Predicting Faulty Water Pumps in Tanzania

by Kaitlyn Zeichick

Source: Tobias Hayes

Outline

Context

Data

Approach 1

Methods
Results

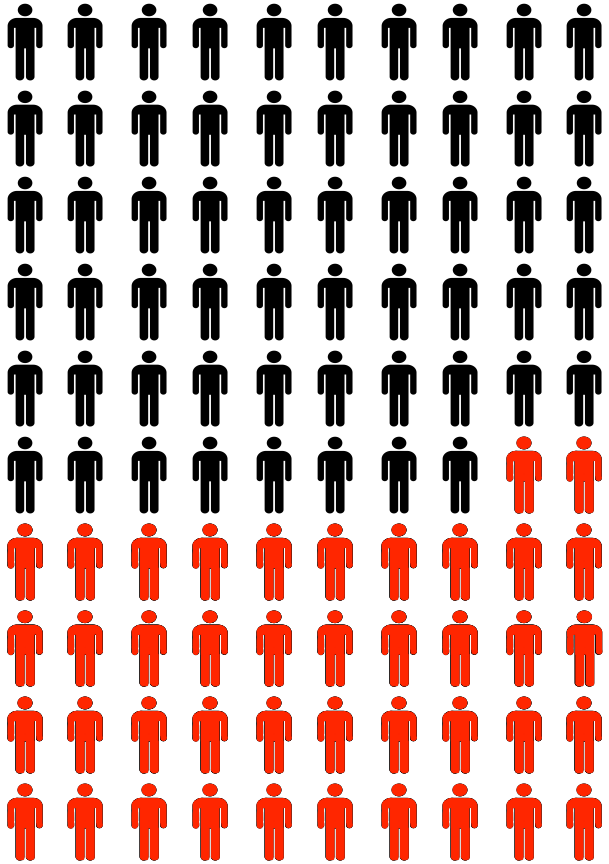
Approach 2

Methods
Results

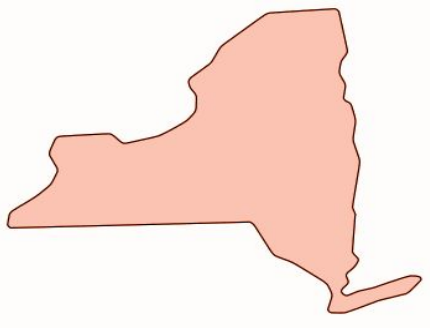
Conclusions/ Next Steps

Tanzania Water Crisis

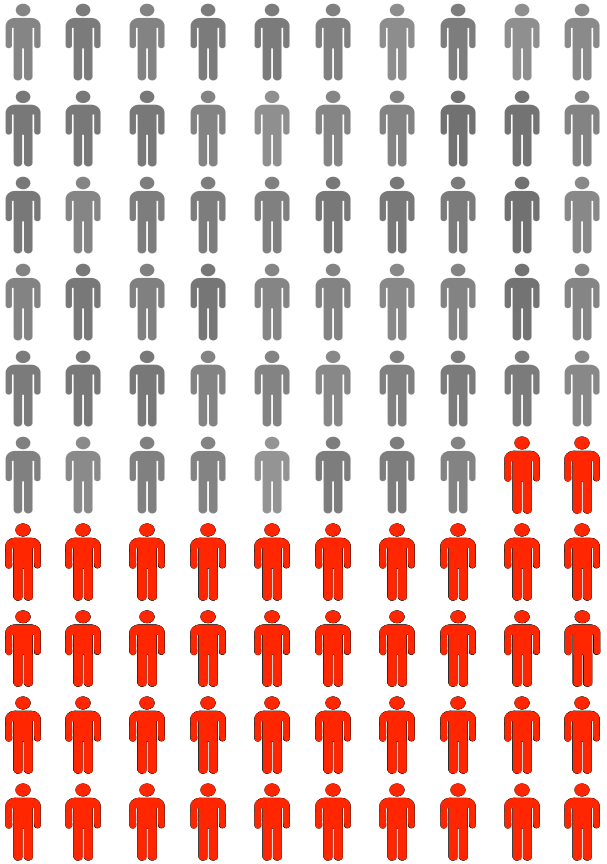
- 24 million people without basic access to safe water



Tanzania Water Crisis



=



Consequences

- Malaria
- Cholera

Solution : Water Pumps!

