Analyzing the Impact of Water Pollution on Waterborne Disease Prevalence

Exploring the relationship between water quality and public health outcomes (2000-2025)

Introduction

- Objective: Investigate how water pollution impacts the prevalence of waterborne diseases.
- Key Questions:
  - Which water quality parameters are most associated with disease outbreaks?
  - How do socioeconomic factors modulate these relationships?
- Dataset: Kaggle - "Water Pollution and Disease"

Dataset Overview

- Source: Kaggle Dataset
- Key Features:
  - Water Quality Indicators: pH, turbidity, contaminant levels, etc.
  - Socioeconomic Factors: GDP, healthcare access, sanitation coverage.
  - Disease Metrics: Diarrheal, cholera, and typhoid cases.

Exploratory Data Analysis (EDA)

- Steps:
  - Data cleaning and preprocessing.
  - Summary statistics and visualizations.
- Key Insights:
  - High turbidity and low dissolved oxygen correlate with disease prevalence.
  - Socioeconomic factors like sanitation coverage show unexpected trends.

Modeling Approach

- Water Quality Analysis:
  - Features: pH, turbidity, contaminant levels, etc.
  - Target: Disease cases (diarrheal, cholera, typhoid).
- Socioeconomic Analysis:
  - Features: Access to clean water, GDP, urbanization rate.
  - Target: Disease cases.

Model Results

- Water Quality Models:
  - Diarrheal Cases: $R^2 = 0.747$, pH and turbidity significant.
  - Cholera Cases: $R^2 = 0.738$, contaminant level significant.
  - Typhoid Cases: $R^2 = 0.739$, pH and dissolved oxygen significant.
- Socioeconomic Models:
  - Diarrheal Cases: $R^2 = 0.739$, water source and sanitation coverage significant.
  - Cholera and Typhoid: Similar trends with socioeconomic factors.

Conclusions

- Key Findings:
  - Water quality indicators (pH, turbidity) are critical drivers of disease prevalence.
  - Socioeconomic factors (sanitation, healthcare access) show complex relationships.
- Recommendations:
  - Improve water quality monitoring.
  - Address confounding factors in socioeconomic data.

Future Work

- Next Steps:
  - Explore causal relationships using advanced statistical methods.
  - Incorporate additional datasets for validation.
  - Develop predictive models for early disease outbreak detection.

Thank You