Tanzanian Water Pump Project

Machine Learning and Analysis: Data Science

Orlando Vilar, Nov 2022

Introduction

Introduction

Overview & Business Understanding

- House as an asset is deeply ingrained in our culture;
- Which features impact the most housing prices?
- How can we assess housing prices using property's characteristics and location within King County - WA;
- **Relevance**: overall implication to investors decision-making, realtors, contractors, and the general public.

Data & Analysis

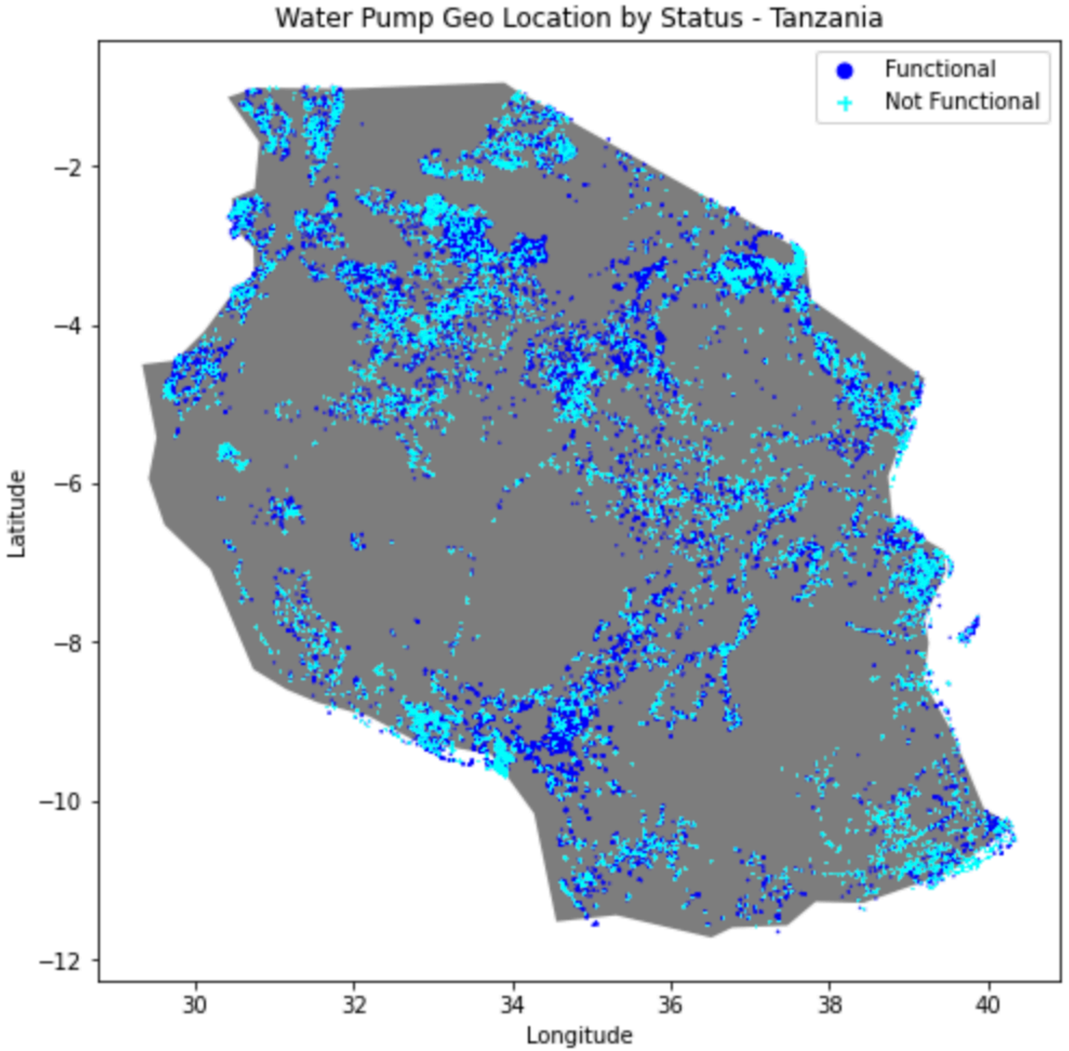
Data Understanding

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