- Pros:
  - Clean
  - Cheap
- Cons:
  - Easily broken

Data

Water Management

Location

Water Usage

Water Related

Pump Features

Functional

Non-Functional

Two Problem Statements

1) WHERE? Locate already existing faulty water pumps.

2) WHY? Identify features that are highly correlated with faulty pumps.

Random Forest: Choosing a Metric

Consequence of many
False Negatives:

Consequence of many
False Positives:

Random Forest: Results

Accuracy

79%

Recall

80%

Logistic Regression

Why and how to water pumps break?

Logistic Regression

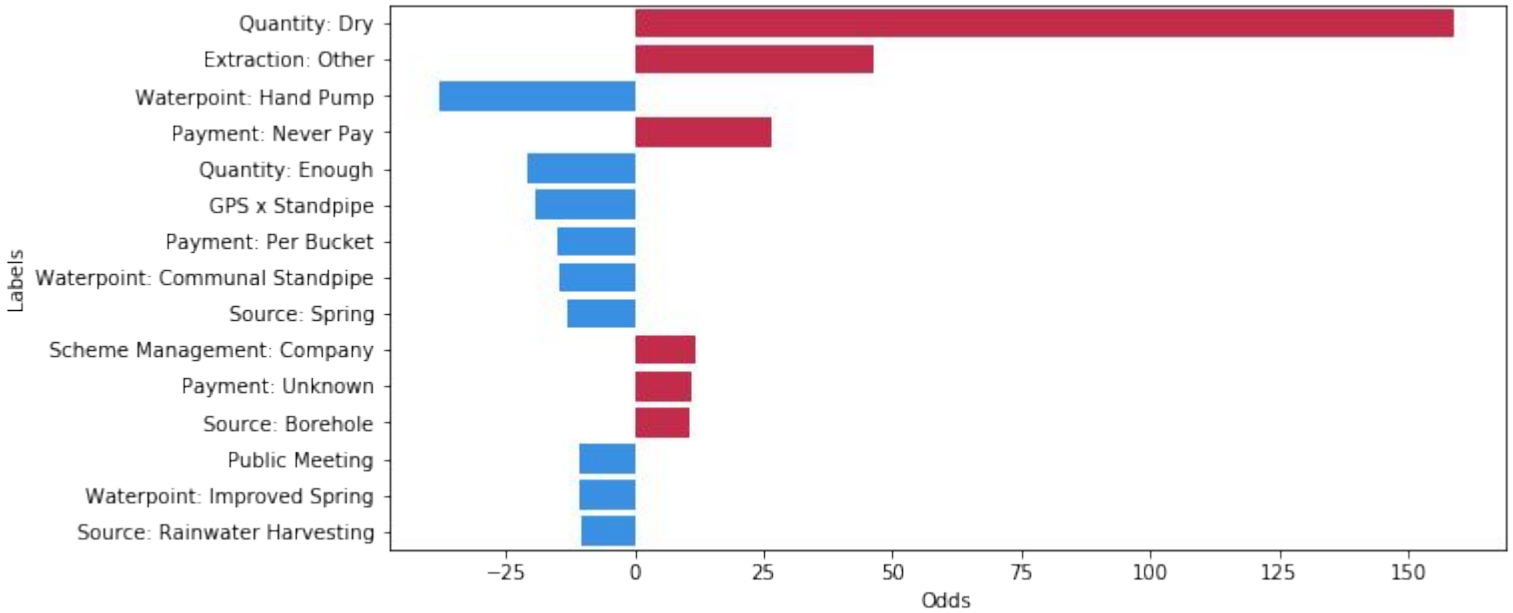
Accuracy

70%

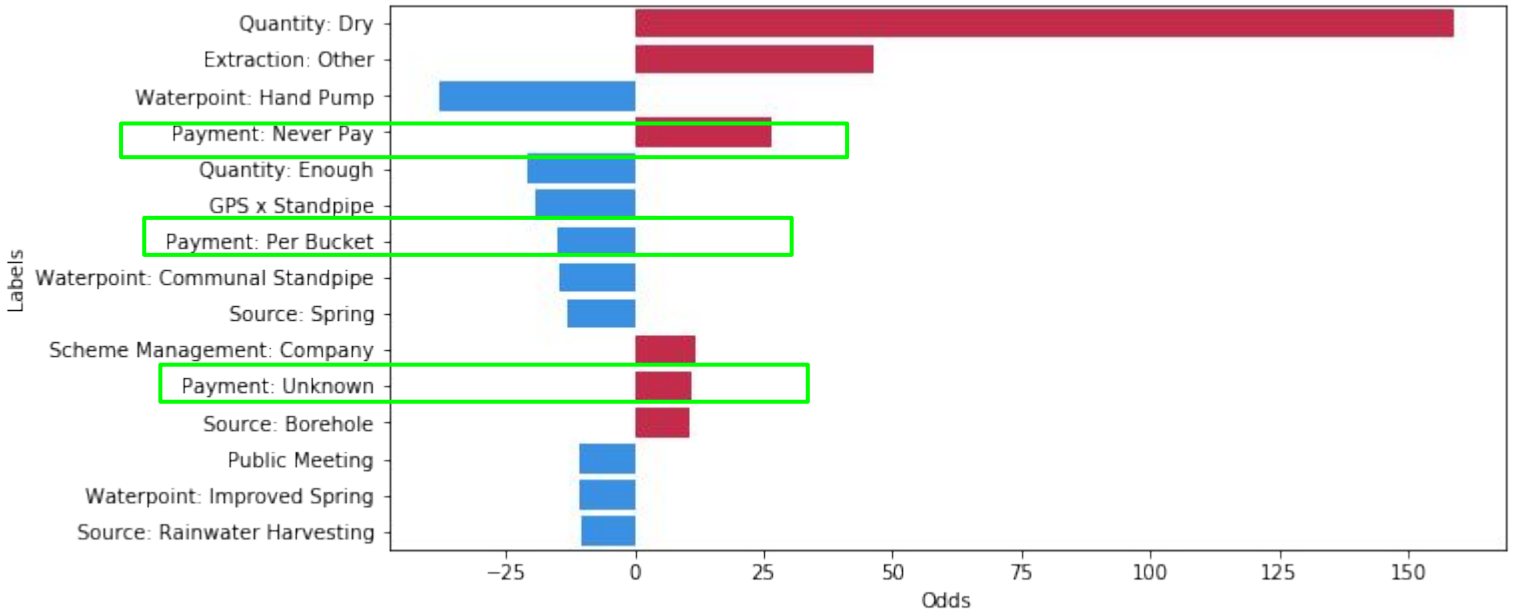
Recall

68%

Logistic Regression: Top 15 Features

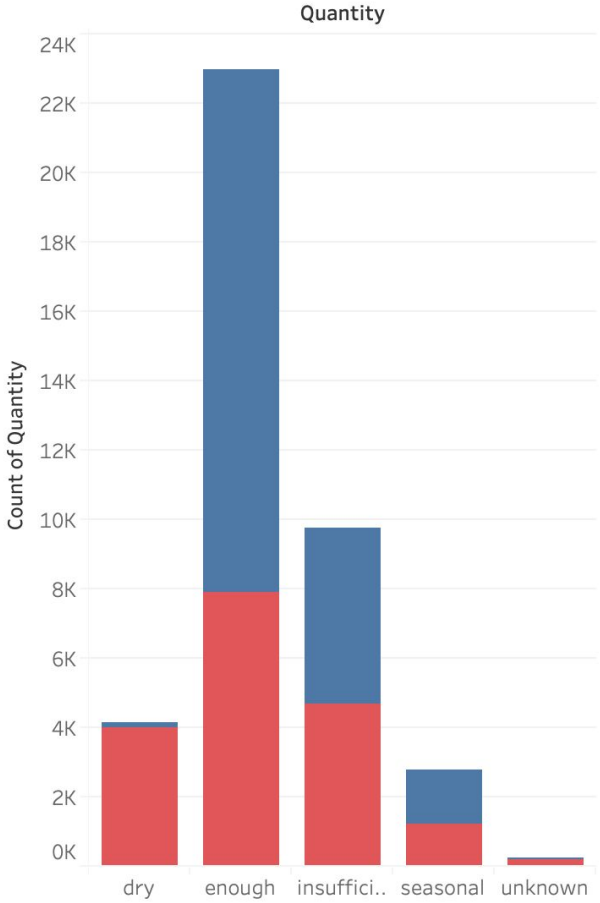


Logistic Regression: Top 15 Features



Logistic Regression: Top Feature

- Increases odds of pump being non-functional by 159%
- Explanation:
  - Dry running
  - Defined as non-functional



