WATER QUALITY ANALYSIS

Project Title: Water quality Analysis

Phase 3: Development part-1

Topic: Start building the water quality analysis by preprocessing the data and performing exploratory data analysis.

Phase 3: submission document

WATER QUALITY ANALYSIS

Introduction:

- Water quality analysis is a critical field of study and application that involves assessing the chemical, physical, biological, and microbiological characteristics of water to determine its suitability for various purposes. This analysis plays a vital role in ensuring the safety and sustainability of water resources for human consumption, agriculture, industry, and aquatic ecosystems. It helps identify potential contaminants, pollutants, and threats to public health and the environment.

- Water quality analysis employs various techniques and instruments, including chemical tests, spectrometry, microbiological assays, and sensors to measure parameters like pH, turbidity, temperature, dissolved oxygen, nutrient concentrations, heavy metals, and more. Continuous advancements in technology and data analysis tools have made water quality monitoring more efficient and accurate, enabling prompt responses to potential water quality issues.

- In summary, water quality analysis is indispensable for safeguarding human health, protecting the environment, and ensuring the sustainability of water resources in various applications, from drinking water to industrial processes and ecological conservation.

Given data set:

0 – 14	47.4 – 323	321 – 61.2k	0.35 – 13.1	129 – 481	18
3.71608007538699	129.42292051494425	18630.057857970347	6.635245883862		59
8.099124189298397	224.23625939355776	19909.541732292393	9.275883602694089		41
8.316765884214679	214.37339408562252	22018.417440775294	8.05933237743854	356.88613564305666	36
9.092223456290965	181.10150923612525	17978.98633892625	6.546599974207941	310.13573752420444	39
5.584086638456089	188.3133237696164	28748.68773904612	7.54486878877965	326.6783629116736	28
10.223862164528773	248.07173527013992	28749.716543528233	7.5134084658313025	393.66339551509645	28
8.635848718500734	203.36152258457054	13672.091763901635	4.563008685599703	303.3097711592812	47
	118.98857909025189	14285.583854224515	7.804173553073094	268.646940746221	38
11.180284470721592	227.23146923797458	25484.50849098786	9.077200016914393	404.04163468408996	56
7.360640105838258	165.52079725952862	32452.614409143884	7.550700906704114	326.62435345560164	42
7.974521648923869	218.69330048866644	18767.65668181348	8.110384501123875		36

Sulfate	# Conductivity	# Organic_carbon	# Trihalomethanes	# Turbidity
ount of Sulfates solved in mg/L	Electrical conductivity of water in μS/cm	Amount of organic carbon in ppm	Amount of Trihalomethanes in μg/L	Measure of light emiting property of water in NTU (Nephelometric Turbidity Units)
– 481	181 – 753	2.2 – 28.3	0.74 – 124	1.45 – 6.74
3.51644134980336	564.308654172439	10.3797830780847	86.99097046150888	2.96313538063164 07
	592.8853591348523	15.180013116357259	56.32907628451764	4.500656274942408
	418.6062130644815	16.868636929550973	66.42009251176368	3.0559337496641685
.88613564305666	363.2665161642437	18.436524495493302	100.34167436508008	4.628770536837084
.13573752420444	398.41081338184466	11.558279443446395	31.997992727424737	4.075075425430034
.6783629116736	280.4679159334877	8.399734640152758	54.917861841994466	2.5597082275565217
.66339551509645	283.6516335078445	13.789695317519886	84.60355617402357	2.672988736934779
.3097711592812	474.60764494244853	12.36381669870525	62.798308962925155	4.401424715445482
.646940746221	389.3755658712614	12.70604896865791	53.928845767512236	3.595017180957615 5
.04163468408996	563.8854814810949	17.92780641128502	71.9766010322191 5	4.37056193665549 7

Necessary step to follow:

outline of the steps to obtain and preprocess a water quality dataset, as well as conduct exploratory data analysis (EDA). Here's a step-by-step guide:

1. Data Collection:

   - First, you need to obtain a water quality dataset. You can find such datasets from various sources, including government agencies, research organizations, or open data repositories.

2. Data Import:

   - Import the dataset into your preferred data analysis tool (e.g., Python with pandas, R, or any other tool you are comfortable with).

3. Data Exploration:

   - Begin by exploring the dataset to understand its structure and contents. Use functions like `head()`, `info()`, and `describe()` to get a feel for the data.

4. Handling Missing Values:

   - Identify missing values in the dataset. You can use functions like `isna()`, `isnull()`, or `info()` to find missing values. Depending on the extent of missing data, you can consider different strategies:

     - Remove rows or columns with a high percentage of missing values.

