TAMILNADU MARGINAL WORKERS ASSESSMENT

Water Quality Analysis – Phase 5 DOCUMENTATION

Team Members:

1.Darshini.N(au613021205003)

2.Dhanusree.K(au613021205008)

3.Vethasundari.N(au613021205059)

4.Ramya.R(au613021205039)

5.Swetha.E(au613021205056)

OBJECTIVE:

The objectives for conducting water quality analysis can vary based on specific needs, research goals, or environmental concerns. Here are common objectives for water quality analysis:

1. Assessment of Potability:

Determine if the water meets the standards for human consumption and is safe to drink without causing health risks.

2. Environmental Monitoring:

Evaluate the impact of human activities or natural occurrences on water bodies, such as rivers, lakes, or oceans, to maintain ecological balance.

3. Compliance with Regulations:

Ensure that water quality adheres to established local, national, or international standards and regulations set by environmental agencies.

4. Identification of Contaminants:

Detect and identify various pollutants, chemicals, heavy metals, pathogens, or other contaminants present in the water that might affect human health or the ecosystem.

5. Source Identification:

Determine the sources of pollution or contamination in the water, helping in the planning and implementation of remediation or mitigation measures.

6.Trend Analysis:

Identify trends and changes in water quality over time, assisting in understanding seasonal variations, long-term alterations, or the impacts of specific events.

7. Risk Assessment and Management:

Assess potential risks associated with water quality issues and develop strategies to manage or mitigate those risks effectively.

8. Public Health Protection:

Protect public health by ensuring that recreational waters, drinking water supplies, and agricultural water sources are safe and free from harmful contaminants.

9. Community Awareness and Education:

Provide information to communities about the importance of water quality and its impact on health, ecosystems, and overall well-being.

10. Support Policy Development:

Contribute data and analysis to support the development of water quality management policies and regulations at local, regional, or national levels.

DESIGN THINKING CHALLENGES:

Design thinking can be an effective approach in the field of water quality analysis to address complex challenges and find innovative solutions. The design thinking process involves several stages, each with its unique focus, and can be adapted to address water quality analysis effectively. Here's how the design thinking process might be applied in this context:

1. Empathize:

Understand the needs and concerns of stakeholders, including local communities, water authorities, environmentalists, and policymakers. Engage with them to comprehend their challenges and aspirations regarding water quality.

2. Define:

Clearly define the problem areas related to water quality. Identify specific issues, such as contamination sources, pollutant types, regulatory challenges, or public health concerns. Formulate precise problem statements that focus on improving water quality parameters.

3. Ideate:

Brainstorm and generate a wide range of potential solutions to address the identified problems. Encourage diverse thinking from multidisciplinary teams or stakeholders. This can involve innovative monitoring methods, novel data analysis techniques, or creative strategies for public engagement.

4. Prototype:

Develop prototypes or models to test and validate the proposed solutions. This might involve creating experimental setups for water quality testing, designing new data visualization tools, or implementing pilot projects to assess the feasibility of proposed solutions.

5. Test:

Test the prototypes and solutions in real or simulated environments. Collect and analyze data to evaluate their effectiveness in improving water quality. Gather feedback from users, stakeholders, and experts to refine and improve the solutions.

6. Implement:

Based on the successful test results, move forward with the implementation of the most effective solutions. This might involve the deployment of new technologies, policies, or methods for water quality analysis, management, and public awareness.

7. Iterate:

Continuous improvement is key. Gather feedback, monitor the implemented solutions, and be ready to iterate or adapt based on ongoing evaluations and changing environmental or regulatory conditions.

By applying design thinking principles in water quality analysis, you can foster a more human-centric and innovative approach to addressing challenges. This process can lead to the development of more effective, user-friendly, and sustainable solutions to improve and maintain water quality.

DEVELOPMENT PHASES:

In the context of water quality analysis, the development process typically involves various phases, including planning, data collection, analysis, and reporting. Here's an overview of the development phases for water quality analysis:

1. Problem Identification and Planning:

Objective Setting: Define the specific goals and objectives of the water quality analysis. Determine what aspects of water quality you intend to assess or improve.

