Data Collection and Preprocessing Phase

Date	17 Dec 2025
Team ID	xxxxxx
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

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	I have **6 CSV files** related to global energy consumption and power generation. Example shapes: • *Continent_Consumption_TWH.csv* → 31 rows × 12 columns • *Country_Consumption_TWH.csv* → 255 rows × 12 columns • *nonRenewablesTotalPowerGeneration.csv* → 232 rows × 2 columns • *renewablePowerGeneration97-17.csv* → 434 rows × 43 columns • *renewablesTotalPowerGeneration.csv* → 124 rows × 6 columns • *top20CountriesPowerGeneration.csv* → 20 rows × 6 columns
Data Cleaning	• Checked for missing values (some files may have NA for certain energy sources). • Removed duplicated country/year entries if found. • Standardized column names (e.g., remove spaces, unify "TWh" suffix). • Corrected inconsistent values like **0.000** where energy source was not available.
Data Transformation	• Filtered data for specific years (e.g., 1990–2017). • Sorted countries by highest/lowest power generation. • Pivoted *renewablePowerGeneration97-17* to create year-wise

	energy category tables. • Created calculated columns such as: ◦ **Total_Renewables = Hydro + Solar + Biofuel + Geothermal** ◦ **Per_Capita_Consumption (if population added)**
Data Type Conversion	• Converted **Year → Whole Number** • Converted all energy columns (**TWh values → Decimal Number**) • Standardized text columns (**Country**, **Continent**) to Text type
Column Splitting and Merging	• Split combined columns if needed (e.g., "Hydro(TWh)" → "Hydro"). • Merged datasets using keys like **Year**, **Country**, **Continent**.
Data Modeling	• **Relationships created:** ◦ Country table ↔ Renewable generation table (Country) ◦ Country table ↔ Non-renewable table (Country) ◦ Year dimension ↔ All fact tables • **DAX Measures:** ◦ *Total Renewable Power = SUM(Table[Total (TWh)])* ◦ *Total Non-Renewable Power = SUM(Table[Value])* ◦ *Percentage Renewable = Total Renewable / Total Power*
Save Processed Data	• Exported cleaned data to **Power BI PBIX**. ✓ Energy_Model.pbix