www.ijarp.com ISSN 2456-9992 Page: 01-14

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 09

WATER QUALITY ANALYSIS USING PYTHON AND POWER BI

***Prof. R. O. Tagade, Tanmay Balu Dongare**

Department of Computer Engineering Adsul's Technical Campus, Savitribai Phule Pune University (SPPU) Ahilyanagar, Maharashtra 414005, India.

Article Received: 09 August 2026, Article Revised: 29 August 2026, Published on: 19 September 2026

***Corresponding Author: Prof. R. O. Tagade**

Department of Computer Engineering Adsul's Technical Campus, Savitribai Phule Pune University (SPPU) Ahilyanagar, Maharashtra 414005, India.

Doi: https://doi-doi.org/101555/ijarp.4605

ABSTRACT

Water quality assessment requires the examination of several physical and chemical parameters collected from different sources and geographical locations. When these observations are stored across multiple related datasets, manual analysis becomes difficult and it can be challenging to identify relationships, trends, and variations in quality. This research paper presents an interactive water quality analysis framework using Python and Microsoft Power BI. The work follows a structured analytical workflow consisting of dataset preparation, data inspection, cleaning, exploratory analysis, visualization, relational data modeling, dashboard development, and testing.

The implementation uses five related datasets: Water_Sample, Location_Master, Water_Source_Master, Physical_Parameters, and Chemical_Parameters. Python, Pandas, and Matplotlib are used for data loading, inspection, cleaning, descriptive analysis, and graphical representation. Microsoft Power BI is then used to establish relationships between tables and present the results through KPI cards, a donut chart, a water-source column chart, a WQI trend line chart, geographical analysis, a decomposition tree, slicers, and conditional formatting.

The supplied dashboard contains 500 water samples, an average Water Quality Index (WQI) of 93.63, an average pH of 7.29, and 398 samples categorized as Excellent. The displayed quality-status distribution is 79.6% Excellent, 20.2% Good, and 0.2% Moderate. The dashboard also provides state-wise, city-wise, source-wise, and month-wise analysis. The work demonstrates how combining programmatic data processing with interactive business-intelligence visualization can transform structured environmental records into an accessible analytical system.

KEYWORDS: Water Quality, Water Quality Index, WQI, Python, Pandas, Matplotlib, Power BI, Data Visualization, Environmental Data, Dashboard, Data Analytics.

I. INTRODUCTION

Water is a fundamental natural resource required for domestic use, agriculture, industry, and ecosystem sustainability. The quality of available water can vary according to source, location, environmental conditions, and chemical and physical characteristics. Consequently, water-quality assessment requires systematic collection and analysis of multiple parameters rather than relying on a single measurement.

The project on which this paper is based is titled "Water Quality Analysis Using Power BI and Python." The supplied project report describes a workflow that begins with structured datasets and proceeds through inspection, cleaning, analysis, visualization, Power BI modeling, dashboard development, testing, and documentation. The report identifies Python as the main environment for data analysis and Power BI as the interactive visualization platform.

A major challenge in environmental data analysis is the volume and variety of observations. A single water sample may have an identification code, collection date, location, water source, physical measurements, chemical measurements, and a quality index. When these attributes are stored in separate but related tables, an analytical system must preserve the relationships while also making the information understandable to users.

The proposed framework addresses this challenge by separating data-processing activities from visualization activities. Python is used to inspect and prepare the datasets, while Power BI provides the interactive presentation layer. This division makes the workflow systematic and allows the dashboard to represent several analytical dimensions at the same time.

A. Problem Statement

Water-quality records distributed across multiple tables can be difficult to analyze manually. Users may need to compare locations, sources, dates, WQI values, pH values, and other parameters simultaneously. A suitable system is therefore required to integrate the datasets, perform basic data validation and analysis, and provide an interactive dashboard through which users can explore the available information.

B. Objectives

- Load and inspect structured water-quality datasets using Python.
- Identify fields, data types, missing values, and relationships among datasets.
- Prepare data for analysis and visualization.
- Analyze physical parameters such as temperature, turbidity, TDS, and conductivity.
- Analyze chemical parameters including pH, hardness, alkalinity, chloride, nitrate, phosphate, sulfate, fluoride, iron, lead, and arsenic.
- Visualize water-quality distributions using Python.
- Create a relational data model in Power BI.
- Develop an interactive dashboard using KPI cards, charts, map visualization, slicers, and a decomposition tree.
- Study WQI variation by state, city, water source, and collection period.
- Test the dashboard for filtering, relationships, field selection, readability, and output correctness.

