Data Cleaning Report

Dataset

- cleaned_global_water_consumption.csv (Original)
- cleaned_global_water_consumption2.csv (Modified)

Modifications Made

1. Computed Sectoral Usage Percent Sum

New Column Created: Total_Percent

Operation: Summation of sectoral water use percentages:

Total_Percent = Agricultural Water Use (%) + Industrial Water Use (%) + Household Water Use (%)

Purpose: To validate the internal consistency of water use reporting.

2. Calculated Actual Water Usage by Sector

New Columns Created: Agricultural_Water_BCM, Industrial_Water_BCM, Household_Water_BCM

Operation: Converted percentages into actual water use (BCM):

Sector_BCM = Total Water Consumption × Sector Percentage ÷ 100

Purpose: To enable meaningful absolute comparisons across countries and sectors.

3. Verified Sectoral Summation vs. Reported Total

New Columns Created: Sector_Sum, Difference

Operation: Compared the sum of sectoral water use to reported total:

Sector_Sum = Agricultural_Water_BCM + Industrial_Water_BCM + Household_Water_BCM

Difference = Reported Total - Sector_Sum

Purpose: To detect any inconsistencies or rounding issues in sector-level data.

4. Created Groundwater Stress Metric

New Column Created: Groundwater_Stress_Index

Operation: Computed as Groundwater Depletion Rate ÷ Rainfall Impact

Purpose: To assess groundwater stress by comparing depletion to rainfall-based replenishment.