# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 24 July 2025 |
| Team ID | |
| Project Title | Predicting plant growth stages with environmental and management data using power bi |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing Template**

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

| **Section** | **Description** |
|---|---|
| Data Overview | Dataset contains **193 rows** and **7 columns**:<br>Soil_Type, Sunlight_Hours, Water_Frequency, Fertilizer_Type, Temperature , Humidity, and Growth_Milestone. |
| Data Cleaning | No missing values found in any column. All values are complete and consistent. No duplicate check needed based on initial review. |
| Data Transformation | Could involve creating derived columns like Water_Frequency_Days or grouping Growth_Milestone into low/medium/high categories for analysis. |
| Data Type Conversion | Most columns are correctly typed: Temperature, Humidity, and Sunlight_Hours as floats, Growth_Milestone as integer, and others as categorical/objects. |
| Column Splitting and Merging | No combined columns are present, but possible merging:<br>e.g., combining Soil_Type and Water_Frequency for interaction effects. |

| | |
|---|---|
| Data Modeling | This flat file could be linked to reference tables (e.g., Fertilizer_Info, Soil_Properties) in a dashboard or schema. |
| Save Processed Data | After preprocessing, data can be saved as **processed_data.csv** or loaded into Power BI for dashboard creation. |