# Data Collection and Preprocessing Phase

| Date | 25 December 2025 |
|---|---|
| Project Title | Predicting Plant Growth Stages with Environmental and Management Data Using Power BI. |
| Maximum Marks | 2 Marks |

**Project Overview:**

1. This project aims to develop an interactive Data Analytics and Prediction Dashboard using Microsoft Power BI to address the high rate of crop failure in agricultural operations. Currently, the farming process suffers from a 50% failure rate (based on historical data of 193 trials) due to a reliance on intuition rather than data-driven decision-making.

2. By analyzing the plant_growth_data.csv dataset, this project will identify the critical "Golden Rules" for plant success-specifically quantifying the impact of **Organic vs. Chemical fertilizers** and optimizing **Watering Frequency** for different **Soil Types** (Loam, Sandy, Clay).

3. The final deliverable will be a visual dashboard that empowers agricultural managers to:

   - **Monitor** real-time success rates.

   - **Diagnose** the root causes of failure (e.g., identifying that Clay soil with Chemical fertilizer is a high-risk combination).

   - **Predict** the likelihood of a plant reaching its "Growth Milestone" based on current environmental conditions (such as Temperature, Humidity) and care inputs.

**Data Collection Plan:**

| S.No. | What data will be collected? | What data will be collected? | How will the data be collected? | When will the data be collected? |
|---|---|---|---|---|
| 1. | Environmental Conditions Data | To analyze how natural factors like **Temperature**, **Humidity**, and **Sunlight Hours** impact plant growth success. | **Download** from the project data archive (archive.zip). | 17/12/2025 |
| 2. | Agricultural Management Logs | To compare the effectiveness of different care inputs, specifically **Fertilizer Type** (Organic vs. Chemical) and **Watering Frequency**. | **Extracted** from the primary dataset (plant_growth_data.csv). | 17/12/2025 |
| 3. | Growth Outcome Metrics | To serve as the "Target Variable" (Growth_Milestone) for the Power BI predictive model. | **Included** in the source CSV file. | 17/12/2025 |

**Raw Data Sources:**

| Dataset Name | Source | Description | Format | Size | Accessibility |
|---|---|---|---|---|---|
| **Plant_growth_ data.csv** | **Open-Source Repository (e.g., Kaggle)** | This dataset contains **193 records** of plant growth trials which includes the 7 columns and the success indicator (0/1). | **CSV** | **15KB** | **Public** |