Resource Planning: Allocate resources, including personnel, equipment, and budget, needed for the analysis. Plan the timeline and milestones for the project.

2. Data Collection:

Sampling Design: Develop a systematic sampling plan based on the objectives. Decide on the sampling locations, frequency, and methods to collect water samples.

Data Acquisition: Collect water samples from various sources such as rivers, lakes, groundwater, or treatment plants. Ensure proper storage and labeling of samples.

3. Data Preprocessing and Quality Control:

Sample Processing: Conduct initial processing of collected samples, including filtration or preservation, as necessary.

Quality Assurance/Quality Control (QA/QC): Perform quality control measures to ensure the accuracy and reliability of collected data. This includes calibration of instruments, blanks, and duplicate samples.

4. Laboratory Analysis:

Testing Procedures: Analyze the collected water samples in a laboratory using appropriate testing methods for various water quality parameters, such as pH, dissolved oxygen, turbidity, and specific contaminants.

Data Recording: Record and manage the data obtained from laboratory analysis accurately.

5. Data Interpretation and Analysis:

Data Evaluation: Interpret the results of the laboratory analysis to assess compliance with regulatory standards or specific project objectives.

Statistical Analysis: Apply statistical methods to analyze the data, identify trends, correlations, outliers, and patterns in the water quality parameters.

6. Report Generation and Communication:

Report Preparation: Document the findings, methodology used, results, and conclusions in a comprehensive report.

Visualization and Communication: Use data visualization techniques to present the results in an understandable and informative manner. Communicate findings to stakeholders, decision-makers, or the public.

7. Recommendations and Action:

Actionable Insights: Provide recommendations or action plans based on the analysis. This may involve suggesting corrective measures, policy recommendations, or further investigation.

8. Monitoring and Feedback:

Continual Monitoring: Implement monitoring programs for ongoing assessment of water quality parameters. Collect feedback on the effectiveness of any recommended actions and adjust strategies as needed.

ANALYSIS OBJECTIVES:

In water quality analysis, the objectives are defined goals that guide the assessment, evaluation, and understanding of the quality and characteristics of water. These objectives help in determining the purpose of the analysis, guiding the collection of data, and formulating strategies for managing and maintaining water quality. Some common objectives in water quality analysis include:

1. Assessment of Potability:

Determine if the water meets standards for human consumption and is safe to drink without causing health risks.

2. Environmental Monitoring:

Assess the impact of human activities or natural occurrences on water bodies, such as rivers, lakes, or oceans, to maintain ecological balance.

3. Compliance with Regulations:

Ensure that water quality adheres to established local, national, or international standards and regulations set by environmental agencies.

4. Identification of Contaminants:

Detect and identify various pollutants, chemicals, heavy metals, pathogens, or other contaminants present in the water that might affect human health or the ecosystem.

5. Source Identification:

Determine the sources of pollution or contamination in the water, aiding in planning and implementing remediation or mitigation measures.

DATA PREPROCESSING:

Perform any other preprocessing steps that are specific to your dataset and analysis goals. This may include scaling numeric features, handling outliers, or creating new features.

Saving Preprocessed dataset:

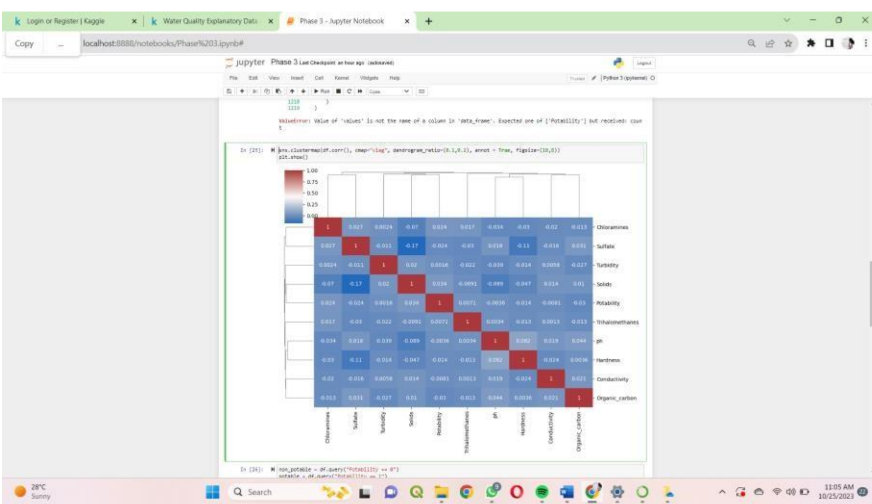
In this step, if we made substantial changes to the dataset and want to save the preprocessed version, you can use the following Code.