II. BACKGROUND AND CONCEPTUAL BASIS

Water-quality analysis combines measurements from several categories. Physical parameters describe observable or measurable characteristics such as temperature, turbidity, dissolved solids, and conductivity. Chemical parameters provide information about acidity or alkalinity and dissolved chemical constituents. A WQI provides a compact way to represent water-quality information when the underlying dataset already contains an index value.

The present project does not describe the mathematical derivation used to calculate every WQI value. Instead, the supplied Water_Sample dataset contains a Quality_Index (WQI) field and a Water_Quality_Status field. Therefore, this paper treats the WQI values as dataset-provided analytical values and focuses on their exploration and visualization.

A. Role of Python

Python provides a programmable environment for data inspection and preparation. Pandas supports DataFrame-based manipulation, while Matplotlib supports graphical representation. In this project, Python is used to load CSV files, inspect records, check missing values, generate descriptive information, and create preliminary graphs before the data are modeled in Power BI.

B. Role of Power BI

Power BI provides an interactive environment for combining related datasets and presenting analytical results. The dashboard developed for this project uses KPI cards for summary values, a donut chart for quality-status distribution, a column chart for water sources, a line chart for monthly WQI, a map and location analysis, a decomposition tree, and slicers for filtering.

III. RELATED WORK AND LITERATURE REVIEW

Water Quality Index approaches have been widely discussed as methods for converting several water-quality measurements into a more interpretable index. Published reviews emphasize that WQI models can differ in the parameters selected, weighting methods, sub-index calculations, aggregation procedures, and treatment of uncertainty. Consequently, a WQI value should always be interpreted together with information about the methodology and standards used.

International drinking-water guidance provides a broader context for water-quality monitoring and emphasizes the importance of controlling chemical and microbiological hazards. National standards similarly provide specifications for drinking water. In an analytical dashboard, these standards can potentially be used as reference thresholds; however, the current project primarily visualizes the supplied dataset rather than implementing a complete standards-compliance engine.

Data analytics and business-intelligence tools have also become important for environmental datasets. Programmatic tools such as Python and Pandas provide reproducible data preparation, while visualization platforms provide interactive exploration. The combination used in this project is therefore appropriate for demonstrating a practical workflow from raw tabular data to an interactive dashboard.

Research Gap Addressed

The project focuses on the practical integration of five related water-quality datasets into a single analytical workflow. Instead of presenting only isolated graphs, it connects sample, location, source, physical, and chemical information and makes the results explorable through Power BI filters and visual components. This integration is the central contribution of the implemented project.

IV. DATASET DESCRIPTION

The project report identifies five datasets, each containing 500 records. Their structures are summarized below.

Dataset	Records	Purpose
Water_Sample	500	Sample identifiers, location/source links, collection date, WQI and quality status
Location_Master	500	Location, city, state, latitude and longitude information
Water_Source_Master	500	Water-source name and source type
Physical_Parameters	500	Temperature, turbidity, TDS and conductivity
Chemical_Parameters	500	pH and selected chemical measurements

A. Water_Sample Dataset

The Water_Sample table contains Sample_ID as the primary identifier and links to station, location, source, and equipment information. It also includes Collection_Date, Quality_Index (WQI), and Water_Quality_Status. Example records in the project report include WQI values of 97.38, 81.10, 88.86, 96.77, and 93.02 for the first five samples.

B. Location_Master Dataset

The Location_Master table contains Location_ID, City, State_ID, State, Latitude, and Longitude. These fields support geographical analysis and allow the dashboard to connect water samples with their corresponding locations.

C. Water_Source_Master Dataset

The Water_Source_Master table contains Source_ID, Water_Source, and Source_Type. Examples in the supplied dataset include Godavari River, Hussain Sagar Lake, Groundwater Well, Public Water Supply, and Community Borewell.

D. Physical_Parameters Dataset

Physical measurements include Temperature in degrees Celsius, Turbidity in NTU, TDS in mg/L, and Conductivity in µS/cm. These variables can be used to examine the physical characteristics of sampled water.

E. Chemical_Parameters Dataset

Chemical measurements include pH, hardness, alkalinity, chloride, nitrate, phosphate, sulfate, fluoride, iron, lead, and arsenic. The supplied report provides sample records for each of these fields.

