

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	3 Marks

Data Quality Report Template

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle - Renewable Power Generation Dataset	Missing generation capacity values for certain regions and years in the dataset	Moderate	Applied forward-fill method for consecutive missing years and mean imputation based on regional averages for isolated gaps
Kaggle - Renewable Power Generation Dataset	1,247 duplicate records due to multiple entries for same country-year-energy type combinations	High	Identified duplicates using composite key (Country+Year+EnergyType), retained most recent entries, removed exact duplicates

Kaggle - Renewable Power Generation Dataset	Outliers in generation data exceeding realistic capacity limits for certain regions	Moderate	Applied statistical outlier detection (values beyond 3 standard deviations), validated against reasonable benchmarks, corrected or removed anomalies
Kaggle - Renewable Power Generation Dataset	Data type inconsistencies with numeric values stored as text format	Low	Converted all numeric columns to appropriate data types (decimal, integer) using Power Query data type conversion