PROJECT REPORT

On

WATER QUALITY ANALYSIS USING PYTHON

By

Kavya sri Gullipalli - AP21110010060

Under the guidance of

K. Meenakshi, APSSDC

K. Narmada, APSSDC

Prepared in the partial fulfillment of the

Summer Internship Course

At

Andhra Pradesh State Skill Development Corporation

(APSSDC)

3rd floor,Infosight Building, survey No. 78/2,Near Pathuru Road

Junction,NH-16 Service Road, Tadepalli - 522501,Guntur District,Andhra

Pradesh

ACKNOWLEDGEMENTS

We would like to express my special thanks to our trainers Meenakshi and Narmada Mam for their time and efforts that they provided throughout the internship. Your useful advice and suggestions were really helpful to us during the project's completion. In this aspect, we are eternally grateful to you.

We are highly indebted to APSSDC management for their guidance and constant supervision as well as for providing necessary information regarding the project and also for their support in completing the project.

Tables of Contents

Abstract

The project aims to leverage the capabilities of python programming for comprehensive analysis of water potability , with focus on extracting insights into chemical forms present in different water samples. The dataset includes ph, hardness, solids, chloramines, sulphate, Conductivity, Organic carbon, Trihalomethanes, turbidity, and Potability . The analysis provides the information of the water samples which are suitable for drinking and unsafe for drinking.

The dataset is from kaggle.com to analyze the water samples contents. This enables a detailed analysis in the investigation of water,of different chemical forms. This project involves rigorous data processing, exploratory data analysis, and statistical techniques to extract meaningful insights, patterns and trends from the dataset, and to ensure accuracy data cleaning is performed. The provided result will help the user to improve their drinking habits.

Introduction

Water is one of the most fundamental and essential resources for all forms of life. Ensuring the safety and quality of drinking water is of paramount importance for public health and well-being. In this analysis, we delve into a comprehensive examination of water potability using a dataset sourced from Kaggle. This dataset contains various parameters that play a role in determining the potability of water, allowing us to explore the factors that contribute to safe and suitable drinking water.

The dataset comprises a collection of samples, each characterized by a set of features that include both chemical and physical attributes. These features encompass a wide range of measurements, such as pH levels, concentrations of dissolved solids, turbidity, and various ions. The primary objective of this analysis is to develop a deep understanding of the relationships between these features and the potability of water, with the ultimate goal of building a predictive model that can assess the potability of water samples based on their properties

System Requirements

- PC (Minimum 4GB Ram and 256 SSD)
- OS (Windows, MAC)
- Python (Version 3.10)
- Jupyter Notebook (IDE)

Uses of Data Analysis Library

Data Analysis libraries are essential tools for working with and manipulating data in various formats. These libraries provide a wide range of functions and methods that enable you to perform tasks such as data cleaning, transformation, exploration, visualization, and more.

Pandas: Libraries like pandas allow you to read data from various file formats such as CSV, Excel, and more. This is the first step in the data analysis.

Numpy: It is a versatile tool for various data manipulation and analysis tasks.

Matplotlib: A popular data visualization library that allows you to create a wide range of interactive animated visualizations for conveying insights and information from your data.

Plotly: It is an interactive data visualization library for python that allows the creation of custom, interactive visualizations. Plotly visualizations can be explored, zoomed and interacted directly.

Data Cleaning and Preprocessing: Data is rarely clean and ready for analysis right away. So, data analysis libraries provide tools to handle missing values, duplicate records, incorrect data types and other data quality issues.
Data Transformation: You can reshape data using Pandas. This is crucial for preparing data in the format needed for analysis.

Exploratory data analysis (EDA): Libraries like Pandas, NumPy enable you to create summary statistics, scatter plots, and other visualizations that help you understand the data's characteristics and relationships.

Data Visualization: Libraries like Matplotlib, Plotly allow you to create a wide range of charts, graphs, plots and visualizations to effectively communicate insights from your data.

