**PHASE:3**

**Water Quality Analysis**

## TEAM MEMBERS:

NAME: Archana R REG NO:721221104007

NAME: Illakiya R K REG NO:721221104020

NAME: LitheeshKumar G REG NO:7212211040032

NAME: Mohammed Harshad N

REG NO:721221104038

NAME: Nathish T

REG NO:721221104041

# Introduction:

Water quality analysis is a critical aspect of environmental science and public health, aiming to assess the safety and health of water sources for various purposes, such as drinking, agriculture, industrial use, and aquatic ecosystems. Data analytics plays a vital role in this field by enabling the collection, processing, interpretation, and visualization of data related to water quality. It helps researchers, environmentalists, and policymakers make informed decisions about managing water resources and protecting public health.

# Data Collection:

* Water quality data is collected from various sources, including rivers, lakes, reservoirs, groundwater,and treatment facilities.
* Data may include measurements of physical, chemical, and biological parameters,such as such temperature,PH,turbidity, dissolved oxygen, nutrients, heavy metals, and microbial contaminants.
* Sensors, monitoring stations, and sampling methods are used to collect data over time, providing a detailed picture of water quality dynamics.

**DatasetLink:** [**https://www.kaggle.com/datasets/adityakadiwal/water-potability**](https://www.kaggle.com/datasets/adityakadiwal/water-potability)

**Data Preprocessing and Cleaning:**

* + Clean the collected data to ensure its quality and accuracy.

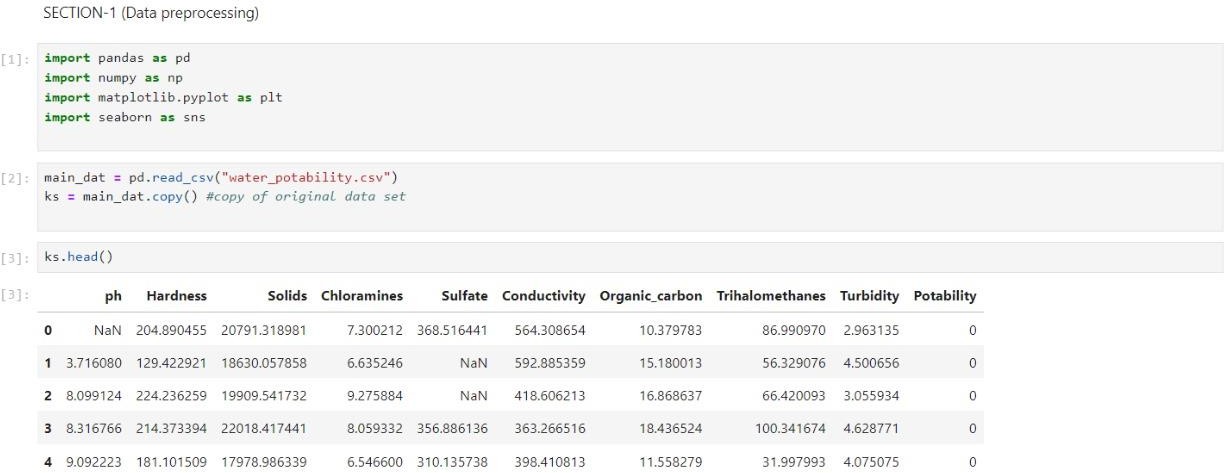
**#importing data set** import pandas as pd import numpy as np

