**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 June 2025 |
| Team ID | xxxxxx |
| 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**: 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:  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\_growth\_data.csv** or loaded into Power BI for dashboard creation. |