**Data Collection and Preprocessing Phase**

| Date | 10 Oct 2025 |
| --- | --- |
| Team ID | SWUID20250216152 |
| Project Title | Predicting plant growth stages with environmental and management factors |
| Maximum Marks | 3 Marks |

**Data Quality Report**

The dataset demonstrates high data quality and readiness for visualization and predictive modeling. Minor formatting issues were resolved, and validation confirmed consistent, reliable data.  
 This clean dataset ensures accurate results during Power BI analysis and model prediction for plant growth stages

| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| --- | --- | --- | --- |
| Plant Growth Dataset | Minor inconsistencies in data formatting (e.g., inconsistent capitalization in categorical columns like Soil\_Type or Fertilizer\_Type). | Low | Standardized text format using lowercase conversion and value normalization before analysis. |
| Plant Growth Dataset | Potential noise or minor variation in environmental readings (Temperature and Humidity) due to sensor precision. | Moderate | Applied statistical smoothing and validated sensor calibration to maintain reliable readings. |
| Plant Growth Dataset | The dataset lacks a few secondary attributes (e.g., CO₂ levels, soil pH variations) that could further improve predictions. | Low | Proceeded with available attributes and included placeholders for future dataset expansion. |
| Plant Growth Dataset | No missing values found, but validation performed to ensure data consistency across all records. | Low | Verified all columns and confirmed data completeness using Power BI and Python checks. |

| Plant Growth Dataset | Outlier check performed for numeric columns — no major outliers detected. | Low | IQR-based outlier detection used to validate numeric stability. |
| --- | --- | --- | --- |