V. SYSTEM ARCHITECTURE AND WORKFLOW

The implemented architecture can be represented as a layered analytical workflow:

DATASETS → PYTHON INGESTION → DATA INSPECTION → CLEANING → EXPLORATORY ANALYSIS → POWER BI IMPORT → RELATIONSHIPS → VISUALIZATION → INTERACTIVE FILTERING → ANALYTICAL OUTPUT

A. Data Ingestion Layer

CSV files are loaded into Python DataFrames. Each dataset is inspected for its number of records, fields, data types, and missing values. The purpose of this stage is to identify structural problems before visualization.

B. Data Preparation Layer

The preparation stage checks column names, missing cells, data types, and consistency. The documented project dataset summary reports zero missing cells for the five listed datasets.

C. Analytical Layer

Descriptive statistics and frequency-based analysis are used to understand the distribution of values. Graphs are generated for water-quality status, state distribution, water sources, turbidity, and pH.

D. Visualization Layer

The prepared data are imported into Power BI and related using appropriate identifiers. Dashboard visuals then provide summary indicators and detailed analytical views.

VI. IMPLEMENTATION METHODOLOGY

A. Dataset Loading

The first implementation task is to load the CSV files using Pandas. The analyst verifies that the files can be opened and that expected columns are available. This prevents later visualization errors caused by incorrect field selection.

B. Data Inspection

The inspection process includes displaying initial records, reviewing column information, calculating descriptive statistics, and checking missing values. The project report records the use of head, column information, descriptive statistics, and missing-value checks.

C. Data Cleaning

Data cleaning focuses on missing values, inconsistent records, data types, and appropriate column selection. Cleaning is important because a visualization can appear correct while still being based on an incorrectly typed or incomplete field.

D. Exploratory Visualization

Python visualizations provide an initial analytical view before Power BI dashboard construction. Graphs can be used to examine distributions and identify unusual patterns that may require further inspection.

E. Relational Modeling

The Power BI model connects the five tables through common identifiers. Sample-level records are associated with location, source, physical parameters, and chemical parameters. This structure allows the dashboard to display information from multiple tables in a coordinated manner.

F. Dashboard Construction

The dashboard is organized around summary KPIs and supporting analytical visuals. Users can move from overall sample counts and average values to state, source, city, and temporal analysis.

VII. DASHBOARD DESIGN AND VISUAL OUTPUTS

The supplied dashboard screenshots show the final Power BI interface. The first dashboard view contains a title area, water-source filters, four KPI cards, a state-wise WQI bar chart, and an overall water-quality-status donut chart.

