**WATER QUALITY ANALYSYS**

* The project involves analyzing water quality data to assess the suitability of water for specific purposes, such as drinking. The objective is to identify potential issues or deviations from regulatory standards and determine water potability based on various parameters. This project includes defining analysis objectives, collecting water quality data, designing relevant visualizations, and building a predictive model

**ABSTRACT:**

* Water quality analysis is to measure the required parameters of water, following standard methods, to check whether they are in accordance with the standard.
* Colorimeters and Photometers are used to analyze samples of water, suspended sediment, and bottom material for their content of inorganic and organic constituents.
* Commonly used methods include chelating ion-exchange and, for trace organic analysis, solvent extraction, carbon adsorption, and resin adsorption using nonionic macroeticular resins. Minor variations in microbiological analyses can cause significant changes in results.

**OBJECTIVES:**

* Water quality analysts monitor and analyze the chemical, physical, and biological components of water to ensure that it meets safety standards for human consumption
* Water has a neutral pH of 7, which indicates that it is neither acidic or basic. The scale ranges from 0 (very acidic) to 14 (very basic). It is normal for water to have a range of between 6.5 and 8.5 on the scale. PH in water may fluctuate with differing environmental factors.
* 0 to 60 mg/L (milligrams per liter) as calcium carbonate is classified as soft; 61 to 120 mg/L as moderately hard; 121 to 180 mg/L as hard; and more than 180 mg/L as very hard.
* The objective of the water quality analysis project is to assess and improve the quality of water in a specific area, such as a river, lake, or reservoir. This project aims to provide accurate and timely data on water quality parameters, identify potential sources of contamination, and develop strategies to maintain or enhance water quality.
* improving water quality for safe drinking water, protecting aquatic ecosystems, or supporting recreational activities.
* Develop a prototype of the water quality monitoring system, which may include sensors, data collection methods, and a data management platform.

**ANALYSIS APPROCH:**

1. Sample Collection: Collect water samples from various sources, ensuring they are representative of the area of interest. Properly label and store the samples to prevent contamination.

2. Physical Parameters: Measure physical characteristics like temperature, turbidity, color, and odor. These can provide initial insights into water quality.

3. Chemical Analysis: Conduct chemical tests to determine the levels of key constituents such as pH, dissolved oxygen (DO), nutrients (nitrogen and phosphorus), heavy metals, and organic contaminants. This helps assess pollution levels and potential health risks.

4. Biological Assessment: Examine the presence and diversity of aquatic organisms, such as macroinvertebrates and algae. These can indicate the ecological health of the water.

5. Microbiological Testing: Test for the presence of bacteria, viruses, and other microorganisms. High levels of certain pathogens can pose health risks.

6. Suspended Solids and Sediment Analysis: Determine the amount of suspended solids and sediment in the water, which can affect water clarity and aquatic habitats.

7. Taste and Odor Analysis: Evaluate the taste and odor of the water, which can be indicative of contamination or the presence of certain chemicals.

8. Statistical Analysis: Analyze the data using statistical methods to identify trends, correlations, and anomalies. This helps in understanding the overall water quality.

9. Regulatory Compliance: Compare the results with local, state, and national water quality standards and regulations to assess compliance.

10. Geographic Information Systems (GIS): Use GIS to map and visualize water quality data spatially, aiding in identifying pollution sources and patterns.

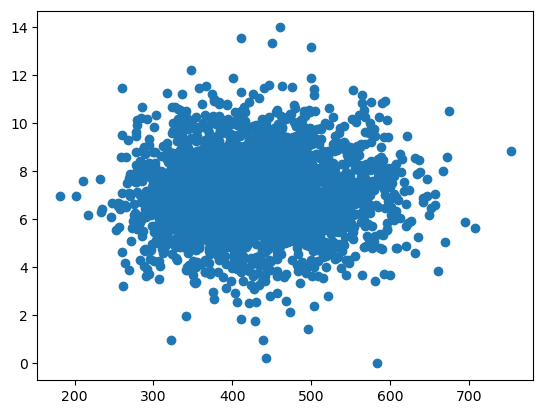
11. Long-Term Monitoring: Establish regular monitoring programs to track changes in water quality over time, which can be critical for detecting trends and addressing issues.

