Data Analysis on the availability of drinking water

Analyzing Water Quality Data to Address Urban Pollution
(SDG 6: Clean Water and Sanitation)

Concept of the Project
Urban pollution is a pervasive issue that affects millions of people worldwide, leading to serious health problems and environmental degradation. This project aims to analyze the quality of water currently. By leveraging data analysis tools and methodologies, the project seeks to propose actionable solutions that align with Sustainable Development Goal 6 (SDG 6): Clean Water and Sanitation. This SDG aims to make drinking water safe and sustainable.

Problem Statement
Access to safe and affordable drinking water is a fundamental human right and a critical component of public health. Despite significant progress, many communities, especially in developing countries, still face severe challenges in accessing clean drinking water. The lack of safe drinking water leads to numerous health problems, including waterborne diseases such as diarrhea, cholera, dysentery, typhoid, and hepatitis A. These diseases are particularly detrimental to vulnerable populations, including children, the elderly, and those with compromised immune systems. This project seeks to address this problem by leveraging available data and technology to develop a solution to monitor drinking water quality, provide timely alerts, and recommend actions to improve water safety, ultimately contributing to better health outcomes and sustainable development.

Objective of the Project
The primary objective of this project is to develop a comprehensive data-driven platform to ensure the safety and accessibility of drinking water in underserved regions.
 The specific objectives are:
- To collect and analyze drinking water quality data from reliable sources.
- To identify the primary sources of water pollution in urban areas.
- To develop predictive models for future pollution levels based on current data.
- To propose actionable solutions and policy recommendations to mitigate urban pollution.
- To assess the potential impact of these solutions on achieving SDG 6.

Data source used
The project will use a water quality data set taken from kaggle.

Features

The key features of the dataset will include:
- Pollutants: Levels of various pollutants
- Potability: If the water is drinkable or not

Tool for Analysis

The following tools and technologies will be used for data analysis:
1. Python: For data cleaning, analysis, and visualization, using libraries such as Pandas, NumPy, Matplotlib, and Seaborn.
2. Google Colabs: For developing predictive models and machine learning algorithms.

Hypothesis

The hypothesis of the project is providing data-driven insights and recommendations to policymakers and local authorities will support informed decision-making and lead to more effective interventions and investments in water infrastructure and public health measures. Additionally, improving the monitoring and management of drinking water quality will lead to a reduction in the incidence of waterborne diseases, thereby improving public health outcomes in the targeted regions.

Methodology

The project will be conducted in the following phases:

Data Collection:
- Gather water quality data from the aforementioned sources.

Data Cleaning and Preprocessing:
- Handle missing values, outliers, and inconsistencies in the data.
- Standardize data formats and integrate datasets from different sources.

Exploratory Data Analysis (EDA):
- Perform descriptive statistical analysis to understand the distribution and variability of clean drinking water.

Source Identification:
- Use correlation analysis and regression models to identify potential sources of pollution.
- Analyze the impact of different factors (e.g., traffic density, industrial activity) on pollution levels.

Predictive Modeling:
- Develop machine learning models to predict future pollution levels based on given data.
- Validate and test the models using appropriate metrics.

Solution Development:
- Based on the analysis, propose solutions such as stricter regulations.
- Assess the feasibility and potential impact of these solutions.

Reporting and Presentation:
- Compile the findings into a comprehensive report.
- Create visualizations and interactive dashboards to present the results.
- Develop policy briefs and recommendations for stakeholders.

Probable Outcome

The expected outcomes of the project are:

- Comprehensive Analysis: A detailed analysis of water quality data identifying key sources of pollution.
- Predictive Models: Reliable models for predicting future pollution levels and assessing the impact of potential interventions.
- Actionable Solutions: Data-driven solutions and policy recommendations to reduce urban pollution.
- Impact Assessment: Evaluation of the potential impact of proposed solutions on achieving SDG 6.
- Awareness and Engagement: Increased awareness among policymakers and the public about the sources and impacts of urban pollution, and the benefits of proposed interventions.

By addressing urban pollution through data analysis and evidence-based solutions, this project will contribute to creating sustainable and healthier urban environments, aligning with the objectives of SDG 6: Clean Water and Sanitation.