import matplotlib.pyplot as plt import seaborn as sns

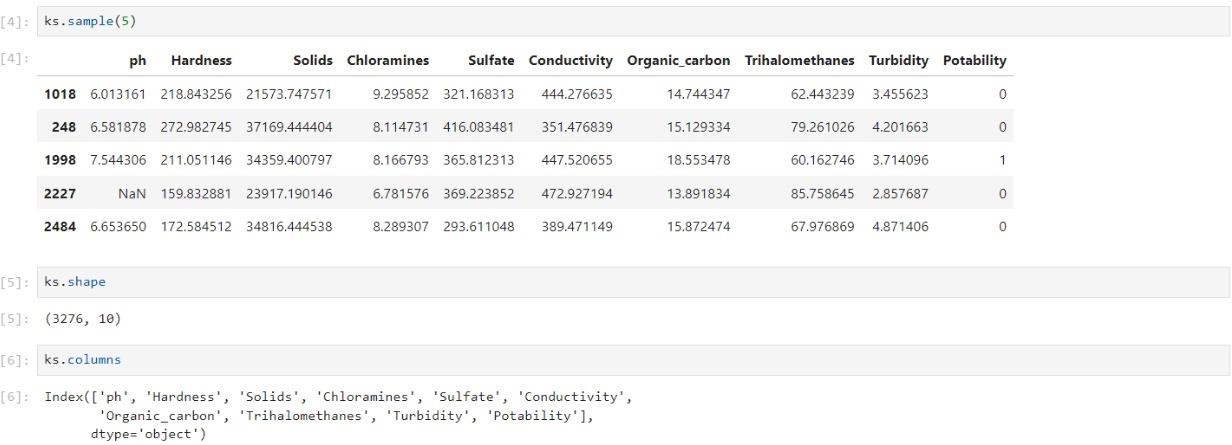
main\_dat = pd.read\_csv("water\_potability.csv")

ks = main\_dat.copy() #copy of original data set ks.head()

