Data Collection and Preprocessing Phase

Date	30 July 2025
Team ID	xxxxxx
Project Title	**Global Energy Trends**: A Data-Driven Analysis of Consumption Patterns & Renewable Transition (1990-2020)
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	*"Six datasets of global energy trends and covering Continent and country-level energy consumption, production, and renewable and non-renewable ,* **45+ countries** (1990–2020) **5 energy sources** (Hydro,Solar,Wind,Biofuel,Geothermal) **Key Metrics:** Consumption (TWh), *These are brief overview of datasets which underpin our analysis of global energy patterns and transitions to renewable in Power BI."*
Data Cleaning	*Reviewed datasets for missing values and removed rows with nulls. Eliminated duplicate entries. Corrected rounding issues in consumption figures. Ensured key columns ("Year", "Index") remained unchanged for modeling and relational integrity.*
Data Transformation	*Performed "Unpivot Columns" in Power Query to transform wide data (continents/groups as columns) into long format. Renamed columns for clarity ("Continent", "TWH"), duplicated queries to preserve raw data, and maintained key columns (Year, Index) for relationships. Transformation prepares data for flexible analysis and visualization in Power BI.*

Data Type Conversion	Converted date fields to proper date format and numeric fields to appropriate decimal/integer types.
Column Splitting and Merging	*In both* **Country_Consumption** *and* **Country_Generation** *datasets, all columns (except* **Index** *and* **Year**) were merged to create a single Combined Metrics *column for streamlined analysis. This helped simplify the structure while retaining key identifiers for filtering and comparison.*
Data Modeling	Created a new column named **Index** in each datasets to serve as a unique identifier. Established relationships between multiple datasets (e.g., consumption, generation, renewable trends) using the **Index** field in Power BI's model view. This enabled accurate joining and filtering across tables for unified analysis**.**
Save Processed Data	Cleaned and transformed data stored in Power BI data model (.pbix) for use in dashboard creation and future updates.