Introduction

This dataset provides information about various water quality parameters and whether the water is potable (safe for drinking) or not. It includes different physical and chemical properties of water that impact its drinkability.

Attributes Details

1. **Ph:**

* Description: Measures the acidity or alkalinity of water.
* Range: Typically between 0-14, where 7 is neutral.
* Missing Values: Yes (some values are missing).
* Importance: Extreme pH values can be harmful to health.

1. **Hardness:**

* Description: Indicates the concentration of dissolved calcium and magnesium in water.
* Units: mg/L (milligrams per liter).
* Missing Values: No.
* Importance: Hard water can cause scaling in pipes but is not harmful to health.

1. **Solids:**

* Description: Total dissolved solids (TDS) in water.
* Units: ppm (parts per million).
* Missing Values: No.
* Importance: High TDS may indicate contamination.

1. **Chloramines:**

* Description: Amount of chloramines used in water treatment.
* Units: mg/L.
* Missing Values: No.
* Importance: Disinfectant used to kill bacteria.

1. **Sulfate:**

* Description: Sulfate concentration in water.
* Units: mg/L.
* Missing Values: Yes.
* Importance: Excess sulfate can cause health issues like diarrhea.

1. **Conductivity:**

* Description: Measures water's ability to conduct electricity.
* Units: µS/cm (microsiemens per cm).
* Missing Values: No.
* Importance: Indicates the presence of dissolved salts.

1. **Organic Carbon:**

* Description: Organic carbon content in water.
* Units: mg/L.
* Missing Values: No.
* Importance: High levels indicate contamination.

1. **Trihalomethanes:**

* Description: Byproducts of chlorine disinfection.
* Units: µg/L (micrograms per liter).
* Missing Values: Yes.
* Importance: High levels may be harmful to health.

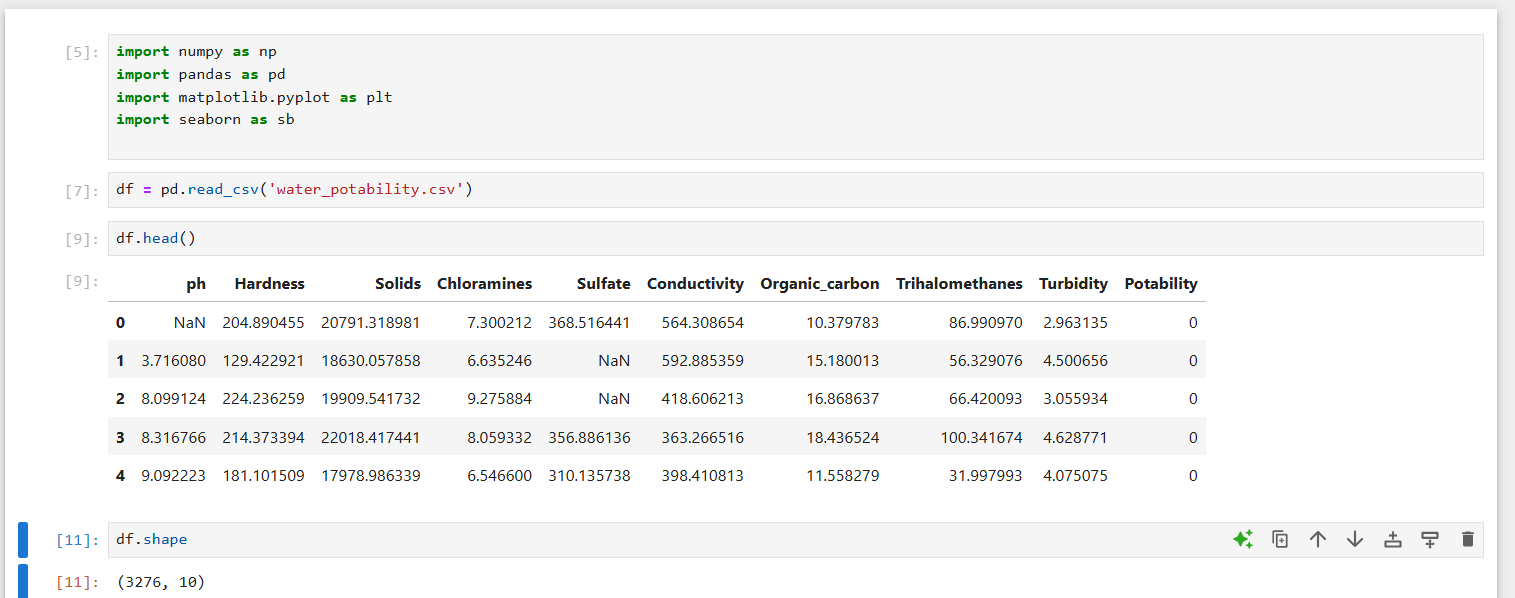
1. **Turbidity:**

* Description: Measures water clarity.
* Units: NTU (Nephelometric Turbidity Unit).
* Missing Values: No.
* Importance: High turbidity can indicate contamination.