Code:

```
# Save thepreprocesseddatasettoanewCSVfile
df.to_csv('preprocessed_dataset.csv', index=False)
import seaborn as sns import
matplotlib.pyplot as plt
sns.heatmap(cor,annot=True,cmap='coolwarm')
plt.show()
```

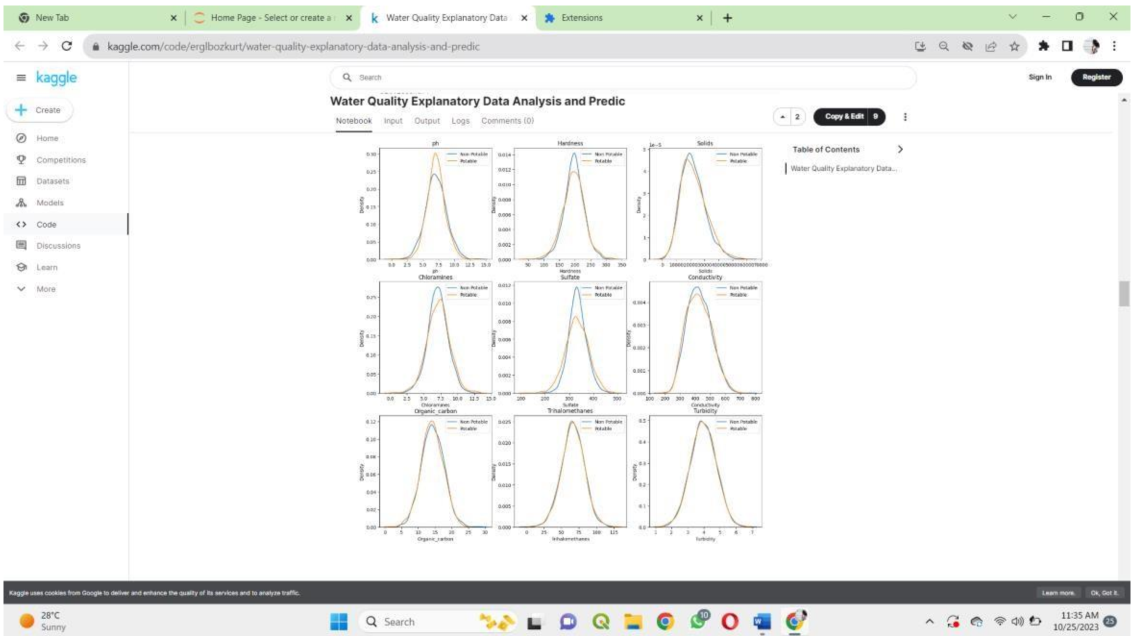
DATA VISUALIZATION:

```
sns.clustermap(df.corr(), cmap="vlag", dendrogram_ratio=(0.1,0.2), annot = True,
figsize=(10,8)) plt.show()
z
```



```
non_potable = df.query("Potability == 0")
potable = df.query("Potability == 1")
```

```
plt.figure(figsize = (15,15)) for ax, col in
enumerate(df.columns[:9]):
plt.subplot(3,3, ax + 1) plt.title(col)
sns.kdeplot(x = non_potable[col], label = "Non Potable")
sns.kdeplot(x = potable[col], label = "Potable") plt.legend()
plt.show()
```



Importing the libraries:

Here, for preprocessing the dataset and manipulate the data, pandas is the library used to frame the data.

