# Data Collection and Preprocessing Phase

| | |
|---|---|
| Date | 28-07-2025 |
| Team ID | HK |
| Project Title | Predicting Plant Growth Stages with Environmental and Management Data Using Power BI |
| Maximum Marks | 10 Marks |

## Data Exploration and Preprocessing

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

## Data Exploration and Preprocessing

| Section | Description |
|---|---|
| **Data Overview** | The dataset contains 193 records and 7 columns, including: Soil Type, Sunlight Hours, Water Frequency, Fertilizer Type, Temperature, Humidity, and Growth Milestone. These fields are used to understand the relationship between environmental and input factors on plant growth. |
| **Data Cleaning** | Minor text inconsistencies in categorical fields were normalized (e.g., “organic” vs. “Org”). - All entries verified for logical accuracy (e.g., temperature range and humidity values). |
| **Data Transformation** | Used Power Query for: Filtering data by soil type and fertilizer, sorting by growth milestones, creating new calculated columns (e.g., **Growth_per_Hour = Growth_Milestone / Sunlight_Hours**), Pivoting to analyse fertilizer performance across soil types |
| **Data Type Conversion** | Converted Soil Type, Fertilizer Type, and Water Frequency to text format. - Ensured Temperature, Humidity, Sunlight Hours, and Growth Milestone are in numeric format. |
| **Column Splitting and Merging** | No splitting required. - Merged environmental metrics (Humidity, Temperature, Sunlight Hours) to form an Environmental Score for advanced insights. |

| | |
|---|---|
| **Data Modelling** | Single-table model used (no complex relationships needed).<br>DAX measures created for insights: Average Growth, Growth Rate per Temperature, Max Growth by Soil Type - Interactive slicers and filters added for soil, water, and fertilizer type. |
| **Save Processed Data** | Cleaned dataset saved within Power BI (.pbix) file. - Backup version of the processed data exported to Excel and CSV for reuse and external analysis. |