

Tanzanian Water Pump Project

Machine Learning and Analysis: Data Science

Orlando Vilar, Nov 2022

Introduction

Introduction

Overview & Business Understanding

- 1/6 of the world population lack access to safe water;
- The average African uses 5 gallon of water daily;
- How can we predict whether a Tanzanian water pump is functional or not?
- **Relevance:** overall implication to public policy, governmental agencies, NGOs, general public.

Data & Analysis

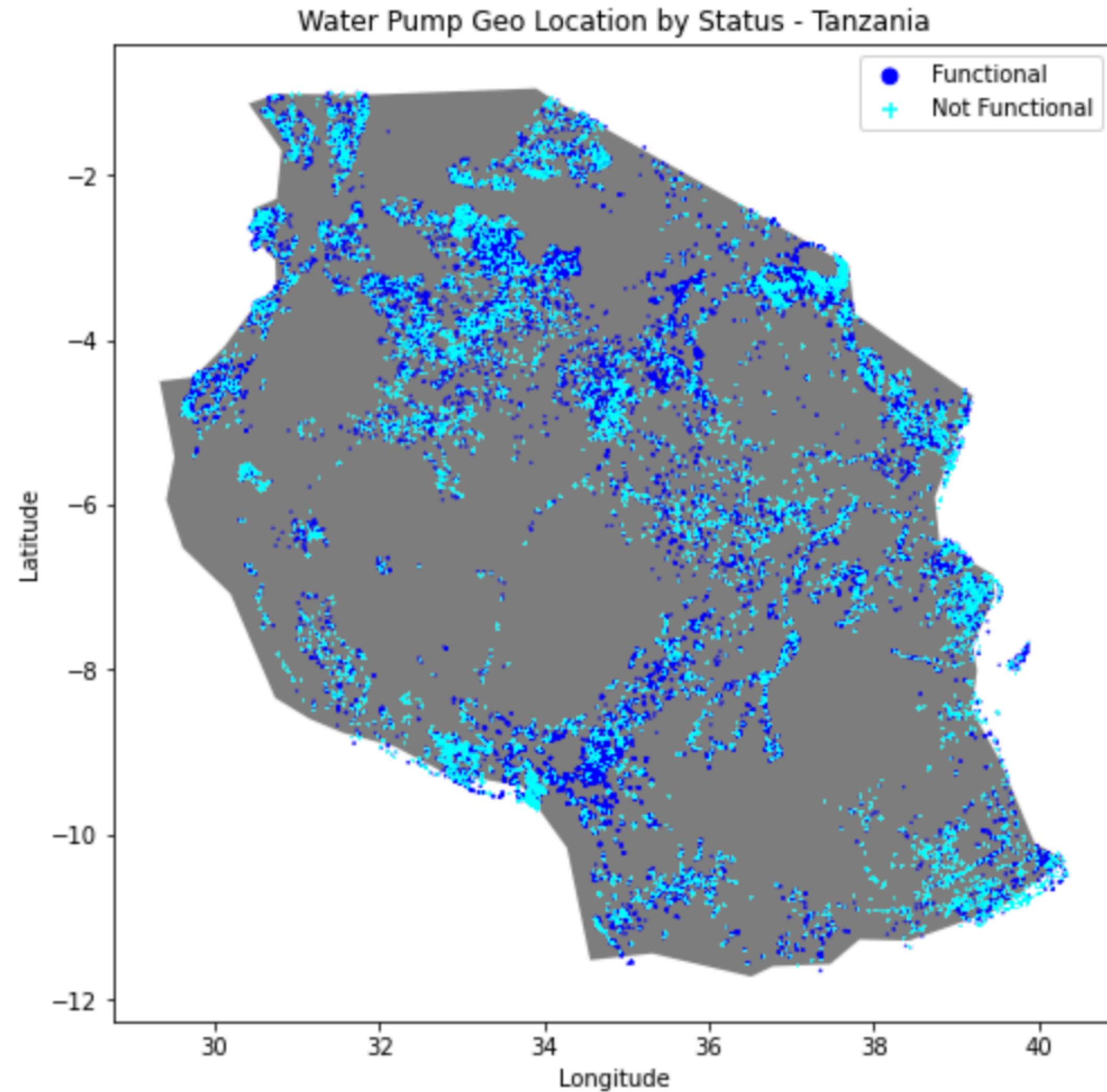
Data Understanding

Sources of Information

- Databases:
 - Taarifa and Tanzanian Ministry of Water (from DrivenData);
- Variables:
 - Status, location, extraction type, source, quantity, construction year, management, etc.
- Additional feature:
 - Geo-plotting of water pumps by status.

Tanzanian Water Pump by Status

Geolocation



Closing Remarks

Strategies

Key insights

- **Prophylactic:** the best model is able to predict whether a pump is working or not by 80%;
 - This can lead to better planning on when to fix functional pumps;
 - It also indicates which ones should be fixed right now;
- **Expansion:** areas with less pumps can be used for expansion and diminishing traveling distance/time.

Limitations

Further analysis

- Adding demographic data about each specific area;
- Predicting pumps that are functional and need repair;
- Matching with more robust numerical data can improve the models;
 - The data is noisy, discretion and refinement is advised.

Thank you!

<https://github.com/ovilar>