Data Collection and Preprocessing Phase

Date	2 October 2025
Team ID	SWUID20250206509
Project Title	Global Energy Trends: A Comprehensive Analysis of Key Regions and Generation Modes using Power BI
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 dataset contains global energy generation statistics from IEA and EIA sources, covering 2005-2025. It includes data across six major regions (North America, Europe, Asia-Pacific, Middle East, Africa, South America) and eight generation modes (coal, natural gas, oil, nuclear, hydro, solar, wind, other renewables). The dataset comprises approximately 500,000+ records with columns for region, year, energy type, generation capacity (TWh), percentage mix, and carbon emissions.
Data Cleaning	Handled missing values in generation capacity and emissions columns using forward-fill and mean imputation methods. Removed 1,247 duplicate records from multiple reporting sources. Corrected inconsistent country naming conventions and fixed data entry errors including negative values and outliers exceeding capacity limits.
Data Transformation	Used Power Query to filter data for 2005-2025, sorted chronologically by year and region. Pivoted energy generation modes for easier comparison. Created calculated columns for Renewable Energy Percentage, Fossil Fuel Percentage, Year-over-Year Growth Rate, and Carbon Intensity Index. Grouped data by region for aggregates and comparative analysis.

Data Type Conversion	Converted 'Year' to whole number, generation capacity to decimal, percentages to percentage format with 2 decimals, dates to DateTime format, and categorical fields to appropriate data types for optimization.
Column Splitting and Merging	**Column Operations**: Split 'Location' column into separate 'Country' and 'Region' columns using delimiter. Merged 'Month' and 'Year' columns into a single 'Date' column for time-series analysis. Split 'Energy_Type_Details' column to separate 'Primary_Energy_Type' and 'Sub_Category'. Merged multiple renewable energy columns (Solar_PV, Solar_Thermal) into consolidated 'Solar' column. Created combined 'Country_Region' column for hierarchical drill-down functionality. Split 'Capacity_Unit' to standardize all measurements to TWh (Terawatt-hours).
Data Modeling	Define relationships between tables and create measures.
Save Processed Data	Save the cleaned and processed data for future use.