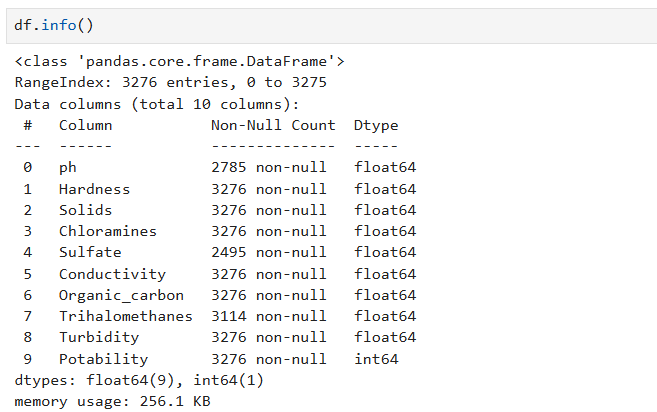
1. **Potability:**

* Description: Indicates if water is safe to drink.
* Values: 1 = Potable (safe), 0 = Not potable (unsafe).
* Missing Values: No.
* Importance: Determines overall water quality.

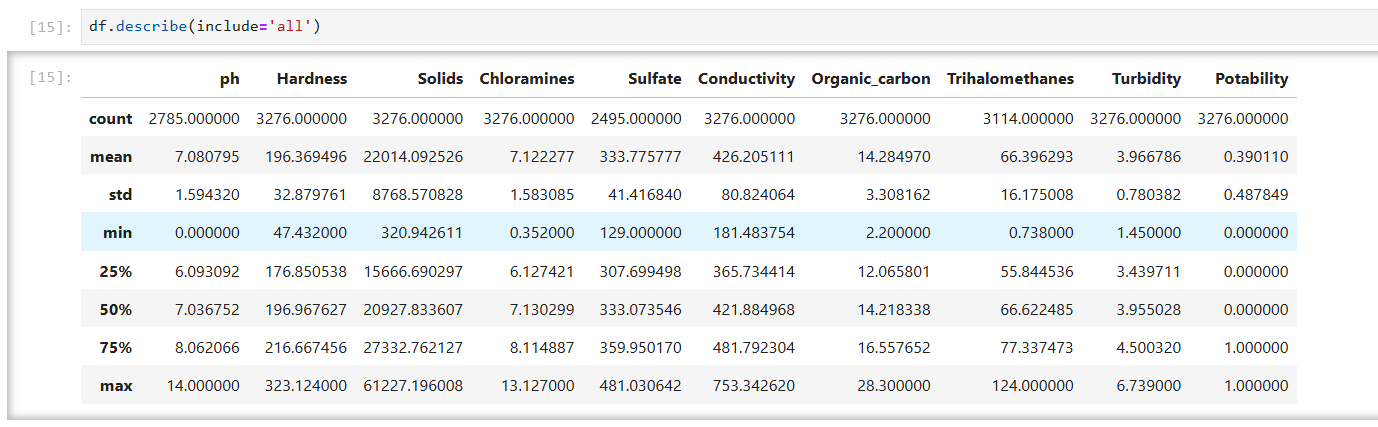
**Imports library and dataset:-**



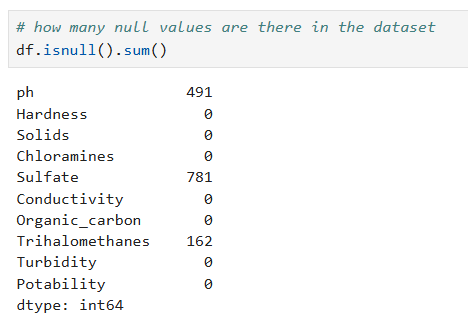