     - Impute missing values using methods such as mean, median, or machine learning-based imputation techniques.

Program:

```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
# Load the dataset
data = pd.read_csv("water_quality_dataset.csv")
# Data Exploration
print(data.head())  # View the first few rows
print(data.info())  # Get data info, check for missing values
# Handling Missing Values (if needed)
data.dropna(inplace=True)  # Removing rows with missing values


# Handling Outliers (if needed)

# Use statistical methods or visualizations to identify and address outliers

# Data Visualization

# Visualize parameter distributions
plt.figure(figsize=(12, 6))
sns.set(style="whitegrid")
sns.histplot(data['parameter_of_interest'], kde=True)
plt.title('Distribution of Parameter of Interest')
plt.xlabel('Parameter Values')
plt.ylabel('Frequency')
plt.show()


# Correlation Analysis
correlation_matrix = data.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```

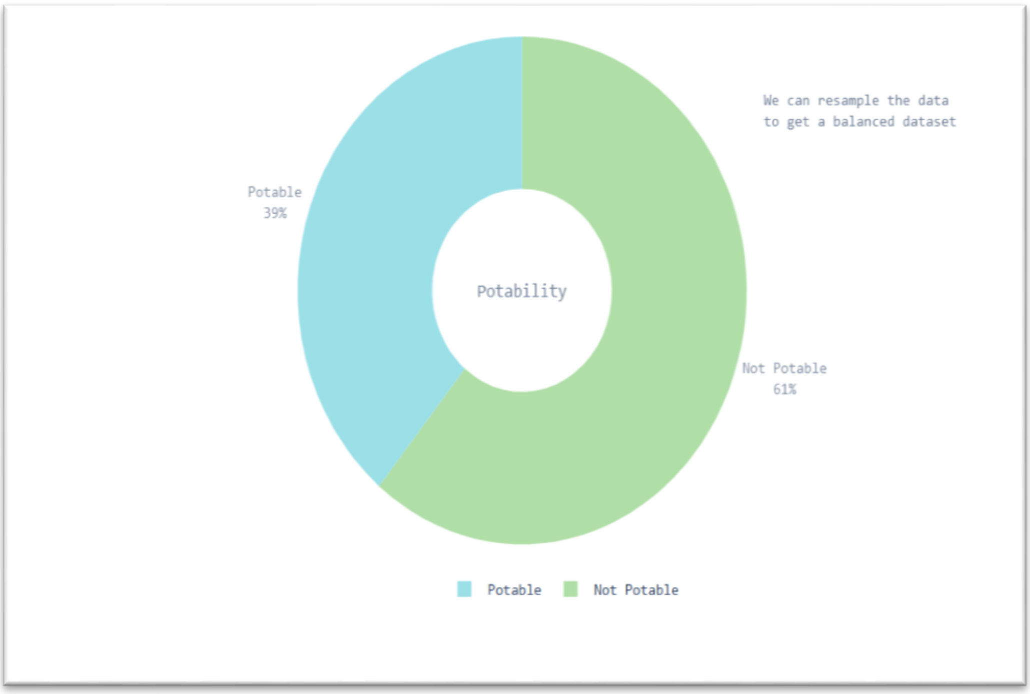
OUT:

 # Deviations from Standards

 # Compare the data to water quality standards (e.g., using a reference dataframe)

reference_data = pd.read_csv("water_quality_standards.csv") # Load standards data

OUT:



Merge the reference data with your water quality data (assuming common columns)

merged_data = pd.merge(data, reference_data, on='common_column')

Compare parameters with standards

deviations = merged_data[merged_data['parameter_of_interest'] > merged_data['standard']]

Visualize deviations

plt.figure(figsize=(12, 6))

sns.scatterplot(x=deviations.index, y='parameter_of_interest', data=deviations, hue='parameter_name')

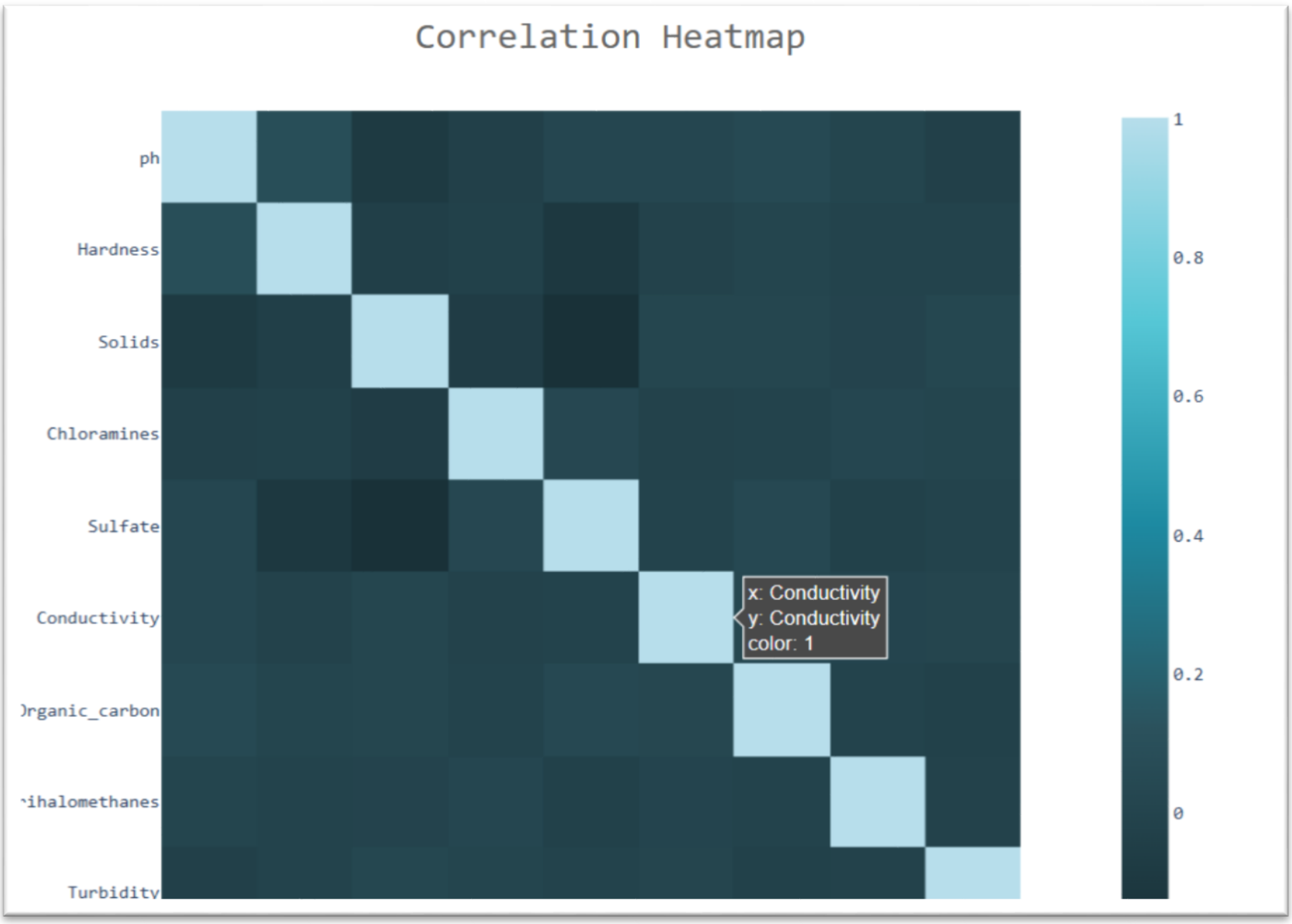
plt.title('Deviations from Standards')

plt.xlabel('Data Points')

plt.ylabel('Parameter Values')

plt.show()

OUT:



CONCLUSION:

In conclusion, harnessing data analytics with Cognas for water quality analysis is a powerful approach that brings several key benefits:

1. Enhanced Decision-Making: The integration of Cognas and data analytics allows for more informed and data-driven decision-making. Through advanced analytics and visualization tools, water quality data can be transformed into actionable insights, helping authorities and organizations respond to water quality issues promptly.

2. Predictive Maintenance: With the help of Cognas, predictive analytics can be applied to anticipate equipment failures or water quality fluctuations. This proactive approach enables maintenance teams to address issues before they become critical, reducing downtime and improving overall water quality management.

3. Data Integration:Cognas can seamlessly integrate data from various sources, including IoT sensors, historical records, and environmental data. This holistic view of water quality parameters enables a more comprehensive analysis, helping identify complex correlations and factors affecting water quality.

4. Efficiency and Automation:Cognas-based data analytics can automate routine data processing and analysis tasks, saving time and resources. It allows for continuous monitoring and real-time alerts, improving operational efficiency and reducing the risk of human error.

However, it's important to be mindful of potential challenges such as data privacy and security, ensuring the availability of reliable data sources, and providing adequate training and resources for staff to effectively utilize Cognas and data analytics tools.

In summary, the use of Cognas in water quality analysis through data analytics offers a multifaceted approach to monitoring, analyzing, and improving water quality. It empowers organizations to make well-informed decisions, ensure the efficiency of water quality management, and ultimately contribute to the protection of this vital resource for both environmental and human well-being.