## OUTPUT:

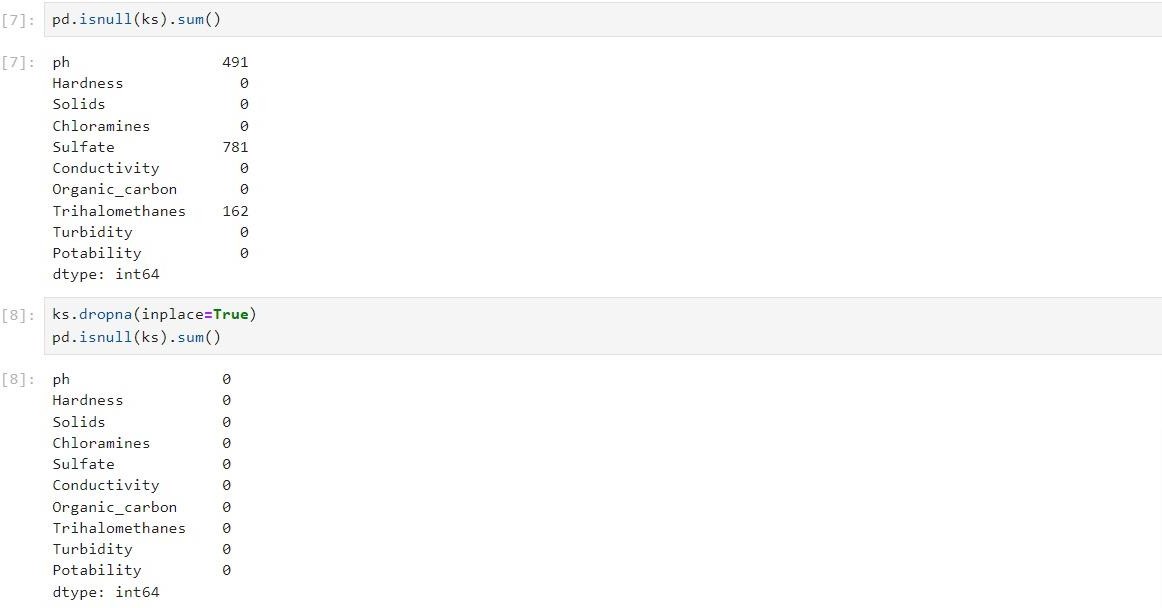


ks.sample(5) ks.shape ks.columns **OUTPUT:**



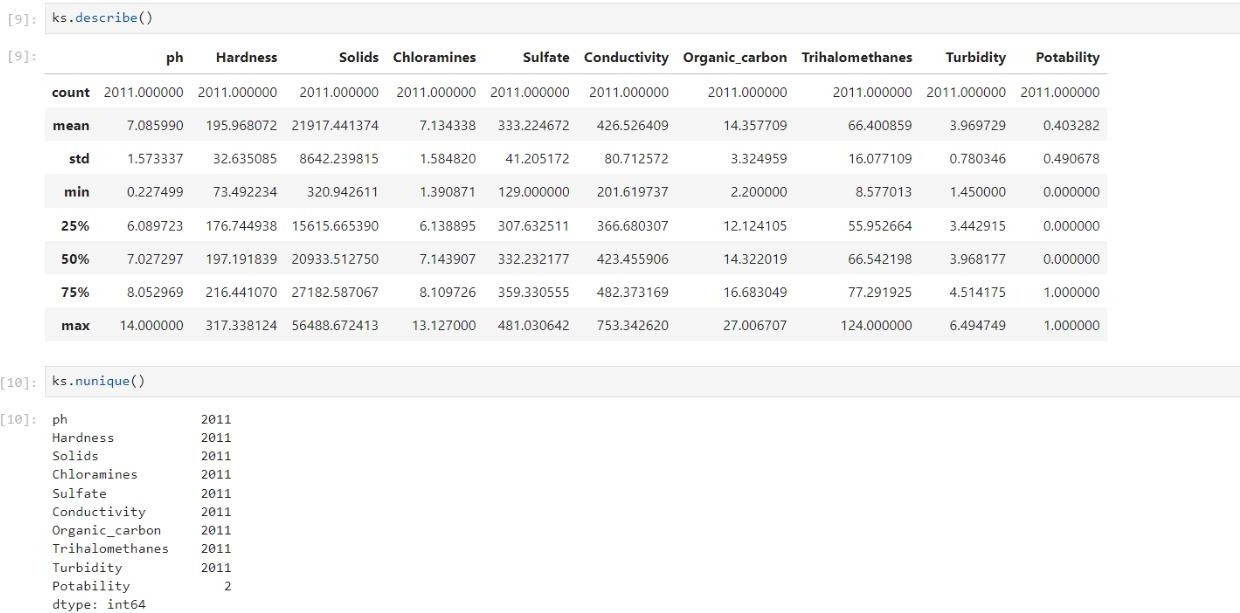
pd.isnull(ks).sum() ks.dropna(inplace=True) pd.isnull(ks).sum()