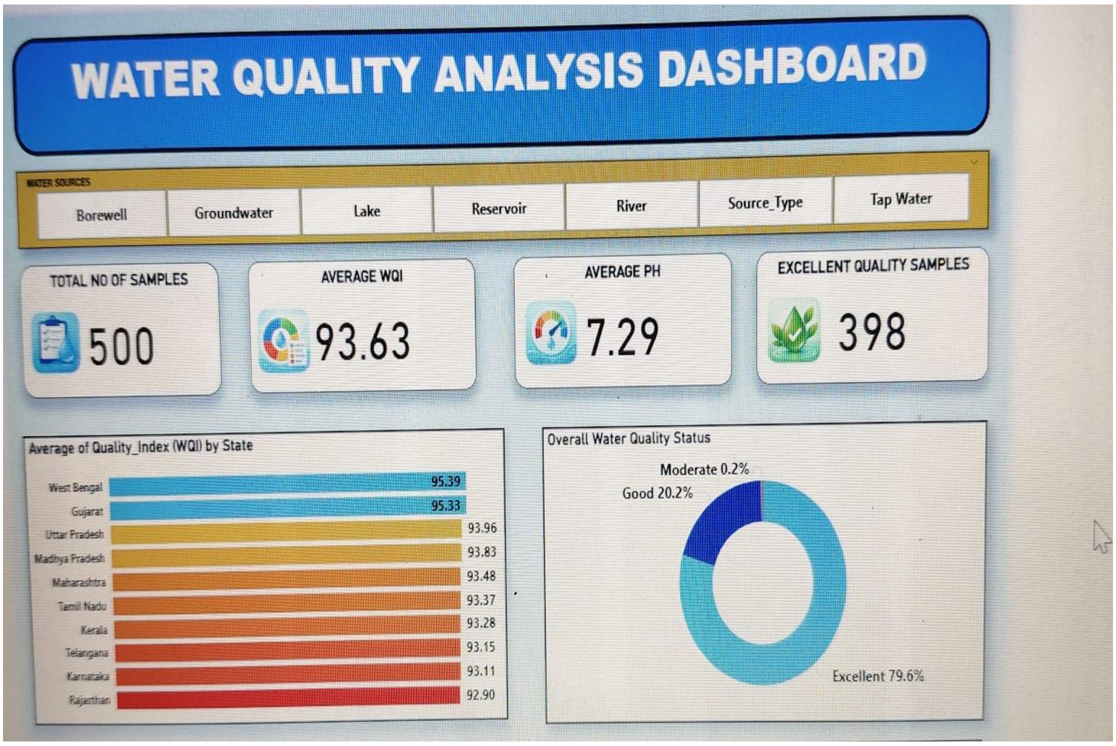

Fig. 1. Final Power BI dashboard overview showing KPI cards, source filters, state-wise WQI and overall quality status.

The second supplied screenshot shows water samples by water resource, a state/city analysis section, and a monthly WQI line chart. These views provide additional analytical detail beyond the headline KPIs.

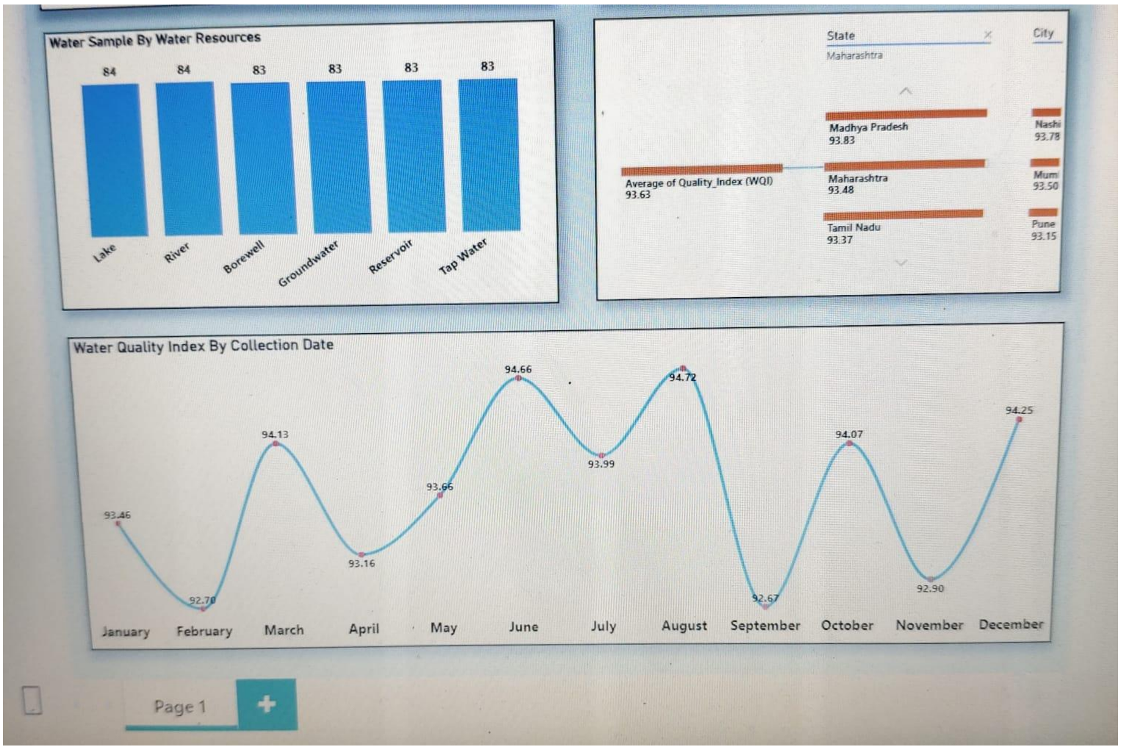

Fig. 2. Water-source distribution, state/city analysis, and monthly WQI trend.

A. KPI Cards

The dashboard presents four major indicators: total number of samples, average WQI, average pH, and excellent-quality sample count. These indicators provide an immediate summary of the selected dataset.

B. Water Quality Status

The donut chart separates the samples into Excellent, Good, and Moderate categories. In the supplied dashboard, the displayed distribution is 79.6% Excellent, 20.2% Good, and 0.2% Moderate.