Conclusions

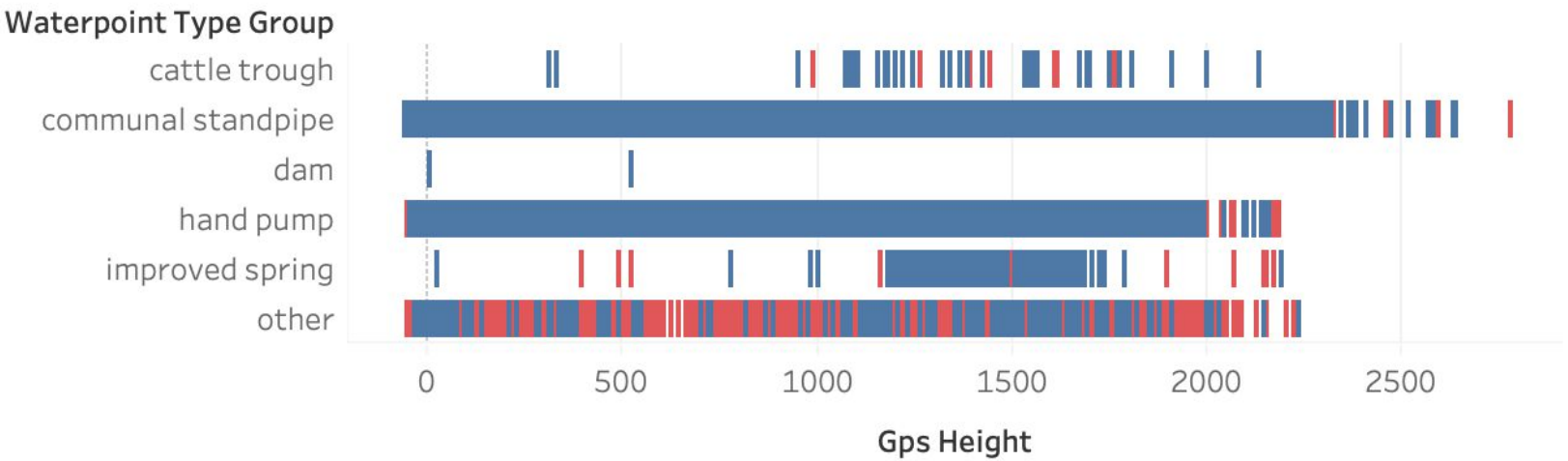
- Identify broken water pumps:
  - Random Forest - 79% accuracy
- Identify important features:
  - Logistic Regression - 70% accuracy
  - Quantity: Dry

