**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 July 2025 |
| Team ID | SWTID1750360304 |
| Project Title | Power consumption analysis for households |
| Maximum Marks | 2 Marks |

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

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 Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This machine learning project aims to analyze household power consumption patterns to optimize energy efficiency and reduce costs. By leveraging smart meter data, weather information, and appliance-level measurements, the system will identify wasteful consumption, predict peak demand, and provide actionable recommendations for households and utility providers. |
| Data Collection Plan | The primary dataset comes from Kaggle, containing detailed power consumption records from smart meters collected over 47 months. This comprehensive dataset includes minute-by-minute measurements of active/reactive power, voltage, current intensity, and breakdowns of energy usage by specific appliance categories. Additional weather data may be integrated to analyze how external temperature affects consumption patterns. |
| Raw Data Sources Identified | The Kaggle dataset provides the core foundation, featuring 2,075,259 measurements with timestamps, power readings in kilowatts, voltage in volts, current in amperes, and three sub-metering categories tracking kitchen, laundry, and HVAC systems. A derived field calculates energy use from unmonitored appliances. The data includes approximately 1.25% missing values that will be addressed during preprocessing to ensure analysis accuracy. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | This dataset contains minute-level electricity consumption records from a single household collected over 47 months (December 2006 - November 2010). The measurements capture detailed power usage patterns with timestamps, including both main power metrics and appliance-specific consumption data. | https://www.kaggle.com/uciml/electric-power-consumption-data-set | CSV | 132.96 MB | Public |