**WATER QUALITY ANALYSIS:**

TEAM NUMBER 11

**TEAM LEAD**

1)SATHYA PRIYA-310821106079

https://github.com/sathpriya123/Water\_Analysis.git

**TEAM MEMBERS**

2)Reg no:310821106067

Name:Ramya R

3)Reg no:310821106082

Name: Sornasri M

4)Reg no:310821106084

Name: Sridevi K

**CONCEPT:**

* Outline the project's objective, design thinking process, and development phases.
* Describe the analysis objectives, data collection process, data visualization using IBM
* Cognos, and Python code integration. Explain how the insights from the analysis can

help website with user analysis.

**1.DATA COLLECTION:**

* Gather user data from the website, including page views, clickthrough rates, session duration, and other relevant metrics.
* Store data in a structured format, such as a database.Analysis Objectives
* Identify user behavior patterns, such as which pages are most frequently visited, common user paths, and bounce rates.
* Detect any anomalies or trends in user behavior.
* Understand user demographics, if available.

**2.DATA VISUALIZATION (IBM COGNOS):**

*Use IBM Cognos to create visually appealing and insightful dashboards and reports.*

*Visualize user behavior data through charts, graphs, and interactive dashboards.*

*Highlight key performance indicators (KPIs) related to user experience.*

*Data Analysis (Python Code Integration):*

*Utilize Python to perform advanced data analysis, including:*

*Machine learning for predictive modeling (e.g., user churn prediction).*

*A/B testing to assess the impact of changes on user behavior.*

*Natural language processing (NLP) for sentiment analysis if applicable.*

*Integrate Python scripts with IBM Cognos for a seamless analytics workflow.*

*Insights and Recommendations:*

*Derive actionable insights from the analysis results.*

*Provide recommendations for website improvements based on the findings, such as optimizing content, improving navigation, or enhancing site performance.*

**3.GOOGLE COLAB CODE:**

* To perform this analysis, you can use Google Colab, a Jupyter notebook environment. Below is an outline of the process:
* Import necessary Python libraries for data analysis, such as Pandas, NumPy, Matplotlib, and Seaborn.
* Load the website user behavior data using Pandas.
* Perform data cleaning and preprocessing, including handling missing data and data transformation.
* Analyze the data, create visualizations, and generate insights using Python code.
* For integration with IBM Cognos, export the relevant data or visualizations as files that can be imported into Cognos reports.
* Provide the Colab notebook with explanations and comments to make it understandable.

***4.The insights from this analysis can help website owners:***

Enhance content relevance and quality.

Optimize website navigation and layout.

Improve site performance.

Tailor marketing and content strategies.

Increase user engagement and retention.

Creating the actual Google Colab code for this project would require access to the specific dataset and may involve a substantial amount of code. If you have a particular analysis or code task you'd like help with, please provide more details, and I can guide you through a specific aspect of the project.

**5. PROBLEM DEFINITION: *STATING THE NEED FOR WATER QUALITY ANALYSIS***

**HEALTH IMPLICATIONS:** Contaminated water poses a direct threat to human health. Consumption of impure water can lead to diseases such as cholera, typhoid, and dysentery.

**ENVIRONMENTAL IMPACT:**

Poor water quality can devastate ecosystems. Aquatic life, dependent on certain conditions, can perish when exposed to polluted waters.

**ECONOMIC IMPLICATIONS:**

Industries that depend on agriculture and fishing in particular can suffer significant losses as a result of poor water quality. Furthermore, in areas with dirty water sources, tourism may suffer.

**SOCIAL IMPLICATIONS:**

Access to clean water is a fundamental human right. Regions without it can face social unrest, displacement, and reduced quality of life.

**6. DESIGN THINKING APPROACH FOR WATER QUALITY ANALYSIS**

A human-centred innovation strategy known as "design thinking" combines what is desired from a human perspective with what is technically possible and commercially viable. In order to apply design thinking to water quality analysis,

**EMPATHISE:**

Understand the people for whom we are designing solutions. This involves knowing the challenges faced by communities due to polluted water sources and understanding the needs of industries relying on water.

