**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 6 August 2025 |
| Team ID | xxxxxx |
| Project Title | Predicting Plant Growth Stages Using Environmental & Management Data in 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** |
| Dataset | * Missing values for some environmental parameters (e.g., temperature, humidity). * Duplicate records due to repeated data logging. * Inconsistent units (e.g., temperature in both °C and °F). * Text instead of numeric values in measurement columns. * Misaligned table headers or shifted columns during data import. * Outliers in growth rate values affecting prediction accuracy. | Low/ Moderate/ High | * Handle Missing Values: Use mean/median imputation for numeric data, forward-fill for time-series gaps. * Remove Duplicates: Identify and delete duplicate rows using unique plant IDs and timestamps. * Standardize Units: Convert all measurements to a single unit (e.g., °C for temperature, % for humidity). * Correct Data Types: Convert text-based numbers to numeric format, and ensure dates are in proper date format. * Fix Header Alignment: Re-import data with proper delimiter settings or manually adjust column headers. * Treat Outliers: Use statistical methods (e.g., IQR or Z-score) to detect and cap/replace extreme values. |
|  |  |  |  |