**Information of dataset:-**



**Describe the dataset :-**

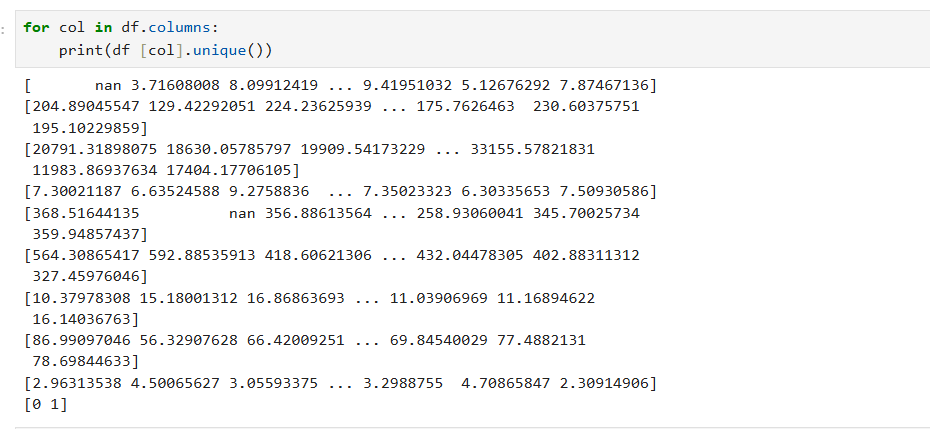


**How many null values are there in the dataframe:-**





**Checking Unique Values in each column of a dataframe:-**



**Handling Missing Values in dataframe:-**

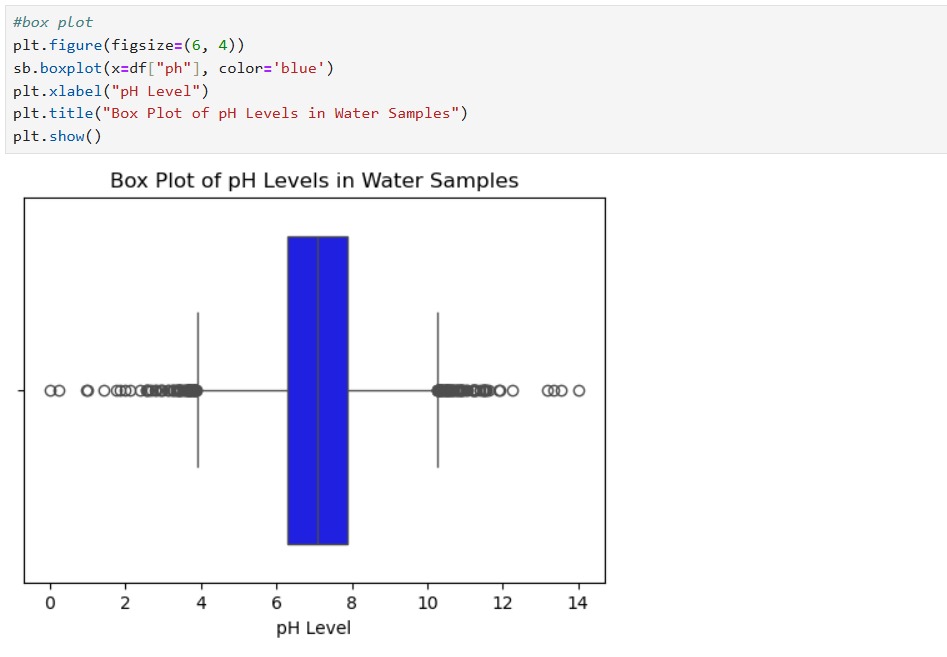