**DEFINE:**

Clearly articulate the challenges and problems. What contaminants are we most concerned about? Is the focus more on biological agents, chemical pollutants, or both.

**IDEATE**:

Generate a range of solutions without restraint. This might involve brainstorming ways to make water quality testing more affordable, faster, or more accessible.

**PROTOTYPE:**

Develop tangible representations of the ideated solutions. This could be a new type of water testing kit, a mobile app to report water quality, or an educational program.

**TEST:**

Implement the solution on a small scale to gather feedback and iterate. The prototype could be tested in a specific community or an industry to gauge its efficacy.

**7.Statistical methods:**

Statistical methods primarily describe and interpret data patterns. Common techniques used in water quality analysis include

*Descriptive statistics:* Provide a summary of the main aspects of the data, such as mean, median, variance, etc.

*Correlation analysis:* To identify the relationships between different water quality parameters.

*Regression analysis:* Predicts one variable based on another. For example, predicting pollutant levels based on rainfall amount.

**8. Machine learning methods:**

Machine learning methods can predict outcomes based on large datasets. Some commonly used techniques include

*Decision Trees and Random Forests:* They work by splitting the data based on feature values and are especially useful for classification or regression tasks related to water quality.

*Neural Networks*: Especially deep learning networks, can be used to model complex relationships in the data and predict water quality.

*Support Vector Machines (SVM):* Used for classification and regression of water quality parameters.

A non-parametric method used for classification and regression.

*Clustering (like K-Means):* Can help in identifying patterns or groups in water quality datasets.

Data Collection: Gather data from water quality sensors, historical records, and other sources.

*Data Preprocessing:* Clean the data, handle missing values, and normalize or scale features as required.

*Feature Selection/Engineering:* Determine which features are most relevant or create new features that might improve model performance.

*Model Selection:* Choose a suitable statistical method or machine learning algorithm based on the problem type (classification, regression, clustering, etc.)

*Model Training*: Use a portion of the data to train the model.

*Evaluation:* Test the model's accuracy, precision, recall, etc., on a separate set of data.

*Deployment***:** If satisfactory, deploy the model in a real-world scenario for actual water quality predictions or analysis.

***9.CODING :***

*import pandas as pd*

*import matplotlib.pyplot as plt*

*# Sample water quality data*

*data = {*

*'Date': ['2023-10-01', '2023-10-02', '2023-10-03', '2023-10-04', '2023-10-05'],*