Code:

```
import pandas as pd
```

Loading the dataset:

In this step,we are framing the data into the table using Dataframe in pandas,display the head or 5 rows of the dataset.

Code:

```
# Replace with the actual filename
```

File_path="C:\Users\darsh\Downloads\water_potability.csv"

EXPLORATORY DATA:

After framing data,the first few or five row of the data in displayed using the head() function
Code: print(data.head())

Output:

```
ph    Hardness        Solids  Chloramines     Sulfate  Conductivity  \
0      NaN  204.890455  20791.318981     7.300212  368.516441    564.308654
1      3.716080  129.422921  18630.057858     6.635246        NaN    592.885359
2      8.099124  224.236259  19909.541732     9.275884        NaN    418.606213
3      8.316766  214.373394  22018.417441     8.059332  356.886136    363.266516 4  9.092223
       181.101509  17978.986339     6.546600  310.135738    398.410813

Organic_carbon  Trihalomethanes  Turbidity  Potability
0      10.379783       86.990970   2.963135         0
1      15.180013       56.329076   4.500656         0
2      16.868637       66.420093   3.055934         0
3      18.436524      100.341674   4.628771         0
4      11.558279       31.997993   4.075075         0
```

Checking the missing values**:**

In this step, the missing values or null values, if it present in the data are separated and number of nu ll values are shown through this code.

Code:
print("Missing values:\n", df.isnull().sum()) Output:

```
Missing values:
ph              491 Hardness
0
Solids            0
Chloramines         0
Sulfate           781
```

Conductivity 0
Organic_carbon 0
Trihalomethanes 162
Turbidity 0 Potability
0
dtype:int64

Check datatype:
In this step, the data type of the columns are discussed Code:
print("Data Types:\n", df.dtypes)

Output:
Data Types:
ph float64 Hardness
float64
Solids float64
Chloramines float64
Sulfate float64
Conductivity float64
Organic_carbon float64 Trihalomethanes
float64
Turbidity float64 Potability
int64
dtype: object

Check basic statistics:
the statistics of the columns such as count, mean, std, min, max, 25%, 50%, 75% are shown through the d escribe() function command.

Code:
print("Summary Statistics:\n", df.describe())

Output:
Summary Statistics:
ph Hardness Solids Chloramines Sulfate \
count 2785.000000 3276.000000 3276.000000 3276.000000
2495.000000 mean 7.080795 196.369496 22014.092526 7.122277
333.775777 std 1.594320 32.879761 8768.570828 1.583085

41.416840 min 0.000000 47.432000 320.942611 0.352000
129.000000 25% 6.093092 176.850538 15666.690297 6.127421
307.699498
50% 7.036752 196.967627 20927.833607 7.130299 333.073546
75% 8.062066 216.667456 27332.762127 8.114887 359.950170
max 14.000000 323.124000 61227.196008 13.127000
481.030642

Conductivity Organic_carbon Trihalomethanes Turbidity Potability count
3276.000000 3276.000000 3114.000000 3276.000000 3276.000000
mean 426.205111 14.284970 66.396293 3.966786 0.390110
std 80.824064 3.308162 16.175008 0.780382 0.487849 min
181.483754 2.200000 0.738000 1.450000 0.000000 25%
365.734414 12.065801 55.844536 3.439711 0.000000
50% 421.884968 14.218338 66.622485 3.955028 0.000000
75% 481.792304 16.557652 77.337473 4.500320 1.000000
max 753.342620 28.300000 124.000000 6.739000
1.000000

PREDICTIVE MODELING FOR POTABILITY:

Predictive modeling for water potability involves using historical water quality data to predict whether a given water sample is safe for human consumption. To create a predictive model for water potability, the process typically involves the following steps:

Steps for Building a Predictive Model for Water Potability:

1. Data Collection and Preprocessing:

Gather a dataset containing water quality parameters, such as pH, hardness, turbidity, solids, chlorides, sulfates, and other relevant factors. Ensure the data is labeled to denote potable and non-potable samples.

