**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 20 July 2025 |
| Team ID | SWTID1750360304 |
| Project Title | Power consumption analysis for households |
| Maximum Marks | 2 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** |
| Household Power Consumption | 1.25% missing values in measurements | Moderate | Interpolate time-series gaps or flag for exclusion if non-recoverable |
| Household Power Consumption | Duplicate timestamp entries | Low | Remove duplicates with drop\_duplicates() and validate temporal consistency |
| Household Power Consumption | Extreme outliers in power measurements | High | Cap values using IQR or domain thresholds (e.g., 99th percentile) |
| Household Power Consumption | Inconsistent timestamp formats (if any) | Low | Standardize datetime format (e.g., pd.to\_datetime()) |
| Household Power Consumption | Sensor drift (voltage/current anomalies) | Moderate | Apply rolling median correction or consult domain experts |
| Derived "Other Appliances" | Negative values (calculation artifacts) | High | Clamp to zero and validate formula logic |