## OUTPUT:

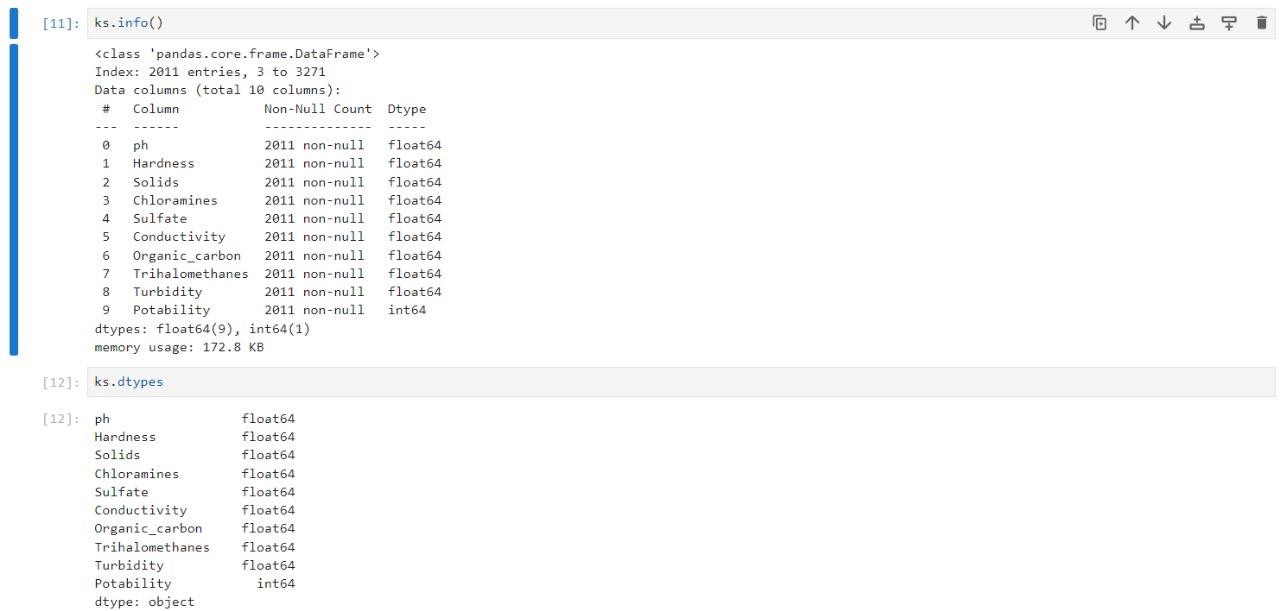


ks.describe() ks.nunique()

## OUTPUT:

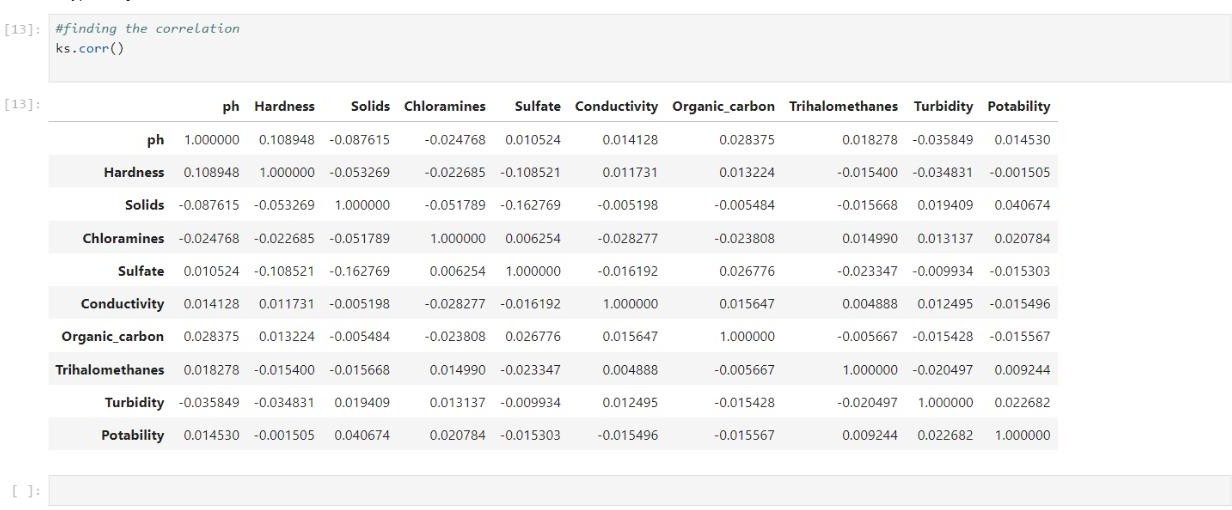


ks.info() ks.dtypes **OUTPUT:**



#finding the correlation ks.corr()

**OUTPUT:**



# Project Conclusion:

In conclusion, data analytics is an indispensable tool in the field of water quality analysis, enabling us to gain valuable insights into the health and safety of water sources. By harnessing the power of data analytics, we can make informed decisions that impact public health, environmental sustainability, and resource management.