# Phase 5: Documentation

*Project Title*: **Water Quality** **Analysis**.

**Objective**:

The objective of this project is to analyze water quality data to assess its potability. The project aims to provide insights into the factors influencing water potability and develop a predictive model to determine if a water source is safe for consumption.

**Design Thinking Process:**

1. ***Empathize***: Understand the need for safe drinking water and its impact on public health.
2. ***Define***: Define the problem of ensuring water potability and its significance.
3. ***Ideate***:Brainstorm approaches to analyze water quality data and develop a predictive model.
4. **Prototype**:Create a data analysis plan and outline the project phases.
5. **Test**: Conduct the analysis and evaluate the model’s performance.
6. ***Implement***: Share the findings and model to promote safe water consumption.

**Development Phases**:

1. ***Data Collection*:**Gather water quality data from various sources, including chemical and physical parameters, along with potability labels.
2. ***Data Preprocessing***:Clean the data by handling missing values and outliers, standardizing units, and encoding categorical variables.
3. ***Exploratory Data Analysis*** (EDA): Explore the dataset to identify trends, correlations, and potential features that affect water potability.
4. ***Data Visualization*:** Create visualizations such as scatter plots, histograms, and heatmaps to better understand the data.
5. ***Feature Selection*:**Determine the most relevant features for the predictive model.
6. ***Predictive Modeling***:Build a machine learning model (e.g., logistic regression, random forest) to predict water potability using the selected features.
7. **Model Evaluation:**Assess the model’s performance through metrics like accuracy, precision, recall, and F1 score.
8. ***Insights and Recommendations:*** Summarize the findings, highlighting factors affecting potability and suggesting improvements to water sources.
9. **Documentation and** ***Submission***: Prepare a comprehensive report and submit it to relevant authorities or stakeholders.

**Analysis Objectives:**

1. Identify key water quality parameters that impact potability.
2. Develop a predictive model to classify water sources as potable or non-potable.
3. Assess the model’s accuracy in determining water potability.

**How Insights Help Assess Water Quality and Potability:**

1. ***Identifying Critical Factors*:** Insights from the analysis can reveal which chemical and physical parameters are most important for determining potability. This information can guide water quality monitoring efforts.
2. ***Early Warning System****:* The model can serve as an early warning system to identify potentially unsafe water sources, allowing timely corrective actions.
3. ***Resource Allocation***:By understanding the factors affecting potability, authorities can allocate resources more efficiently to improve water treatment and quality.
4. ***Public Awareness****:*Sharing the findings with the public can raise awareness about water quality and promote responsible water usage.

**PROGRAM FOR VISUALISATION**:

Import necessary libraries

Import pandas as pd

From sklearn.model\_selection import train\_test\_split

From sklearn.linear\_model import LinearRegression # Example model, choose the appropriate one

From sklearn.metrics import mean\_absolute\_error, mean\_squared\_error, r2\_score

Import matplotlib.pyplot as plt

# Load and preprocess your water quality data

Data = pd.read\_csv(“water\_quality\_data.csv”) # Load your dataset

# Data preprocessing (e.g., handling missing values, feature selection)

# Split the data into training and testing sets

X = data[[‘Feature1’, ‘Feature2’, …]] # Select relevant features

Y = data[‘Target’] # Target variable

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, Unrandom\_state=42)

# Choose and train a model (e.g., Linear Regression)

Model = LinearRegression()

Model.fit(X\_train, y\_train)

# Make predictions on the test data

Y\_pred = model.predict(X\_test)

# Evaluate the model

Mae = mean\_absolute\_error(y\_test, y\_pred)

Mse = mean\_squared\_error(y\_test, y\_pred)

R2 = r2\_score(y\_test, y\_pred)

Print(f”Mean Absolute Error: {mae}”)

Print(f”Mean Squared Error: {mse}”)

Print(f”R-squared: {r2}”)

# Visualization (e.g., scatter plot of observed vs. values)

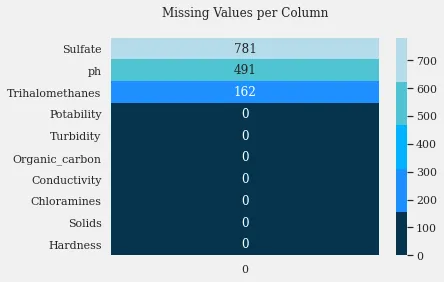
Plt.scatter(y\_test, y\_pred)

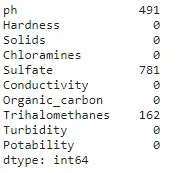
Plt.xlabel(“Observed”)

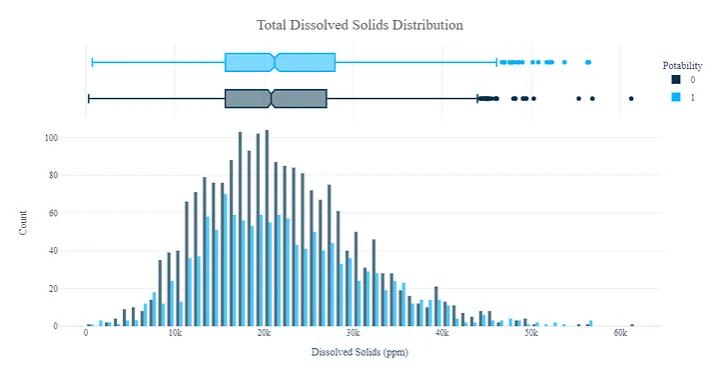
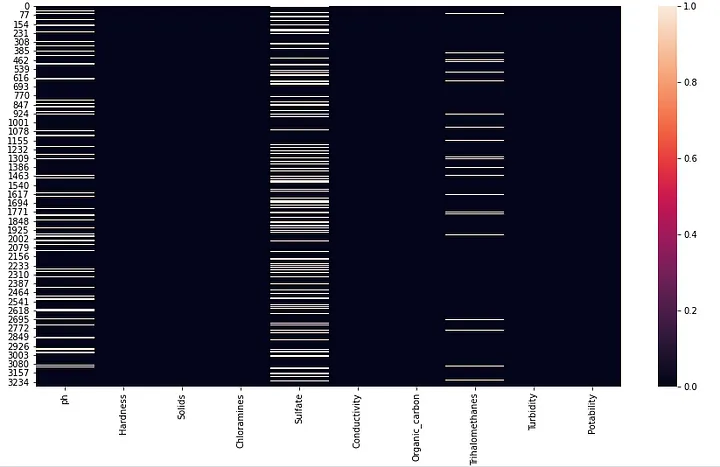
Plt.ylabel(“Predicted”)

Plt.show()

**Output**:









**Conclusions**:

This water quality analysis project focuses on using data-driven insights and predictive modeling to assess water potability. The findings and model can be valuable tools in ensuring safe and clean drinking water for the community, thereby contributing to public health and well-being.