Purpose of the Project

The primary purpose of the water potability analysis project is to comprehensively investigate the factors that contribute to the potability of water and to develop a predictive model that can effectively assess whether a given water sample is suitable for consumption.
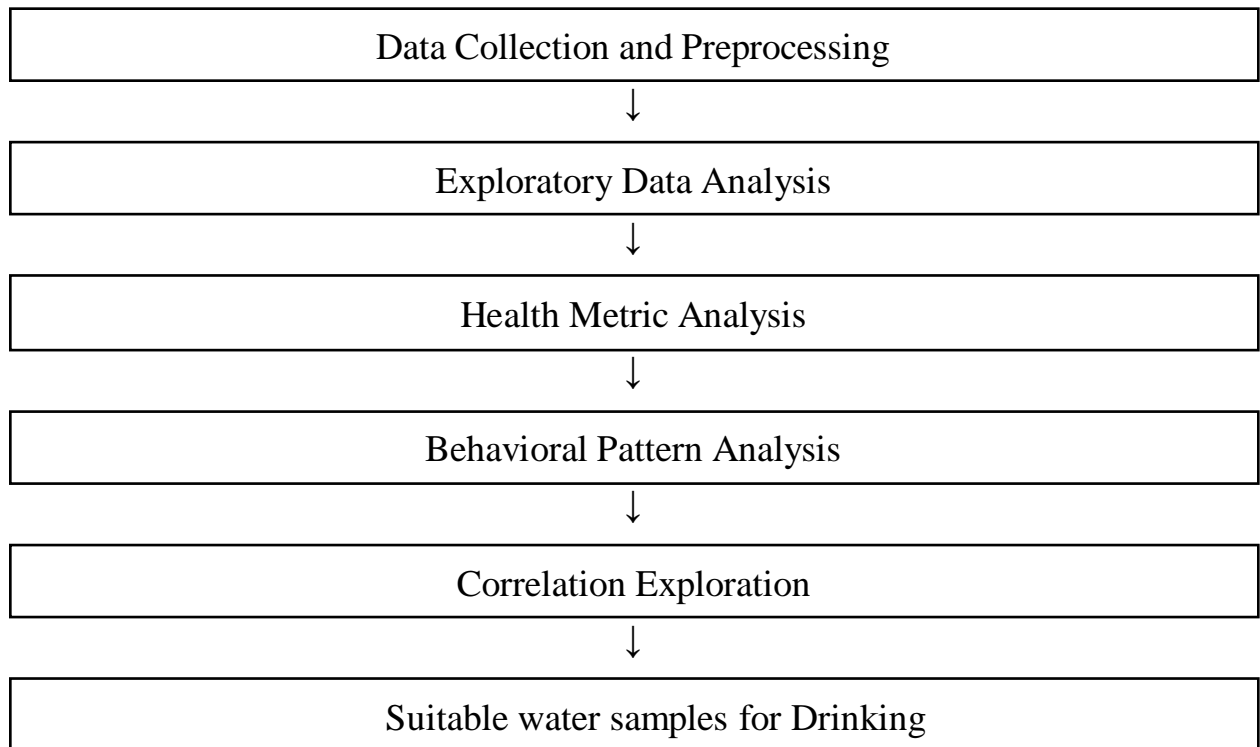
Within this project, every attribute and parameter takes on new significance as a potential contributor to water potability. The chemical composition, mineral content, organic matter levels, and even the microbiological profile of water samples all play a role in shaping their suitability for consumption. By dissecting and understanding these elements, and water quality outcomes.

Objective and Scope of the Project

The primary objective of this project is to conduct a detailed analysis of water potability data. The specific objectives and scope includes:

- Data Collection: Utilizing the data consisting of timestamped records of different water samples.
- Data Preprocessing: Cleaning and preprocessing the data to handle missing values, and ensure data consistency.
- Exploratory Data Analysis: Conducting comprehensive EDA to gain insights into the distribution of chemical forms and assess data quality.
- Correlation Analysis: Perform statistical analysis to uncover correlations between variables.
- Visualization and Reporting: Creating interactive visualizations and reports to communicate insights to users.
- Health Recommendations: Regular water testing, Ph monitoring, and
  - Boiling andtesting improves the water quality

Architecture of project

Data Collection and Preprocessing

↓

Exploratory Data Analysis

↓

Health Metric Analysis

↓

Behavioral Pattern Analysis

↓

Correlation Exploration

↓

Suitable water samples for Drinking

Water Potability Data collection

Collecting data for water potability analysis involves gathering informationabout various parameters that can affect the suitability of water for consumption.

1. Identify Relevant Parameters: Determine the key chemical, physical, and microbial parameters that are indicative of water potability. These may include pH, turbidity, conductivity, dissolved oxygen, total dissolved solids, heavy metal concentrations, microbial counts, and more.

2. Sampling: Collect water samples from various sources such as rivers, lakes, groundwater wells, and municipal water supplies. Ensure that the sampling process is representative of the area and conditions you intend to study.

3. Sample Size and Frequency: Decide on the number of samples to collect and the frequency of sampling. More samples and frequent measurements provide a more accurate representation of water quality fluctuations.