C. State-wise WQI

The state-wise visualization allows the user to compare average WQI values across the displayed states. The dashboard shows values ranging from 92.90 for Rajasthan to 95.39 for West Bengal.

D. Water Source Analysis

The water-source chart contains Lake, River, Borewell, Groundwater, Reservoir, and Tap Water. The displayed counts are 84, 84, 83, 83, 83, and 83 respectively.

E. Monthly WQI Trend

The line chart presents WQI values for January through December. The displayed values show variation over the collection period, with the highest displayed value in August at 94.72 and the lowest displayed value in September at 92.67.

F. State and City Analysis

The second dashboard page provides a state and city analytical view. The supplied screenshot demonstrates filtering for Maharashtra and presents related WQI values for selected states and cities.

VIII. RESULTS

A. Overall Results

Indicator	Value
Total samples	500
Average WQI	93.63
Average pH	7.29
Excellent samples	398
Excellent share	79.6%
Good share	20.2%
Moderate share	0.2%

The results indicate that the dashboard can summarize the complete dataset through a small set of high-level indicators while retaining access to detailed dimensions.

B. State-wise Results

State	Average WQI
West Bengal	95.39
Gujarat	95.33
Uttar Pradesh	93.96
Madhya Pradesh	93.83
Maharashtra	93.48
Tamil Nadu	93.37
Kerala	93.28
Telangana	93.15
Karnataka	93.11
Rajasthan	92.90

C. Water Source Results

Source	Displayed Samples
Lake	84
River	84
Borewell	83
Groundwater	83
Reservoir	83
Tap Water	83

D. Monthly WQI Results

Month	WQI
January	93.46
February	92.70
March	94.13
April	93.16
May	93.66
June	94.66
July	93.99
August	94.72
September	92.67
October	94.07
November	92.90
December	94.25

IX. TESTING AND VALIDATION

Testing was performed at both hardware and software levels. Hardware testing focused on the computer system, keyboard, mouse, display, storage, and internet operation. Software testing focused on CSV loading, field names, data types, missing values, Python graphs, Power BI relationships, filters, KPI cards, charts, map visualization, decomposition tree, and conditional formatting.

Test Area	Test Activity	Expected Result
Dataset loading	Load all five CSV files	All datasets load successfully
Data structure	Check columns and types	Expected fields are available
Missing values	Check dataset completeness	Missing cells are identified
Python visualization	Verify titles, labels and values	Graphs display correctly
Power BI relationships	Check table connections	Relationships operate correctly
Slicers	Apply state/source filters	Visuals update consistently
KPI cards	Check summary calculations	Displayed indicators are correct
Charts	Check labels and values	Visuals correspond to selected data
Dashboard readability	Inspect layout and labels	Information remains understandable
Output documentation	Record final results	Outputs are reproducible and documented

A. Testing Challenges

The project report notes that some visualizations initially produced errors when a field not present in the selected dataset was used. Checking actual DataFrame columns resolved such issues. Establishing relationships between Power BI tables also required identification of suitable key fields. These observations emphasize the importance of validation before dashboard publishing.

X. DISCUSSION

The implemented system demonstrates a practical approach to environmental data analysis. The five-dataset structure separates concerns: location information is maintained independently from sample information, water-source information is maintained independently from physical and chemical measurements, and the Water_Sample table acts as a central analytical table containing the WQI and quality-status fields.

The dashboard provides both summary and detailed analysis. KPI cards answer high-level questions such as how many samples exist and what the average WQI is. Charts then provide additional context by showing water-source distribution, state-wise WQI, monthly variation, and quality-status proportions.

An important aspect of the system is that visualization does not replace the underlying data-validation process. If a field is incorrectly selected or relationships are incorrectly configured, a visually attractive dashboard may still produce misleading results. Therefore, the Python inspection and Power BI relationship-testing stages are important components of the overall workflow.

The displayed results should be interpreted as analysis of the supplied project dataset. The dashboard indicates the distribution of quality categories and WQI values, but the project documentation does not establish that these values represent a real-time national water-quality monitoring network. Similarly, the paper does not claim that the displayed WQI alone determines regulatory drinking-water compliance.

XI. ADVANTAGES

- Provides a centralized analytical dashboard.
- Integrates five related water-quality datasets.
- Supports interactive filtering and exploration.
- Combines programmatic Python analysis with Power BI visualization.
- Provides state, city, source, and temporal views.

