**Data-Driven Decision Support for Water Pump Maintenance in Tanzania: A Predictive Modelling and Analysis Study**

Adam Williams, Filip Kornata

**1. Introduction**

Hello

**2. Dataset Description**

The provided dataset contains information on 59,400 water pumps in Tanzania. The dataset has 40 features containing information on each pump, with information ranging from the pump’s location, water quality, surrounding population size, pump type,

**3. Research Questions**

The goal of this project is to provide insights which could allow a decision maker to better allocate maintenance resources across Tanzania, and possibly lead to improved pump installation factors into the future. We have formulated 4 research questions to provide these insights:

1. *What are the main factors associated with pump failure or malfunction, and how do these vary across different regions of Tanzania?*
2. *Which operators and/or management groups have the highest success rates in maintaining water pumps, and how do these rates vary based on factors which may make pump maintenance easier, such as water cost, pump type, or location remoteness?*
3. *Are there any interactions or correlations between different features of the dataset, such as between the water basin and pump type, that could provide insights into the underlying causes of pump failure?*
4. *How does the age of a water pump relate to its functionality, and is there a point at which pumps become significantly more likely to break down or require replacement?*

**4. Data Pre-Processing**

Hello