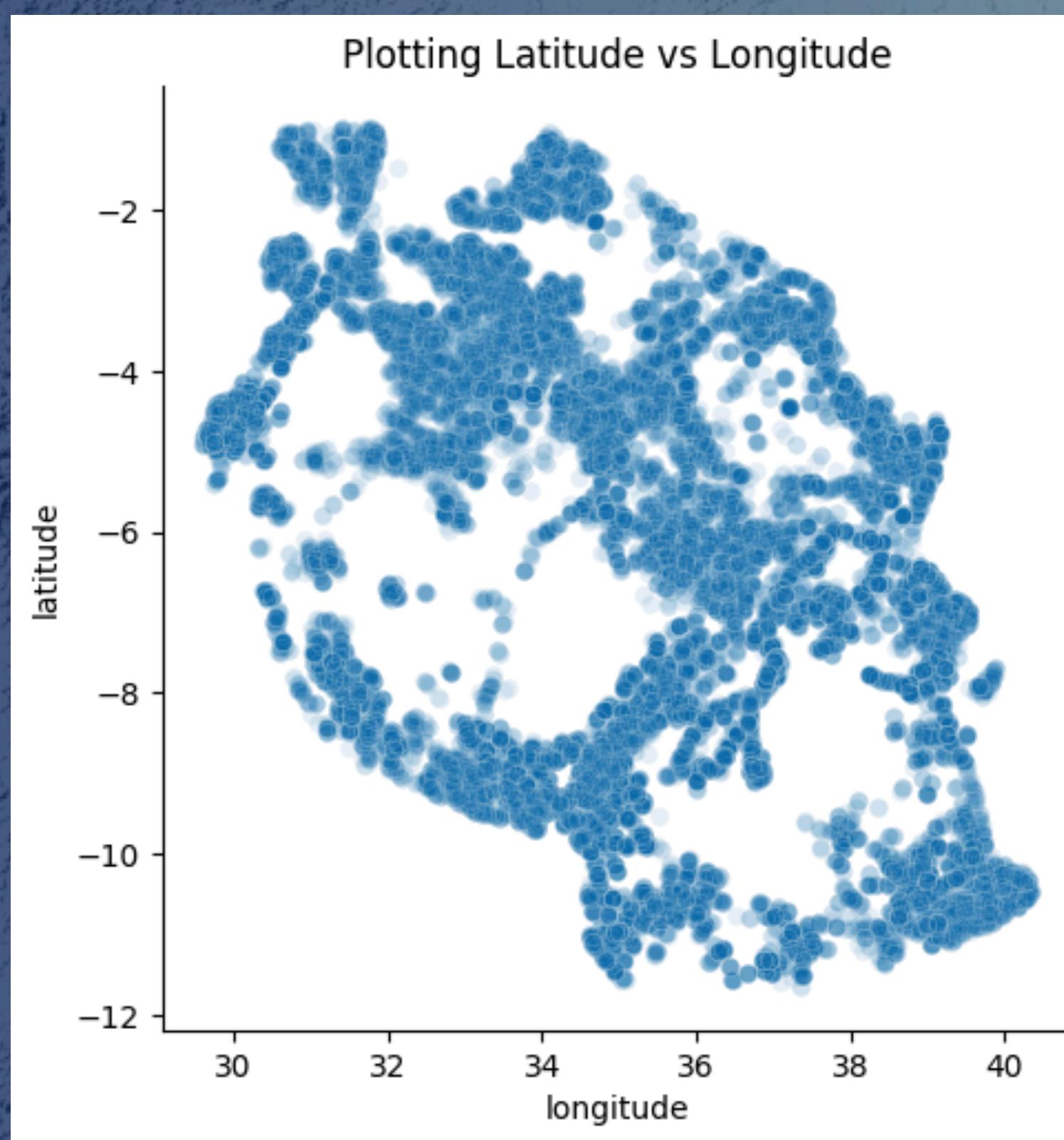
DATASET

VISUAL TWO



DATASET

VISUAL THREE



STRENGTHS

XGBOOST - ACCURACY: 74% & RECALL: 57%

Lowest Number of False Negatives

[Broken wells detected & not ignored]

Highest Degree of Accuracy

[Correctly predicted status most accurately]

Performs Best in Most Important Areas/Metrics

[Best model available so far]

WEAKNESSES

XGBOOST - ACCURACY: 74% & RECALL: 57%

False Negative Rate Not High Enough

[Best result, but lots of room to improve]

Small Sample of Data Used

[More data use & processing time needed]

Not All Features Included

[More data cleaning & filling missing values needed]

CLASS BALANCE

DATA - WEIGHTS OF CLASSES

54% - Functional

40% - Non-Functional

07% - Functional, Needs Repair

Functional, Needs Repair:

Poorest prediction rates, by a large margin

FINAL RECOMMENDATIONS

DATA - USE MORE

The current analysis is an initial result.

Implement more of the dataset by alleviating:

- Spelling errors & inconsistent data
- Long processing times
- Invalid values in some features
- Errors from processing such a large amount of data