

Thomas Adler-Wireframe Deliverable-25/06/2022

Barebone documentation for understanding the start of my project.

1. Intro

Water consumption is one of the most basic needs. Humans need water to function and fuel their body, irrigate agriculture, grow livestock, maintain hygiene, and cure themselves, and much more. Thus, water infrastructure is crucial for a community to survive, develop and thrive.

But what happens when a resource as important as water is hard to or even impossible to access? What happens when the provision of this resource is not sufficient, or the quality of it not high enough? These situations can have a devastating effect on the community's ability to feed itself, stay healthy and prosper. This can lead to protests, revolts, civil conflict and violence and even coups.

2. Question

First, I want to test the assumption that a lower quality of water infrastructure can make a region more likely to experience violence or conflict. Then, I want to predict, using machine learning, which regions are most likely to experience unrest in the future based on the quality of their water infrastructure.

3. Why should I care?

This model will be crucial for governments, public bodies, and water service providers as they will be able to identify which region is most at risk of instability. They can then act on this insight and strategically choose which water infrastructure should be fixed or improved first. This will enable a more efficient deployment of resources and will benefit citizens, enjoying better water provision, and governments, ensuring national and political stability in the country is upheld.

4. Data

a. Independent variables-Water infrastructure

We will be using the [Water Point Data Exchange dataset](#), a global dataset on water infrastructure. The majority of the observations are in sub-Saharan Africa. We will have to explore what variables will be most useful, but will most likely belong to one of the following categories:

- Number of citizens served (usage)
- Number of citizens not served
- Accessibility of the water point
- Technology/complexity of the infrastructure
- Water quality
- Water quantity
- Functioning water point
- Improvement of water point over time

b. Control variables-Demographic and economic indicators

We will also have to control for other variables to attempt to isolate the effect of water

infrastructure on our outcome variable. As of now, the dataset for this is uncertain as we will have to better define what countries and regions we will be focusing on. Relevant controls will most likely include:

- Income
- GDP
- Urbanisation
- Development indicators
- Health provision/Life expectancy
- Education levels

c. Dependent variables-Present of conflict and violence

We will use the [Armed Conflict Location & Event Data Project dataset](#), showing us all battles, protests, riots and violence in our selected region. Luckily, data for Africa is the richest, with entries going up to 1997.

5. Next steps

For all three of our datasets, further cleaning and EDA is needed to choose relevant, timely, consistent and available variables to construct our model. We will then be able to choose the best set of models and start our analysis.