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	10 Marks

Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	The analysis uses six primary CSV datasets containing multi-year energy values measured in terawatt hours (TWh): continent-level consumption, country-level consumption, renewable generation by source, renewable totals by mode, non-renewable totals by mode and renewable breakdown for the top 20 countries. These datasets together enable trend, regional and country-level comparisons of energy production and consumption.
Data Cleaning	Data cleaning focused on removing blanks, trimming whitespace, standardizing column names, converting types and eliminating pre-aggregated “Total” rows that could distort analysis. Wide format tables were unpivoted to long format to enable flexible aggregation and visualization in Power BI.
Data Transformation	Transformations included unpivoting wide tables, creating an Energy_Type flag for mode tables and normalizing source names. Feature engineering produced derived fields and measures such as Total_TWh, Renewable_Share (%) and manual YoY growth calculations based on the Year column.

Data Type Conversion	All date/year fields were converted to whole numbers and energy values to numeric (decimal) types, ensuring correct aggregation and DAX calculations. Text fields such as country and region names were standardized to consistent casing and spacing to avoid grouping errors in visuals.
Column Splitting and Merging	Minimal splitting/merging was required after unpivoting; compound headers were renamed rather than split. Where country names included extra metadata, columns were split by delimiter and cleaned to ensure a single authoritative Country column for mapping and joins.
Data Modeling	The model uses cleaned fact tables for renewables, non-renewables, continent consumption and country consumption, with an optional Dim_Year or Dim_Country lookup for cleaner relationships. In the absence of a full date table, Year was treated as a numeric field and measures were implemented with manual time-intelligence logic.
Save Processed Data	After transformation, cleaned tables were loaded into the Power BI model and the. pbix saved as Global_Energy_Trends.pbix. For reproducibility, processed tables can be exported as CSVs and Power Query M steps documented in the appendix.