Integration Documentation: SDG 6 - Clean Water and Sanitation

---

Project Title: Enhancing Access to Clean Water in Rural Communities

1. Overview of Data Integration

This documentation outlines the steps taken to integrate SQL data with Microsoft Excel for the purpose of analyzing clean water access in rural communities. The process involved designing a relational database to capture water access, infrastructure, and socio-economic factors, exporting data into CSV files, and performing data analysis in Excel. The final analysis is presented through interactive dashboards to visualize key insights.

---

2. Data Sources

- **Regions**: Contains data on rural areas, population, and water sources.
- **Water Sources**: Information on water quality, availability, and types of sources (e.g., wells, piped water).
- **Infrastructure**: Data on the availability of water treatment plants, pipe networks, and renewable water sources.
- **Socio-Economic Factors**: Indicators such as income levels, education, and employment rates in relation to water access.

---

3. Data Export from SQL Database

1. **SQL Queries** were written to extract data from the relational database:
    - Example query for retrieving water access data:

      ```
      sql
      Copy code
      SELECT RegionName, WaterSource, AccessLevel, Infrastructure, Year
      FROM RegionWaterAccess;
      ```

    - The results were exported as **CSV files** for each dataset.
2. **Data Export Process**:
    - SQL queries were executed in SQL Server Management Studio.
    - Results were saved as CSV files and organized into separate datasets: **Regions**, **Water Sources**, **Infrastructure**, and **Socio-Economic Factors**.

---

4. Data Import into Excel

1. **Steps to Import CSV Files into Excel**:
    - Open Excel, go to **Data** > **Get Data** > **From Text/CSV**.
    - Select the CSV file for each dataset (Regions, Water Sources, etc.).
    - Ensure correct data types (e.g., numeric, text) are applied to each column during import.
2. **Data Formatting**:
    - Data was formatted into Excel Tables for easy referencing in pivot tables and charts.
    - Columns such as *Access Level* and *Infrastructure Coverage* were formatted as percentages or numerical values for analysis.

---

5. Creating Relationships in Excel

Using Excel's **Data Model**, relationships between tables were established:

- **Regions[RegionID]** was linked with **WaterSources[RegionID]** and **Infrastructure[RegionID]**.
- This allowed for the creation of pivot tables and cross-table analysis in the final dashboard.

---

6. Data Analysis and Dashboard

- **Pivot Tables** were created to summarize clean water access per region and compare it with socio-economic indicators.
- **Charts and Graphs**: Line graphs, bar charts, and pie charts were used to visualize the data.
- **Interactive Dashboard**: Slicers and filters were added to allow users to explore water access data by region, infrastructure, and water source.

---

7. Data Consistency and Validation

- **Cross-Validation**: Data in Excel was cross-checked with SQL query results to ensure consistency.
- **Data Cleansing**: Any missing or anomalous values were cleaned, and foreign key relationships were validated to ensure referential integrity.

8. Conclusion

The integration process successfully brought SQL data into Excel for detailed analysis of water access in rural communities. Through the use of Excel's data modeling tools and interactive dashboards, key insights on clean water access were uncovered, which will help inform decisions in line with SDG 6.