# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 1 October 2025 |
| Team ID | SWUID20250207636 |
| Project Title | **Global Energy Trends: A Comprehensive Analysis of Key Regions and Generation Modes using Power BI** |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan**

| Section | Description |
|---|---|
| Project Overview | The project is the Global Energy Transition Analytics & Forecasting Platform. The objective was to utilize historical global energy consumption and power generation data to create a comprehensive, interactive dashboard for trend analysis and insights. The project is complete, with the final output delivered as the Global Energy Trends.pbix Power BI file. |
| Data Collection Plan | Data collection is complete. The raw data was obtained from a single, specific Kaggle Global Energy Trends dataset which contained the 6 CSV files used for analysis. These files were subsequently ingested, cleaned, transformed, and modeled within the Power BI environment to produce the final dashboard. This plan documents the raw files used from that original source. |
| Raw Data Sources Identified | The raw data consists of six distinct CSV files sourced from Kaggle, plus the Power BI file which is the project deliverable. |

**Raw Data Sources**

| Source Name | Description | Location/URL | Format | Size | Access Permissions |
|---|---|---|---|---|---|
| **Continent_Consumption_TWH.csv** | Time-series data detailing total energy consumption in Terawatt-hours (TWh), aggregated by major geographic regions (e.g., OECD, BRICS, continents). | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Small (< 1 MB) | Public |
| **Country_Consumption_TWH.csv** | Time-series data detailing total energy consumption in TWh for a comprehensive list of global countries. | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Small (< 1 MB) | Public |
| **Renewable Power Generation97-17.csv** | Time-series data detailing the historical contribution (TWh) of specific major renewable sources between 1990 and 2017. | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Small (< 1 MB) | Public |
| **Renewables Total Power Generation.csv** | Snapshot data summarizing the total global power generation contribution by various renewable modes (e.g., Hydro, Wind, Solar PV). | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Very Small (< 1 MB) | Public |
| **Non Renewables Total Power Generation.csv** | Snapshot data summarizing the total global power generation contribution by various non-renewable modes (e.g., Coal, Natural Gas, Nuclear, Oil). | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Very Small (< 1 MB) | Public |

| | | | | | |
|---|---|---|---|---|---|
| **Top 20 Countries Power Generation.csv** | Country-specific data detailing the power generated by the top 20 countries, broken down by major renewable sources. | Kaggle Dataset (Global Energy Consumption & Renewable Generation \| Kaggle) | CSV | Very Small (< 1 MB) | Public |