4. Laboratory Analysis: Perform tests to analyze the collected water samples. Use appropriate testing methods to measure parameters like pH, alkalinity, hardness, microbial content, and chemical constituents.

5. Quality Assurance: Maintain quality control during the data collection process. Calibrate instruments, follow standard operating procedures, and perform replicate tests to ensure accuracy and reliability.

6. Metadata Collection: Record relevant metadata for each sample, such as location (GPS coordinates), date and time of collection, weather conditions, and any known sources of contamination.

7. Data Management: Organize the collected data in a structured format. Use spreadsheets or databases to store information about each sample and its corresponding parameter values.

8. Data Validation: Clean the collected data by identifying and addressing inconsistencies, outliers, or errors. Data validation helps ensure the accuracy and integrity of your dataset.

9. Data Integration: If applicable, integrate data from different sources and time periods to create a comprehensive dataset that reflects changes in water quality over time.

10. Publicly Available Datasets: You can also explore publicly available water quality datasets provided by governmental agencies, research organizations, or environmental databases. These datasets may include historical water quality measurements from various locations.

Health impacts of Water Quality Analysis

The analysis of water potability data can have various health related impacts and benefits for individuals. By leveraging the insights gained from water samples data analysis, users can make informed decisions to improve overall well-being. Here are some potential health impacts of water quality analysis:

- Protection of Public Health: Accurate water potability analysis ensures that drinking water meets regulatory standards and guidelines established by health authorities. Compliance with these standards safeguards public health by minimizing the risk of water-related illnesses and ensuring the overall well-being of communities.

- Prevention of Waterborne Diseases: Water potability analysis helps identify and quantify potential contaminants in water sources. By detecting harmful microorganisms (such as bacteria, viruses, and parasites) and chemical pollutants (like heavy metals, pesticides, and organic compounds), the analysis contributes to preventing waterborne diseases that can cause gastrointestinal, respiratory, and other health issues.

- Long-Term Health Impacts: Prolonged exposure to unsafe drinking water can lead to chronic health issues, including kidney problems, developmental disorders, neurological effects, and even cancer. Water potability analysis contributes to reducing these long-term health risks.

Advantages

The Water quality data analysis project offers several advantages that can positively influence the individual's healthcare.

Here are some key advantages:

- Health Protection: The primary advantage of water potability analysis is the protection of public health. By identifying contaminants and potential health hazards in drinking water, analysis ensures that water sources meet regulatory standards and are safe for consumption, and related health issues

- Public Awareness: Sharing the results of water potability analysis with the public raises awareness about water quality issues and encourages responsible water usage practices. Informed communities are more likely to actively participate in safeguarding water sources.

- Data-Driven Decision-Making: Water potability analysis provides data-driven insights into the quality of water sources. These insights empower regulatory agencies, policymakers, and communities to make informed decisions regarding water treatment, resource allocation, and public health interventions.

- Customized Treatment Strategies: Analysis of water quality parameters helps tailor water treatment strategies to specific contaminants present in each water source. This ensures efficient and effective treatment, optimizing the removal of pollutants and enhancing water safety.

- Global Health Impact: In regions with limited access to clean water, water potability analysis is crucial for international health initiatives. Accurate analysis supports efforts to provide safe drinking water to underserved populations and improve overall community health.

Conclusion

The water potability analysis stands as a monumental step towards ensuring the purity and safety of our most essential resource - water. Through meticulous examination of diverse parameters, this analysis sheds light on the complex factors that determine water's suitability for consumption. By unearthing the relationships between chemical composition, microbial content, and physical attributes, this undertaking safeguards public health and advances environmental stewardship.

With each pH level measured, every mineral concentration assessed, and all microbial counts analyzed, the analysis transcends data points, transforming into a shield against waterborne diseases and long-term health risks. This comprehensive exploration empowers regulatory bodies, policymakers, and communities to enact strategies that prioritize clean and safe drinking water, particularly for the most vulnerable among us.

Just as water flows through every aspect of our lives, the implications of this analysis ripple through society. It empowers individuals to make informed choices about their water sources, inspiring a collective commitment to responsible water usage and source protection. Moreover, the analysis resonates on a global scale, serving as a foundation for international efforts to provide safe water to underserved communities, thereby addressing a fundamental component of public health on a worldwide stage.

we can proactively mitigate health risks and nurture the vitality of both individuals and ecosystems. As the journey of analysis continues, a reminder of our shared commitment to a healthier, safer world.