Data Collection and Preprocessing Phase

Date	15 December 2025
Team ID	SWUID20250254709 (Anshika Arya)
Project Title	Global Energy Trends: A Comprehensive Analysis of Key Regions and Generation Modes
Maximum Marks	3 Marks

Data Quality Report

This report summarizes the data quality findings across the supplied CSVs, the severity of each issue and the planned resolution steps (Power Query actions).

Data Source	Data Quality Issue	Severity	Resolution Plan
Continent_Consumption_TWH.csv	• Inconsistent column naming, potential trailing spaces. • Some zero or missing values for specific years in small regions.	Moderate	• Trim text, standardize names through Power Query, then rename columns to snake_case. • Inspect missing pattern, if sparse, keep blanks. If critical for an analysis view, document and replace with null or interpolate only for visualization
Country_Consumption_TWH.csv	• Wide format, not analysis friendly, also inconsistent country name spellings. • Missing values in certain country-year cells.	High	• Unpivot all country columns into long format, create a country mapping/lookup to normalize names. • If missing is small, document and leave blank. If many rows missing for some countries, flag and exclude from country-level comparisons or annotate in the report.

renewablePowerGeneration97-17.csv	• Year range mismatch vs other datasets. • Column names include parentheses and inconsistent formatting.	Moderate	• Document exact Year range in metadata. Use year filters in visuals to align comparisons or aggregate to common intersection years when comparing datasets. • Standardize column names in Power Query.
renewables TotalPower Generation .csv	• Presence of pre-aggregated "Total" rows and trailing spaces in Mode of Generation values. • Some numeric columns stored as text with commas.	Moderate	• Remove 'Total' rows from mode tables or keep as separate validation rows, Trim strings, Convert Contribution column to numeric, add Energy_Type column for easier slicers. • Use Replace Values & Change Type to remove commas and convert to Decimal Number.
nonRenew ablesTotal PowerGene ration.csv	• Presence of pre-aggregated "Total" rows and trailing spaces in Mode of Generation values. • Some numeric columns stored as text with commas.	Moderate	• Remove 'Total' rows from mode tables or keep as separate validation rows, Trim strings, Convert Contribution column to numeric, add Energy_Type column for easier slicers. • Use Replace Values & Change Type to remove commas and convert to Decimal Number.
top20Coun triesPower Generation .csv	Ensure numeric total columns have no thousand separators and Country names are consistent.	Moderate	Trim names, convert numeric columns to Decimal, unpivot if required for stacked bar visuals.