12. Interpretation and Reporting: Interpret the findings and prepare reports that communicate the water quality status, potential risks, and recommended actions to relevant stakeholders.

13. Mitigation and Remediation: Implement measures to address water quality issues, which may include treatment, pollution control, and habitat restoration.

14. Public Awareness: Educate the public about water quality concerns and promote responsible water use and conservation.

**VISUALIZATION SELECTION:**



**ALGORITHM:**

**LINEAR REGRESION:**

* Linear regression is a statistical method that can be used to model the relationship between two variables. It can be used to predict the water quality index based on parameters like temperature, dissolved oxygen, pH value, Turbidity, Nitrates and fecal coliform. It relates one or more explanatory variables to a response variable using linear coefficent.
* Urbanization and industrialization have contributed to a deterioration of water quality at an unprecedented pace, contributing to harrowing diseases. The dataset that has been used for WQI assessment consists of a total of 1992 samples with six columns corresponding to six parameters mentioned earlier. Based on the simulation results it has been observed that the proposed regression model is able to predict the WQI effectively.
* It’s important to note that linear regression is just one of many statistical methods that can be used for water quality analysis. Other methods such as logistic regression, decision trees, and neural networks may also be appropriate depending on the specific problem at hand

**STEPS:**

**1.Data Collection:**

* Collect data on the water quality parameters of interest. This data should include both normal and unusual values for the parameters.
* It includes pH value, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic Carbon, Trihalomethanes.

**2.Data Preparation:**

* Prepare the data for analysis by cleaning and organizing it. This may involve removing outliers, filling in missing values, and transforming the data if necessary.

**3.Model Selection:**

* Choose an appropriate linear regression model to fit the data. This may involve selecting a simple linear regression model or a more complex model that includes multiple variables.

**4.Model fitting:**

* Fit the selected model to the data using a statistical software package such as R or Python.

**5.Model Evaluation:**

* Evaluate the performance of the model by examining its goodness of fit and statistical significance.

**6.Prediction:**

* Once the model has been fitted and evaluated, it can be used to predict unusual values in water quality parameters.

**7.Deployment:**

* If the model is satisfied and meets your requirements you can deploy it in a production environment in real time.

**CHECKING THE MISSING VALUES**

import pandas as pd

#This library used to read and write a data frame ,

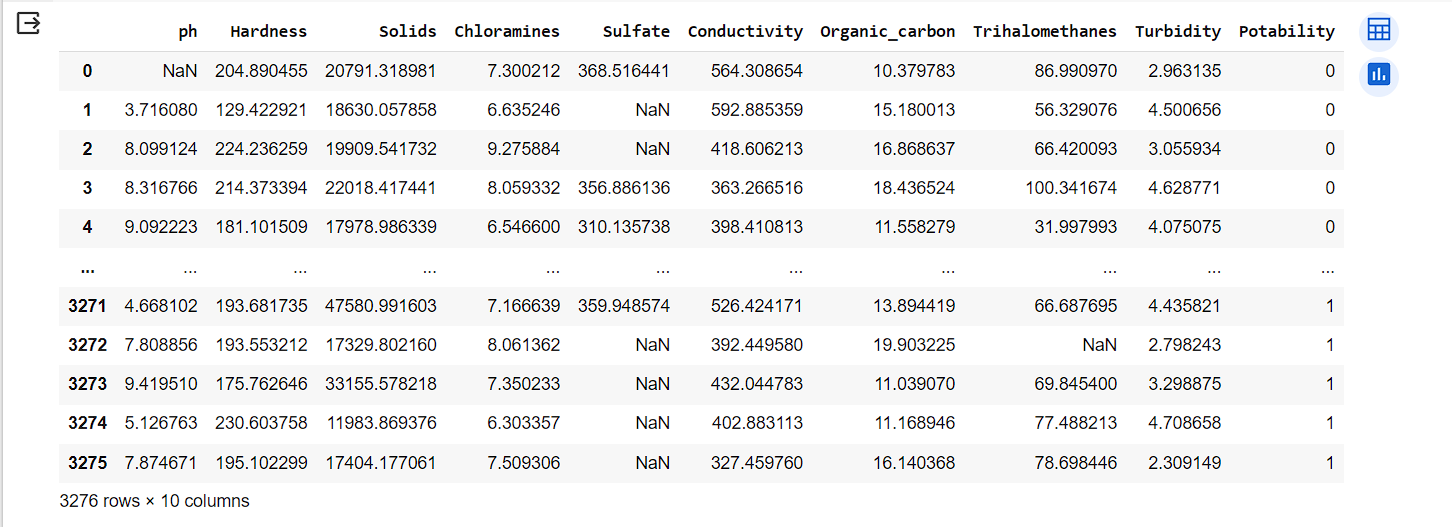
df = pd.read\_csv('/content/water\_potability.csv')

