PHASE:3

Water Quality Analysis

TEAM MEMBERS:

NAME:

R.ALCINSNOWLINA

REG NO:

721221106009

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/aditvakadiwal/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:

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
1018	6.013161	218.843256	21573.747571	9.295852	321.168313	444.276635	14.744347	62.443239	3.455623	0
248	6.581878	272.982745	37169.444404	8.114731	416.083481	351.476839	15.129334	79.261026	4.201663	0
1998	7.544306	211.051146	34359.400797	8.166793	365.812313	447.520655	18.553478	60.162746	3.714096	1
2227	NaN	159.832881	23917.190146	6.781576	369.223852	472.927194	13.891834	85.758645	2.857687	0
2484	6.653650	172.584512	34816.444538	8.289307	293.611048	389.471149	15.872474	67.976869	4.871406	0
ks.s	hape									
(327	6, 10)									
ks.c	olumns									

pd.isnull(ks).sum()

ks.dropna(inplace=True)

pd.isnull(ks).sum()

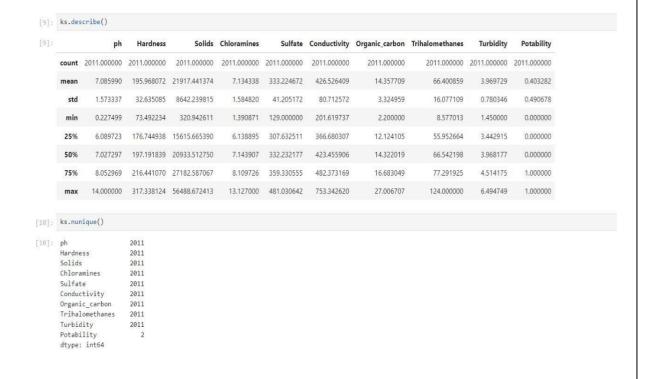
OUTPUT:

```
[7]: pd.isnull(ks).sum()
    Hardness
                     0
    Solids
    Chloramines
    Sulfate
                   781
    Conductivity
    Organic_carbon
     Trihalomethanes 162
                  0
     Turbidity
     Potability
    dtype: int64
[8]: ks.dropna(inplace=True)
    pd.isnull(ks).sum()
[8]: ph
    Hardness
                     0
    Solids
     Chloramines
    Sulfate
    Conductivity
    Organic_carbon
     Trihalomethanes 0
     Turbidity
     Potability
    dtype: int64
```

ks.describe()

ks.nunique()

OUTPUT:



ks.info()

ks.dtypes

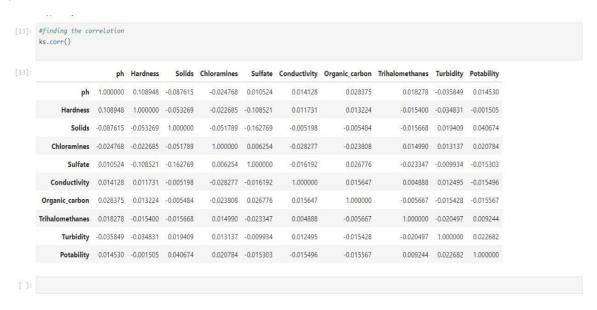
OUTPUT:

```
[11]: ks.info()
                                                                                                                                    回↑↓古早盲
      <class 'pandas.core.frame.DataFrame'>
       Index: 2011 entries, 3 to 3271
      Data columns (total 10 columns):
       # Column
                           Non-Null Count Dtype
                            2011 non-null
                                           float64
           Hardness
                            2011 non-null
                                           float64
           Solids
                            2011 non-null
                                           float64
           Chloramines
                            2011 non-null
           Sulfate
                            2011 non-null
                                           float64
           Conductivity
                                           float64
                            2011 non-null
           Organic_carbon
                           2011 non-null
           Trihalomethanes 2011 non-null
                                           float64
           Turbidity
                            2011 non-null
                                           float64
           Potability
                            2011 non-null
       dtypes: float64(9), int64(1)
       memory usage: 172.8 KB
[12]: ks.dtypes
[12]: ph
                         float64
       Hardness
                         float64
                         float64
      Solids
      Chloramines
                         float64
       Sulfate
                         float64
      Conductivity
                         float64
      Organic_carbon
                         float64
       Trihalomethanes
       Turbidity
                         float64
      Potability
                           int64
      dtype: object
```

#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







