**Data Collection and Preprocessing**

| Date | 16 July 2025 |
| --- | --- |
| Team ID | xxxxxx |
| Project Title | Exploring Global Energy Generation |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing**

| **Section** | **Description** |
| --- | --- |
| Data Overview | The dataset contains global energy generation data from multiple countries over several years. It includes variables such as country, continent, year, energy source (coal, gas, renewables, etc.), and energy generated in Terawatt-hour (TWh). |
| Data Cleaning | * Handled missing values by removing incomplete records or filling them using forward-fill method. * Removed duplicate rows identified by matching country, year, and energy type. * Corrected inconsistent country names (e.g., "United States" vs "USA"). |
| Data Transformation | * Used Power Query to filter out countries with insufficient data. * Sorted data by year for time-series analysis. * Pivoted energy source columns for better comparison. * Created calculated columns for total energy generation per region and renewable share. |
| Data Type Conversion | * Converted year column to Date or Integer type. * Ensured energy values are stored as numeric (float) types. * Corrected any string-to-number inconsistencies in calculated columns. |
| Column Splitting and Merging | * Merged multiple renewable types (solar, wind, hydro) into a single “Renewables” column for simplified analysis. |
| Data Modeling | * Defined relationships between **Country Table**, **Energy Generation Table**, and **Region Table**. * Created measures such as **Total Generation**, **% Renewables**, and **Growth Rate** using Power BI DAX. |
| Save Processed Data | * Saved the cleaned and transformed dataset as an Excel file (.xlsx) and imported into Power BI for further visualization and modeling. Backup version saved in .csv format for reproducibility. |