- Reduces the effort required to inspect multiple CSV files separately.
- Supports repeatable data-processing and visualization workflows.
- Improves presentation of analytical results for project demonstrations and decision-support contexts.

XII. LIMITATIONS

- The supplied project is based on structured datasets and does not describe continuous real-time sensor acquisition.
- The WQI calculation methodology is not fully specified in the supplied report; the paper therefore uses the WQI values already present in the dataset.
- The dashboard is an analytical visualization system and is not itself a laboratory testing system.
- Interpretation of water-quality categories depends on the classification represented in the supplied dataset.
- More extensive external validation would be required before using the system for regulatory or public-health decisions.

XIII. FUTURE SCOPE

The framework can be extended into a larger environmental monitoring platform. IoT sensors could provide real-time measurements for pH, turbidity, temperature, TDS, conductivity, and other parameters. A backend database could store incoming observations and automatically refresh the dashboard. Machine-learning methods could also be investigated for anomaly detection, quality-status prediction, or forecasting. GIS capabilities could be expanded to provide spatial interpolation and hotspot analysis.

A future implementation could additionally incorporate threshold-based alerts, user authentication, mobile access, historical comparison, automated reports, and integration with laboratory information. If drinking-water compliance is required, the dashboard could be extended with explicit standard limits and parameter-level compliance indicators, provided the applicable standard and measurement methodology are clearly defined.

XIV. CONCLUSION

This research presented an integrated framework for water-quality data analysis using Python and Microsoft Power BI. The system organizes five related datasets and provides a structured workflow for data loading, inspection, cleaning, exploratory analysis, relational modeling, visualization, testing, and documentation.

The final dashboard contains 500 water samples, an average WQI of 93.63, an average pH of 7.29, and 398 samples displayed as Excellent. It provides multiple analytical views including quality-status distribution, state-wise WQI, water-source distribution, state/city analysis, and monthly WQI trends. The combination of Python and Power BI demonstrates how environmental data can be transformed into an interactive and understandable analytical interface.

The project also demonstrates that successful dashboard development depends not only on visual design but also on dataset structure, correct relationships, field validation, filtering behavior, and systematic testing. Future work can extend the framework toward real-time IoT integration, machine learning, GIS analysis, automated alerts, and standards-based compliance monitoring.

XV. REFERENCES

1. M. G. Uddin, S. Nash, and A. I. Olbert, “A review of water quality index models and their use for assessing surface water quality,” Ecological Indicators, vol. 122, p. 107218, 2021, doi: 10.1016/j.ecolind.2020.107218.
2. T. Abbasi and S. A. Abbasi, Water Quality Indices. Elsevier, 2012.
3. World Health Organization, Guidelines for Drinking-water Quality: Fourth Edition Incorporating the First and Second Addenda. Geneva: WHO, 2022.
4. Bureau of Indian Standards, IS 10500:2012, Drinking Water — Specification, Second Revision, New Delhi: BIS.
5. J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.
6. The pandas development team, pandas Documentation, Python Data Analysis Library.
7. Python Software Foundation, Python Documentation.
8. Microsoft, Power BI Documentation, Microsoft Learn.
9. Microsoft, Power BI Desktop Documentation, Microsoft Learn.
10. R. Almeida et al., “A recreational water quality index using chemical, physical and microbiological parameters,” Environmental Science and Pollution Research, 2012.
11. R. Abrahão et al., “Use of index analysis to evaluate the water quality of a stream receiving industrial effluents,” Water SA, 2007.

12. A. H. M. J. Alobaidy, H. S. Abid, and B. K. Maulood, "Application of Water Quality Index for Assessment of Dokan Lake Ecosystem, Kurdistan Region, Iraq," Journal of Water Resource and Protection, 2010.
13. S. Tyagi, B. Sharma, P. Singh, and R. Dobhal, "Water Quality Assessment in Terms of Water Quality Index," American Journal of Water Resources, vol. 1, no. 3, pp. 34–38, 2013.
14. P. Sutadian, M. Muttil, A. Yilmaz, and B. Perera, "Development of river water quality indices—a review," Environmental Monitoring and Assessment, 2016.
15. Project Report, Water Quality Analysis Using Power BI and Python, Department of Computer Engineering, Adsul's Technical Campus, SPPU, Academic Project Report, 2025–2026.