**Water Quality Analysis**

**OBJECTIVE:**

*A “Water Quality Analysis Project” is a research or environmental monitoring initiative aimed at assessing and ensuring the safety and purity of water sources, such as rivers,lakes, reservoirs,groundwater,and drinking water supplies.*

**PHASE4**

**1. Data Collection:**

*Start by obtaining the water quality dataset. You may find such*

*datasets on government websites, environmental organizations,*

*or research institutions&#39; data repositories.*

**2.Data Loading:**

*Once you have the dataset, load it into your preferred data*

*analysis environment. Popular tools for this purpose include*

*Python with libraries like Pandas, or R.*

**3.Data Inspection***:*

*Begin by inspecting the dataset to understand its structure and*

*the variables it contains. This typically includes checking the*

*number of rows and columns and the data types of each column.*

**4. Handling Missing Values:**

*missing values.*

*Imputing missing values with statistical measures (e.g., mean,*

*median, mode). Identify and handle missing values. Missing data can significantly impact*

*your analysis, so it&#39;s important to deal with them. Common strategiesIdentify and address outliers in the dataset. Outliers can skew your*

*analysis and should be treated appropriately. Some methods for handling*

*outliers include:*

*missing values.*

*Imputing missing values with statistical measures (e.g., mean,*

*median, mode). Identify and handle missing values. Missing data can significantly*

*include:*

*Removing rows with*

*Using domain knowledge to estimate missing values.*

**5. Handling Outliers:**

*Identify and address outliers in the dataset. Outliers can skew your*

*analysis and should be treated appropriately. Some methods for handling*

*outliers include:*

*Visualizing data using*  
*Water quality analysis is the process of assessing the physical, chemical, and biological characteristics of water to determine its suitability for various purposes, such as drinking, irrigation, industrial processes, and aquatic ecosystems. It is an essential aspect of environmental science and public health. The parameters and methods used in water quality analysis can vary* *depending on the specific objectives and the nature of the water source. Here are some key aspects of water quality analysis:*

**Physical Parameters:**

* + **Temperature:**

*Measuring the water temperature can indicate potential environmental changes and affect the solubility of gases and the growth of aquatic organisms.*

* + **Turbidity:**

*Turbidity measures the cloudiness or haziness of a fluid caused by large numbers of individual particles. High turbidity can affect light penetration and aquatic life*.

* + **pH:**

*The pH level indicates the acidity or alkalinity of the water. It can impact the solubility of chemicals and the health of aquatic organisms*.

* + **Dissolved Oxygen (DO):**

*DO is crucial for aquatic life. Low DO levels can harm fish and other aquatic organisms*.

* + **Chemical Oxygen Demand (COD) and Biological Oxygen Demand (BOD):**

*These parameters measure the organic and inorganic pollution levels in the water*.

* + **Nutrient Levels:**

*Monitoring nutrients like nitrogen and phosphorus is vital because they can lead to water eutrophication and harmful algal blooms.*

* + **Heavy Metals:**

*Testing for metals such as lead, mercury, and cadmium is essential due to their toxicity*.

**PROGRAM**

*import pandas as pd*

*data = {*

*‘Date’: [‘2023-01-01’, ‘2023-01-02’, ‘2023-01-03’, ‘2023-01-04’],*

*‘ProductCategory’: [‘A’, ‘B’, ‘A’, ‘C’],*

*‘Sales’: [100, 150, 120, 80],*

*‘Promotion’: [0, 1, 0, 1]*

*}*

*sales\_data = pd.DataFrame(data)*

*sales\_data[‘Date’] = pd.to\_datetime(sales\_data[‘Date’])*

*sales\_data[‘Year’] = sales\_data[‘Date’].dt.year*

*sales\_data[‘Month’] = sales\_data[‘Date’].dt.month*

*sales\_data[‘Day’] = sales\_data[‘Date’].dt.day*

*sales\_data[‘PreviousDaySales’] = sales\_data[‘Sales’].shift(1)*

*sales\_data[‘RollingMean3Days’] = sales\_data[‘Sales’].rolling(window=3).mean()*

*sales\_data = pd.get\_dummies(sales\_data, columns=[‘ProductCategory’], prefix=‘Category’)*

*print(sales\_data)*

***Output:***

*Date Sales Promotion Year Month Day PreviousDaySales*

*2023-01-01 100 0 2023 1 1 NaN*

*2023-01-02 150 1 2023 1 2 100.0*

*2023-01-03 120 0 2023 1 3 150.0*

*2023-01-04 80 1 2023 1 4 120.0*

*RollingMean3Days Category\_A Category\_B Category\_C*

*0 NaN 1 0 0*

*1 NaN 0 1 0*

*2 123.333333 1 0 0*

*3 116.666667 0 0 1*

**PROGRAM**

*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, r2\_score*

*data = {*

*‘Date’: [‘2023-01-01’, ‘2023-01-02’, ‘2023-01-03’, ‘2023-01-04’],*

*‘ProductCategory’: [‘A’, ‘B’, ‘A’, ‘C’],*

*‘Sales’: [100, 150, 120, 80],*

*‘Promotion’: [0, 1, 0, 1]*

*}*

*sales\_data = pd.DataFrame(data)*

*X = sales\_data.drop(‘Sales’, axis=1)*

*y = sales\_data[‘Sales’]*

*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)*

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

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

***Output:***

*Mean Squared Error: 165.0*

*R-squared (R2) Score: -3.0952380952380956*

*Program:*

*import pandas as pd*

*import matplotlib.pyplot as plt*

*data = {*

*'Date': ['2023-01-01', '2023-01-02', '2023-01-03', '2023-01-04'],*

*'Sales': [100, 150, 120, 80]*

*}*

*sales\_data = pd.DataFrame(data)*

*total\_revenue = sales\_data['Sales'].sum()*

*print("Total Revenue:", total\_revenue)*

*average\_sales = sales\_data['Sales'].mean()*

*print("Average Daily Sales:", average\_sales)*

*sales\_data['Date'] = pd.to\_datetime(sales\_data['Date'])*

*plt.figure(figsize=(8, 4))*

*plt.plot(sales\_data['Date'], sales\_data['Sales'], marker='o')*

*plt.xlabel('Date')*

*plt.ylabel('Sales')*

*plt.title('Sales Trend Over Time')*

*plt.grid(True)*

*plt.show()*

***Output:***

*Total Revenue: 450*

*Average Daily Sales: 112.5*