2. Data Cleaning and Preparation:

Handle missing values, outliers, and inconsistencies in the dataset. Normalize or scale the features for uniformity.

3. Feature Selection and Engineering:

Identify significant features that influence water potability. This could involve correlation analysis, domain knowledge, or feature importance techniques.

4. Data Splitting:

Divide the dataset into training and testing subsets to train and evaluate the model's performance.

5. Model Selection:

Choose an appropriate machine learning or statistical model. Common choices for binary classification tasks, such as predicting potable or non-potable water, include logistic regression, decision trees, random forests, support vector machines, or neural networks.

6. Model Training:

Train the selected model on the training dataset. The model learns the patterns and relationships between the water quality parameters and potability labels.

7. Model Evaluation:

Assess the model's performance using the testing dataset. Metrics like accuracy, precision, recall, F1-score, or area under the ROC curve (AUC-ROC) are typically used to evaluate classification models.

8. Hyperparameter Tuning and Cross-Validation:

Fine-tune the model's hyperparameters to improve its performance. Perform cross-validation to ensure the model's generalizability.

9. Model Validation and Interpretation:

Validate the model's predictions against real-world potability observations. Interpret the model to understand the significance of various features in determining water potability.

10. Deployment and Use:

Deploy the model in a real-time system or decision support tool. It can then be used to predict whether new water samples are potable based on their characteristics.

INSIGHTS FROM ANALYSIS:

Insights obtained from water quality analysis play a vital role in evaluating water quality and determining its potability. These insights help in understanding the factors influencing water quality, identifying potential risks, and ensuring water safety. Here's how these insights aid in assessing water quality and determining potability:

Identification of Contaminants and Parameters:

Contaminant Identification: Insights from the analysis reveal the presence of various contaminants such as heavy metals, microbial agents, chemicals, or organic compounds, helping to identify potential health hazards.

Parameter Assessment: Analysis provides an understanding of key parameters like pH, turbidity, dissolved solids, chlorine levels, microbial counts, and other chemical constituents crucial in assessing water quality.

Compliance with Standards and Guidelines:

Regulatory Compliance: Insights help in comparing water quality parameters against established regulatory standards or guidelines set by health or environmental agencies. Determining compliance ensures that the water meets safety requirements.

Health Risk Evaluation: Analysis insights allow for an assessment of potential health risks associated with certain contaminants or deviations from the acceptable range of parameters.

Source Identification and Risk Mitigation:

Source Identification: Understanding the sources of contamination or pollution helps in implementing measures to mitigate risks. Insights aid in identifying if the contamination is from natural sources, industrial discharges, agricultural runoff, or other human activities.

Risk Mitigation Strategies: Insights guide the formulation and implementation of appropriate interventions or remediation strategies to address the identified contamination sources and mitigate potential risks.

Real-time Monitoring and Early Warning:

Continuous Monitoring: Insights drive the establishment of monitoring programs and early warning systems. This facilitates ongoing surveillance of water quality, ensuring any deviations from safe levels are quickly detected and addressed.

Alert Systems: These insights can contribute to the development of alert systems that notify stakeholders, authorities, or the public about potential water quality issues or impending risks.

Decision Support for Water Treatment and Management:

Water Treatment Guidance: Insights help in determining the most suitable treatment methods and technologies to improve water quality based on the specific contaminants or deviations observed.

Policy Formulation: The data-driven insights influence policy-making by providing evidence for water quality management and governance. Authorities can make informed decisions based on analysis results to ensure safe and potable water for the public.

By leveraging insights derived from thorough water quality analysis, stakeholders, water management authorities, and policymakers can take informed actions, implement appropriate measures, and ensure safe, potable water for consumption and other purposes. This ultimately safeguards public health and environmental well-being.

CONCLUSION:

In conclusion, outline of the project's objective, design thinking process, data preprocessing, exploratory data analysis, data visualization and predictive modeling for potability, By this u can ensure your data is in suitable format and quality for further insights and determine potability