**Visualizing Data Distribution Using Histograms:-**

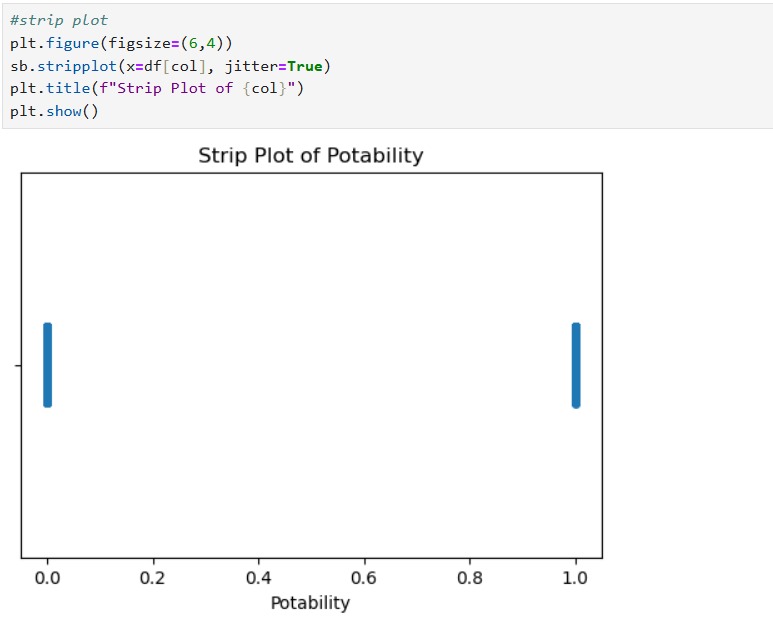


**Univariate Analysis of dataset**

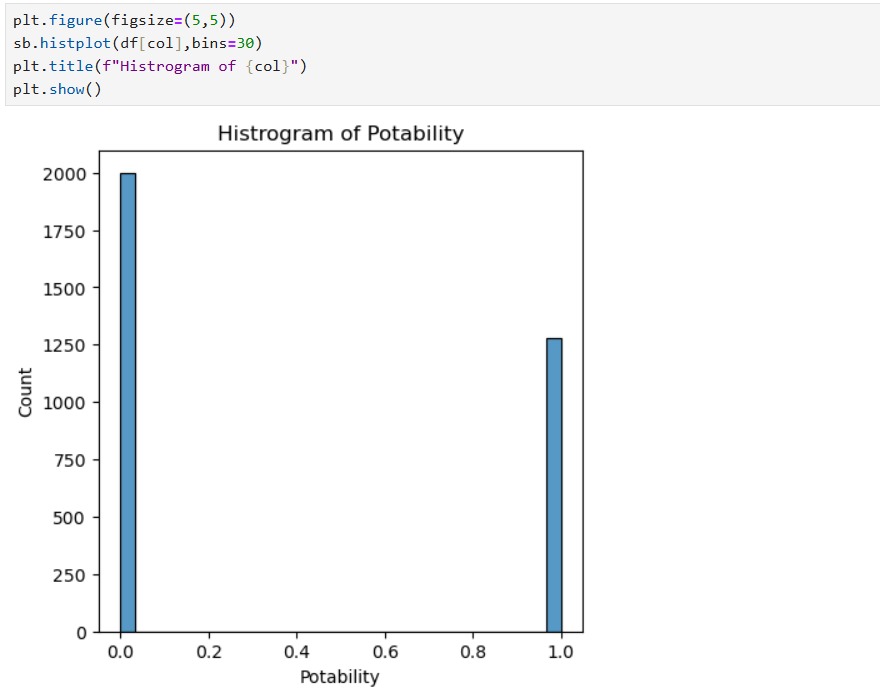
**# Box plot:-**



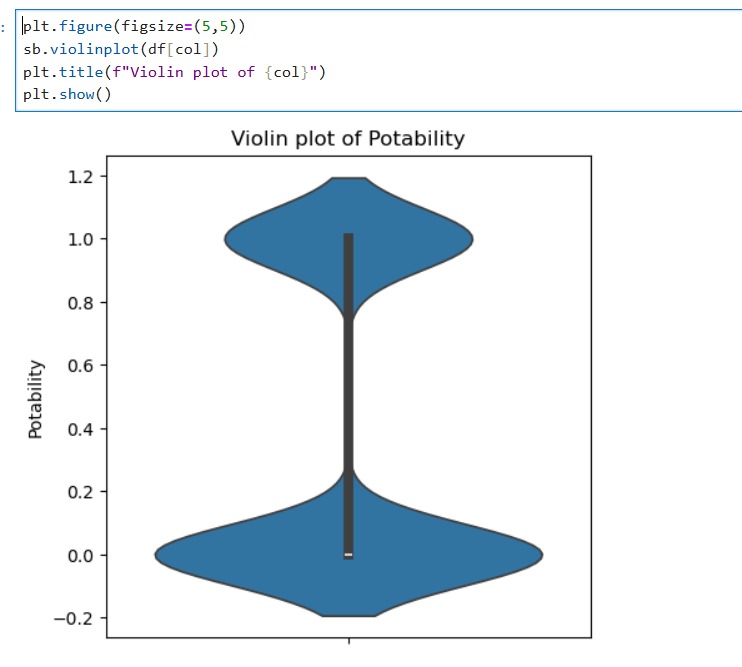
**# Strip plot:-**



**# Histrogram plot:-**



**# Violin graph**



**# Outliers:-**