#upload the dataset using the pandas library

df

#printing the data set

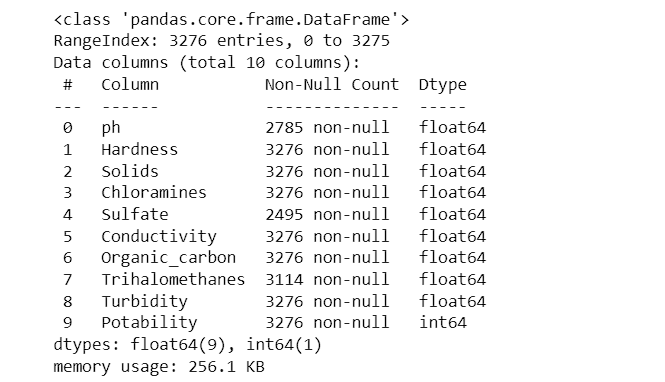
**OUTPUT:**



df.info()

#checking the datatype of the dataset

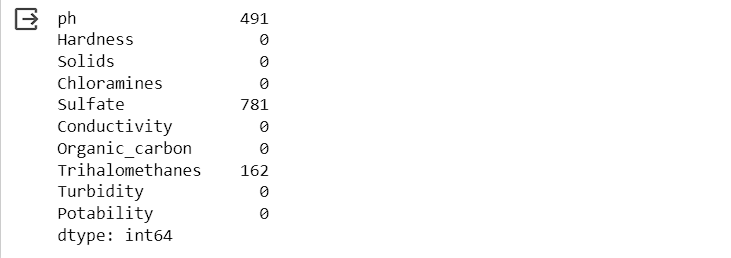
**OUTPUT:**



df.isnull().sum()

#checking the missing values in the dataset

**OUTPUT:**



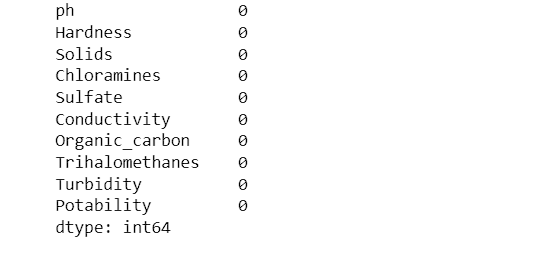
df.fillna(9.5,inplace=True)

#Replace the missing values using fillna() command with the common value 9.5

df.isnull().sum()

#again checking for the missing values

**OUTPUT:**



To replicate a water quality analysis project in Python, gather data from reliable sources and clean it. Perform EDA to understand the dataset and create visualizations. Visualize the data and extract relevant features. Split the data into training and testing sets. Build a predictive model and evaluate its performance. Visualize the model's results and deploy it for production. Monitor the model's performance and update it with new data as needed.

Exploratory Analysis:

**#Importing the libraries**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

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

from sklearn.linear\_model import LinearRegression

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

**#upload the data set using the library**

df = pd.read\_csv("/content/water\_potability.csv",

encoding="ISO-8859-1")

df = df.replace("NA", np.nan)

df = df.dropna()

df = df.astype(float)

df = df[df["Conductivity"] > 0]

X = df[["ph", "Hardness", "Solids", "Chloramines", "Sulfate", "Conductivity"]]

y = df["Turbidity"]

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

model = LinearRegression()

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

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

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

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

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

print(f"R-squared: {r2:.3f}")

print(f"Mean Squared Error: {mse:.3f}")

print(f"Mean Absolute Error: {mae:.3f}")