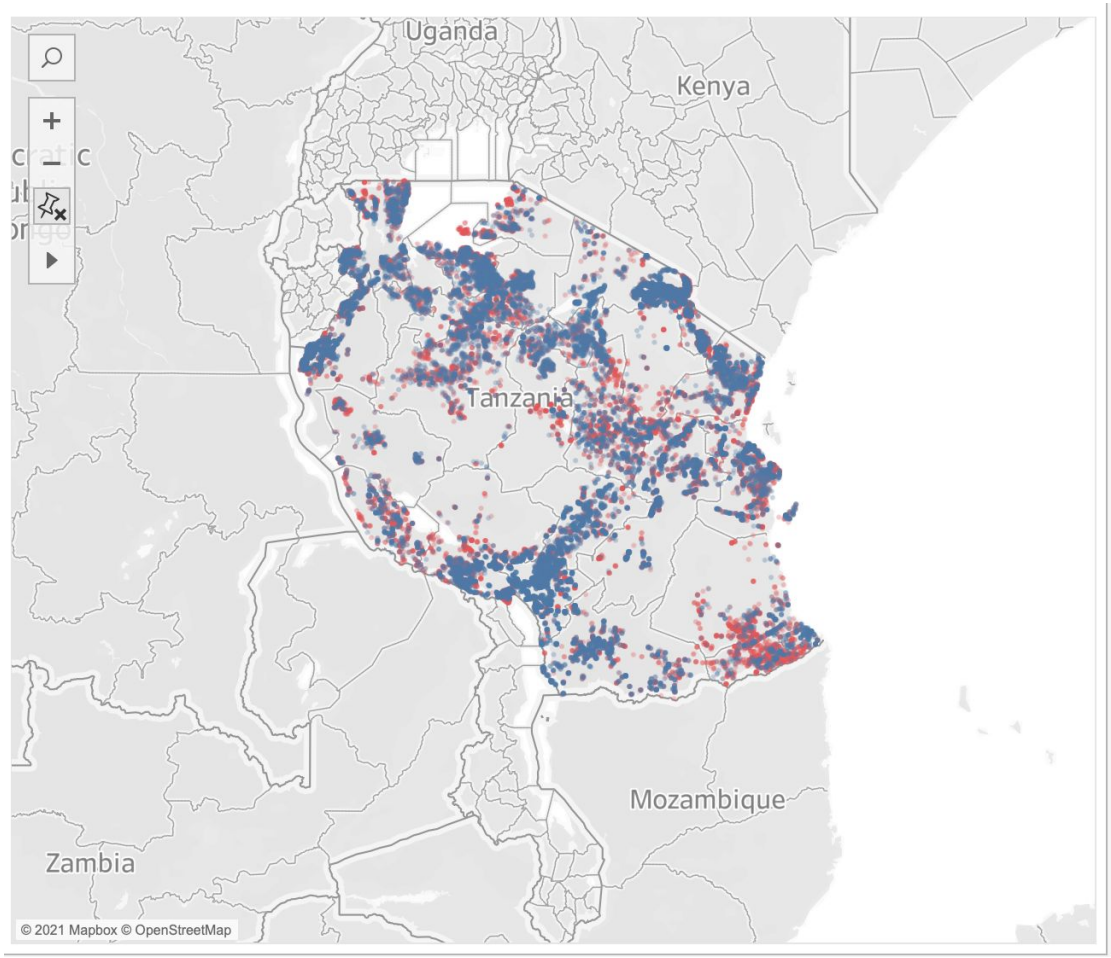
Future work

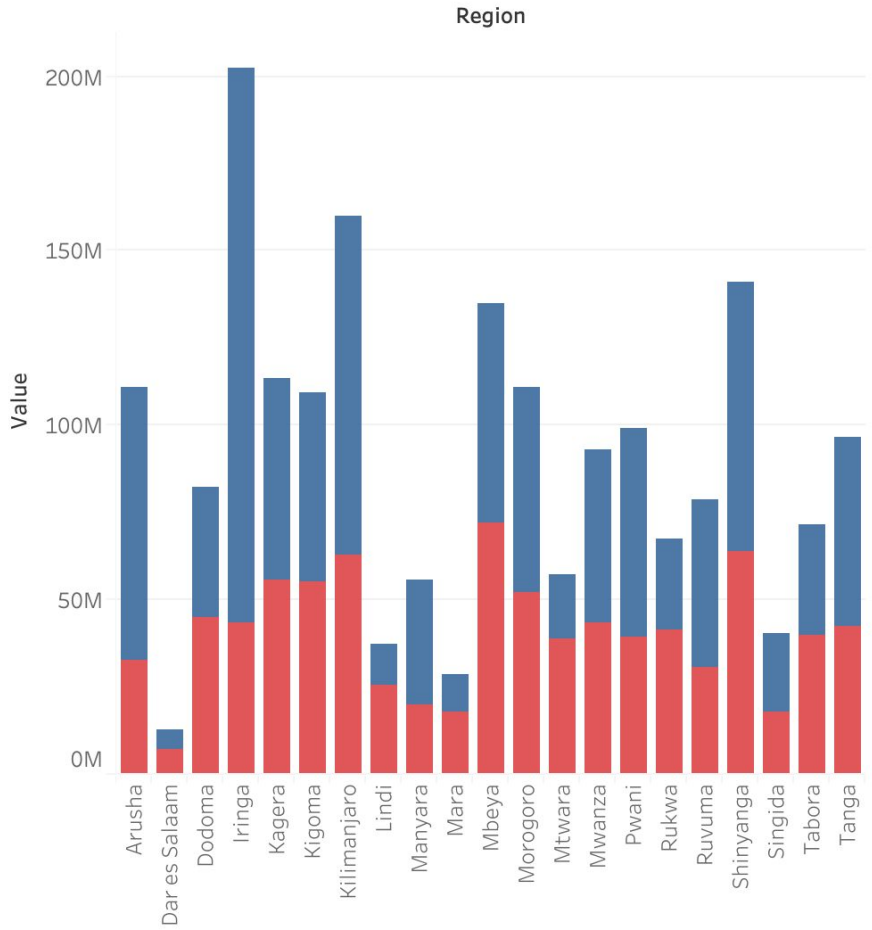
- Improve logistic regression model
  - More feature engineering
- Improve random forest model
  - Ensemble methods
- Interpretation

