# **Water Quality Explanatory Data Analysis**

## **Project Overview**

Access to safe drinking water is essential for public health and overall well-being. This project conducts an exploratory data analysis (EDA) of various water quality parameters to assess their impact on water potability. The analysis examines key indicators such as pH, hardness, total dissolved solids (TDS), chloramines, sulfate levels, and turbidity to understand their implications for human consumption.

## **Goals**

The primary goals of this project are:

* To evaluate the quality of drinking water by analyzing various water parameters.
* To identify trends and correlations between water quality indicators and potability.
* To provide insights that can inform public health policies and interventions aimed at ensuring safe drinking water.

## **Methods**

The following methods were employed in this analysis:

1. **Data Collection**: Water quality data was collected from reliable sources, including public health databases and water quality reports.
2. **Exploratory Data Analysis (EDA)**: Various statistical techniques and visualizations were utilized to analyze the data, including:
   * Descriptive statistics to summarize the data.
   * Correlation analysis to identify relationships between different water quality parameters.
   * Data visualization techniques such as histograms, scatter plots, and heatmaps to illustrate findings.
3. **Statistical Testing**: Appropriate tests were conducted to determine the significance of observed trends and relationships.

## **Results**

Key findings from the analysis include:

* The pH values of the water samples generally fell within the WHO-recommended range, indicating suitable acidity levels.
* Certain water quality parameters, such as turbidity and TDS, showed significant correlations with potability, suggesting potential health risks.
* Visualizations revealed patterns in the distribution of water quality parameters, highlighting areas where further investigation may be warranted.

## **Data Processing**

Data processing steps included:

1. **Data Cleaning**: Missing values and outliers were identified and addressed to ensure the integrity of the analysis.
2. **Data Transformation**: Relevant features were scaled and encoded as necessary to prepare the data for analysis.
3. **Exploratory Analysis**: Initial visualizations were created to understand the distribution and relationships within the dataset.

## **Conclusion**

This project highlights the importance of monitoring water quality parameters to ensure safe drinking water. The insights gained from this analysis can serve as a valuable resource for policymakers, public health officials, and researchers working to improve water safety and health outcomes.