Maji Ndogo Water Infrastructure Project

Introduction

This project aimed to explore a database containing information about water sources, visits to these sources, water quality assessments, and pollution issues. The goal is to extract meaningful insights to address water quality and pollution challenges in a fictional town Maji Ndogo.

Database Overview

The database consists of the following tables:

- Data_dictionary: Contains information about the database schema or data dictionary.
- Employee: Contains information about employees.
- Global_water_access: Contains information about global water access.
- Location: Contains information about locations.
- Visits: Contains information about visits to water sources.
- Water_quality: Contains information about water quality assessments.
- Water_source: Contains information about water sources.
- Well_pollution: Contains information about well pollution.

Technologies Used

- SQL - MySQL Workbench

Tasks Overview (Project 1)

- Get to know our data: Load and examine the database to understand its structure and content.
- Dive into the water sources: Explore the 'water_sources' table to identify the unique types of water sources.
- Unpack the visits to water sources: Analyze the 'visits' table to understand the frequency and distribution of visits to different water sources.
- Assess the quality of water sources: Use the 'water_quality' table to identify frequently visited water sources with decent quality scores.
- Investigate pollution issues: Identify water sources with pollution issues based on 'pollution_tests' results.

Tasks Overview (Project 2)

- Cleaning Our Data: Updated employee data by assigning email addresses for employees based on the format first_name.last_name@ndogowater.gov.

- Honouring the Workers: Finding our best employees by analyzing employee performance data to recognize outstanding workers.
- Analysing Locations: Analyze the 'location' table, 'water_source' table , and the 'global_water_access' table to understand the distribution of water sources, gain insights into the types and locations of water sources, and understand the global water access situation.
- Start of a solution: Rank each type of water source based on the total number of people served by each source to help prioritize which sources should be fixed or improved first, focusing on those that affect the most people.
- Analysing queues: Analyzing the data to understand water collection patterns and peak times for water collection.

Tasks Overview (Project 3)

- Generating an ERD: Analyze the existing database structure to identify tables, relationships, and key attributes.
- Integrating the Report: Integrating the auditor report data into the database by establishing relationships between the auditor report data and other relevant tables in the database to ensure data integrity and consistency.
- Joining Employee Data to the Report: Use SQL JOIN operations to link employee data (e.g., employee ID, name) to the auditor report data, providing context and additional information for analysis.
- Gathering Evidence: Use of a complex SQL query to retrieve specific data from multiple tables, incorporating joins, aggregations, and conditions. Using the query results to gather evidence and insights, supporting decision-making and problem-solving related to water project management.

Tasks Overview (Project 4)

- Joining Pieces Together: Use SQL joins to combine data from different tables to create a comprehensive dataset for analysis.
- Summary Report: Create a summary report highlighting the findings and recommendations from the analysis for stakeholders and decision makers.
- A Practical Plan: Developing a practical plan based on the analysis, outlining actionable steps to address water quality and pollution issues.