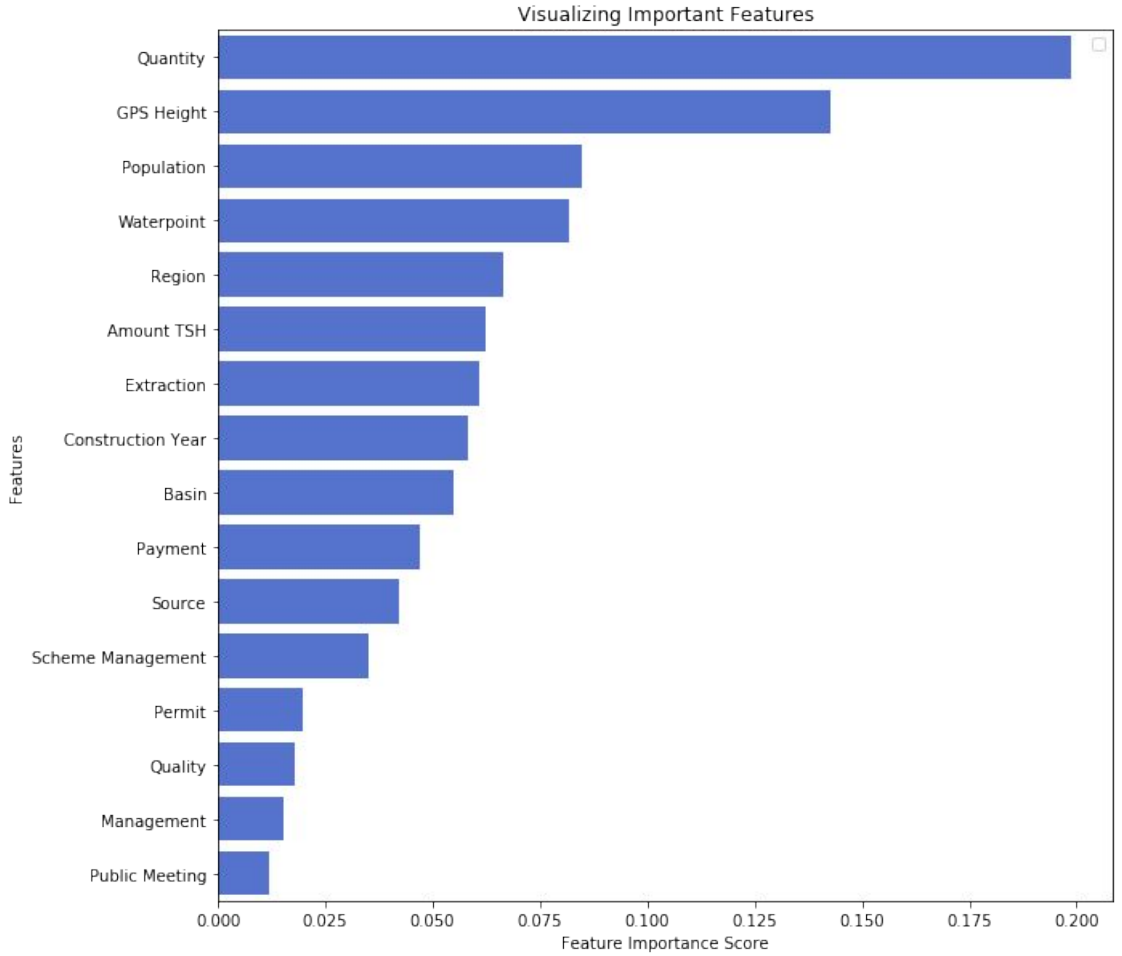
Thank you!

Questions?


Waterpoint Type Group
cattle trough
communal standpipe
dam
hand pump
improved spring
other
0
500
1000
1500
2000
2500
Gps Height


Uganda
Kenya
Tanzania
Mozambique
Zambia
© 2021 Mapbox © OpenStreetMap


Region
Value
200M
150M
100M
50M
0M
Arusha
Dar es Salaam
Dodoma
Iringa
Kagera
Kigoma
Kilimanjaro
Lindi
Manyara
Mara
Mbeya
Morogoro
Mtwara
Mwanza
Pwani
Rukwa
Ruvuma
Shinyanga
Singida
Tabora
Tanga


Visualizing Important Features
Quantity
GPS Height
Population
Waterpoint
Region
Amount TSH
Extraction
Construction Year
Basin
Payment
Source
Scheme Management
Permit
Quality
Management
Public Meeting
Features
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
Feature Importance Score