Project Executive Summary

Aaron, Addison, Christian, Lucas

Our group was given the topic of healthcare. Our goal is to determine the correlation between percentage of underweight adults, access to safe drinking water, and adult mortality rate in countries worldwide in order to use machine learning to predict mortality rates. We'll be gathering data from the World Health Organization's API, simulating real-time data collection with Kafka, and cleaning & transforming data with Python in Azure Databricks. We also plan to forecast future statistics using current statistics.

To include the Census data, a state (or region) of the US with a given population value will be contrasted with a foreign country approximately sharing that population. This will allow us to compare how different circumstances (water availability, for instance) affect the mortality rate, giving us the ability to infer some potential cause and effect relationships between these variables.

Exploratory Questions

1. Is there a correlation between lack of water and adult mortality?
2. Is there a correlation between lack of water and underweight adults?
3. Is there a correlation between underweight adults and adult mortality?
4. Can we predict a trend in adult mortality given past data in water accessibility and malnourishment?
5. Is there another country that is somewhat equivalent to the US in terms of population, but whose mortality rate suffers due to lack of water?
6. Is there another country that is somewhat equivalent to the US in terms of population, but whose mortality rate suffers due to the number of underweight adults?
7. What are the overall trends of each stat? (Over time)
8. Is there another country that is somewhat equivalent to the US in terms of economic status, but whose mortality rate suffers due to lack of water?
9. Is there another country that is somewhat equivalent to the US in terms of economic status, but whose mortality rate suffers due to the number of underweight adults?