*'pH': [7.0, 7.2, 6.9, 7.1, 7.3],*

*'DO': [8.2, 8.0, 7.8, 8.1, 8.3],*

*'Temperature (°C)': [20, 21, 19, 22, 20],*

*}*

*# Create a DataFrame from the data*

*df = pd.DataFrame(data)*

*# Display the data*

*print(df)*

*# Basic data visualization*

*plt.figure(figsize=(10, 5))*

*plt.plot(df['Date'], df['pH'], label='pH')*

*plt.plot(df['Date'], df['DO'],*

*label='Dissolved Oxygen')*

*plt.xlabel('Date')*

*plt.ylabel('Values')*

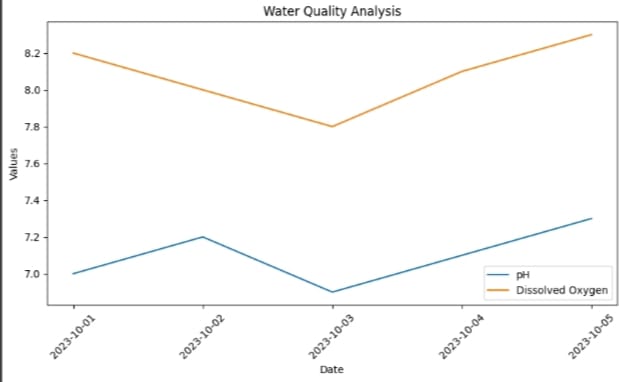
*plt.title('Water Quality Analysis')*