**#ploting the data for scatter plot**

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

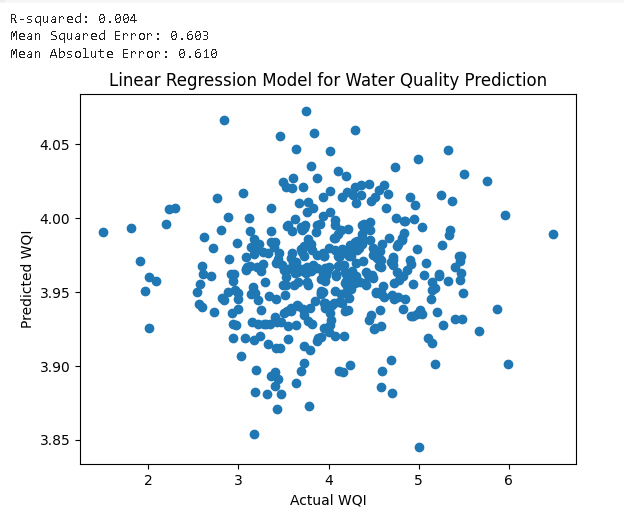
plt.xlabel("Actual WQI")

plt.ylabel("Predicted WQI")

plt.title("Linear Regression Model for Water Quality Prediction")

plt.show()

**#OUTPUT:**



**Virtualization:**

import matplotlib.pyplot as plt

import numpy as np

import pandas as pd

plt.hist(df["Solids"])

**Output:**

A graph with numbers and a black background

Description automatically generated

x=df["Sulfate"]

y=df["Trihalomethanes"]

matrix = np.corrcoef(x, y)

print(matrix)

**Output:**

A black background with white text

Description automatically generated

plt.scatter(df["ph"],df["Solids"])

**Output:**

A blue dot diagram with numbers

Description automatically generated with medium confidence

sns.barplot(x=df["Organic\_carbon"],y=df["Solids"],data=df,palette="plasma")

**Output:**

A screen shot of a computer screen

Description automatically generated

Predict :

import numpy as np

import pandas as pd

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

from sklearn.linear\_model import LinearRegression

from sklearn.metrics import mean\_squared\_error

import matplotlib.pyplot as plt

data = pd.read\_csv("/content/water\_potability.csv")

X = data['Hardness'].values.reshape(-1, 1)

y = data['Turbidity'].values

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=0)

model = LinearRegression()

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

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

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

print(f"Mean Squared Error: {mse}")

plt.scatter(X, y, color='blue')

plt.plot(X\_test, y\_pred, color='red', linewidth=2)

plt.xlabel('Time')

plt.ylabel('Water Quality')

plt.show()

future\_time = np.array([2024, 2025, 2026]).reshape(-1, 1)

future\_quality = model.predict(future\_time)

print("Predicted water quality for future years:")

for year, quality in zip([2024, 2025, 2026], future\_quality):

    print(f"Year: {year}, Predicted Quality: {quality}")

A screen shot of a graph

Description automatically generated

**CONCLUTION:**

* Insights from water quality analysis play a critical role in assessing water quality and determining its potability .

Key Parameters:

* Water quality analysis typically involves measuring various parameters such as pH, turbidity, dissolved oxygen, temperature, and concentrations of specific contaminants (e.g., bacteria, heavy metals, nitrates).Safety for Human Consumption:Parameters like pH, turbidity, and dissolved oxygen are directly related to water's safety for human consumption.Low pH levels or high turbidity can indicate water quality issues .Detection of Contaminants:Water quality analysis can identify the presence of harmful contaminants, such as bacteria (e.g., E. coli), heavy metals (e.g., lead, arsenic), or chemicals (e.g., pesticides).Monitoring for Seasonal Variations:Insights from ongoing water quality analysis can reveal seasonal variations and trends.Early Warning for Pollution Events:Real-time monitoring systems can detect sudden pollution events, such as industrial spills or runoff from agricultural areas.Rapid response to such events can prevent contamination and safeguard potability.Continuous monitoring and adjustments are made to ensure that water leaving treatment plants is potable.