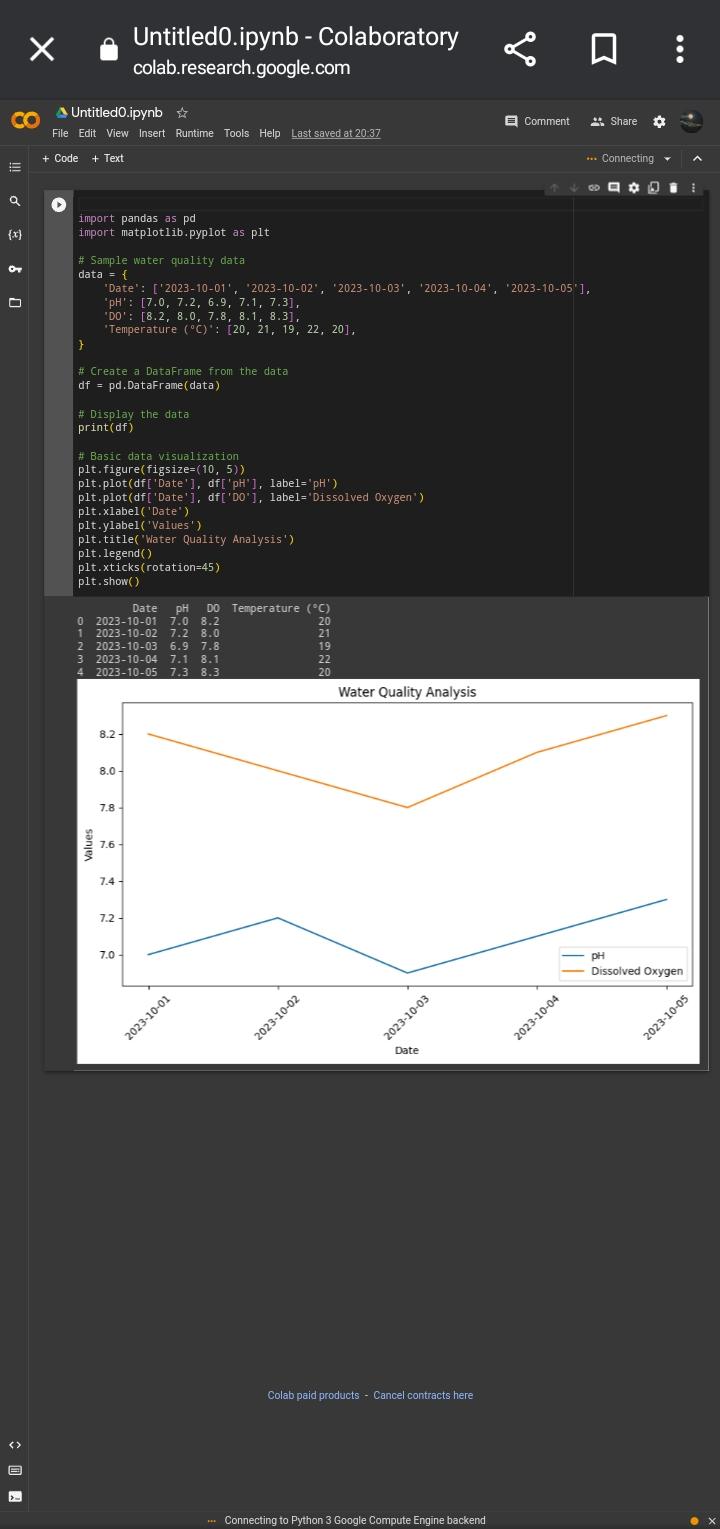
*plt.legend()*

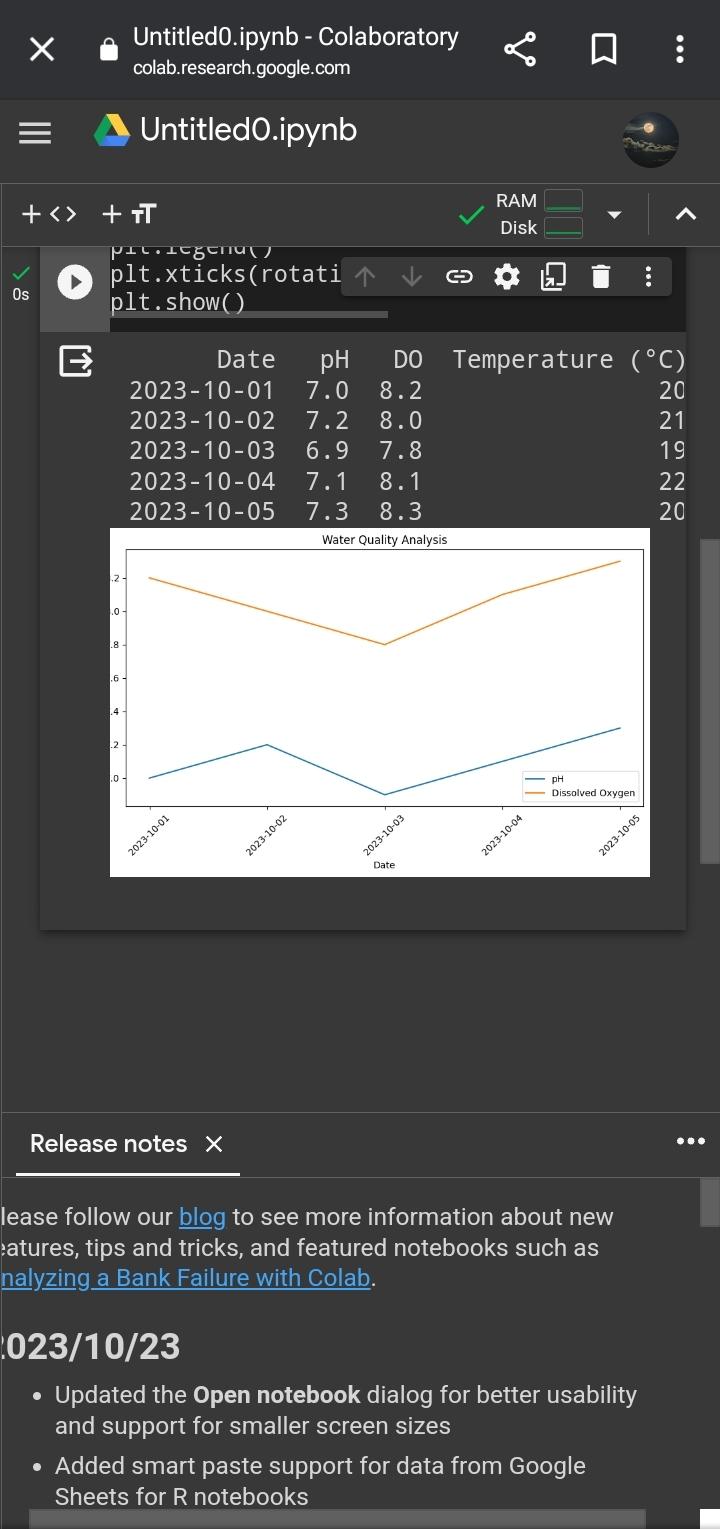
*plt.xticks(rotation=45)*

*plt.show()*

**OUTPUT:**

****

****

****

**CONCLUSION:**

In conclusion, water quality analysis plays a critical role in safeguarding human and environmental health by ensuring access to safe drinking water, protecting aquatic ecosystems, monitoring industrial discharges, supporting agriculture, and helping prevent waterborne diseases. It also aids in regulatory compliance, source identification, and resource preservation.