**Final Project Report**

# 1. Introduction

## 1.1 Project Overview

This project focuses on predicting plant growth stages using environmental and management data. Power BI is used to visualize patterns, support decisions, and guide agricultural planning.

## 1.2 Objectives

The objective is to identify key environmental and operational factors (e.g., soil type, water frequency, temperature, sunlight) that significantly affect plant growth. By visualizing this data, agricultural practices can be optimized.

**2. Project Initialization and Planning Phase**

**Project Overview**

**Objective**To predict plant growth stages using key environmental and management data by leveraging Power BI for interactive data analysis, visualization, and decision-making support.

**Scope**  
This project will focus on analyzing and visualizing environmental (e.g., temperature, humidity, sunlight) and management (e.g., soil type, water frequency, fertilizer) factors influencing plant growth. The scope includes building Power BI dashboards, applying decomposition trees for insights, and providing scenario-based recommendations. The project does not include real-time data capture or physical system deployment.

**Problem** **Statement**

**Description**  
Agriculture companies face difficulty in determining the ideal combination of environmental and resource management factors that lead to optimal plant growth. Raw datasets often lack clarity, making it hard to extract insights that can guide daily decisions in greenhouses or farms.

**Impact**  
Solving this problem enables agriculture-based businesses to optimize growing conditions, standardize best practices across locations, and increase overall plant yield and quality. It also enhances decision-making and promotes sustainable resource use.

**Proposed Solution**

**Approach**

* Import and clean the dataset in Power BI.
* Model the data with calculated columns and DAX measures.
* Create visualizations such as decomposition trees, clustered charts, and scatter plots.
* Analyze scenarios (ABC Greenhouses, GreenEarth Farms, FutureGrow Tech) through dynamic dashboards.
* Draw conclusions to recommend ideal conditions for plant growth.

**Key Features**

* Interactive dashboards for real-time filtering and exploration.
* Decomposition tree to drill down into growth milestone determinants.
* Scenario-based insights for different agri-business use cases.
* Visually engaging charts to simplify complex data for decision-makers.

# Resource Requirements

| Resource Type | Description | Specification/Allocation |
| --- | --- | --- |
| Hardware | Computing Resources | Standard laptop/desktop with 4-core CPU |
|  | Memory | 8 GB RAM |
|  | Storage | 500 GB HDD/SSD |
| Software | Frameworks | Microsoft Power BI |
|  | Libraries | Power Query M, DAX (native to Power BI) |
|  | Development Environment | Power BI Desktop, Excel, GitHub (for version control) |
| Data | Data Source | Internal agricultural dataset from XYZ Company |
|  | Format | CSV/Excel format with tabular data |
|  | Size | ~10,000 records covering environmental and management variables |

| **Sprint** | **Functional Requirement (Epic)** | **User Story Number** | **User Story / Task** | **Story Points** | **Priority** | **Sprint Start Date** | **Sprint End Date (Planned)** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sprint-1 | Project Setup & Planning | USN-1 | As a team, we want to define problem statements and scope for better project direction. | 2 | High | 15/06/2025 | 17/06/2025 |
| Sprint-1 | Dataset Review & Cleaning | USN-2 | As a data analyst, I want to explore and clean the dataset to ensure it is ready for visualization. | 3 | High | 18/06/2025 | 20/06/2025 |
| Sprint-2 | Data Modeling in Power BI | USN-3 | As a Power BI user, I want to structure the data model and define necessary calculated columns and measures. | 2 | High | 21/06/2025 | 22/05/2025 |
| Sprint-2 | Data Visualization Development | USN-4 | As a user, I want to visualize plant growth stages through charts and graphs to understand the data clearly. | 3 | High | 22/05/2025 | 22/05/2025 |
| Sprint-3 | Decomposition Tree Analysis | USN-5 | As a team, we want to use a decomposition tree to analyze growth by soil, water, and sunlight factors. | 2 | Medium | 23/05/2025 | 23/05/2025 |
| Sprint-3 | Dashboard Design & Final Report | USN-6 | As a project team, we want to design a dashboard and finalize the story/report for submission. | 4 | High | 24/05/2025 | 26/05/2025 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Problem**  **Statement (PS)** | **I am**  **(Customer)** | **I’m trying to** | **But** | **Because** | **Which makes me feel** |
| PS-1 | an agricultural manager at a greenhouse facility responsible for maintaining consistent crop growth. | identify the ideal environmental and management conditions—like soil type, watering frequency, sunlight exposure—that lead to optimal plant growth. | I don’t have an effective way to analyze the large amount of growth data collected from different greenhouses and conditions. | the data is complex, varied, and I lack tools that can break it down and reveal actionable insights visually. | frustrated and unsure about how to make informed decisions to improve plant health and yield. |
| PS-2 | a technology lead at an agri-tech company developing smart farming solutions. | validate the impact of real-time sensor-based environmental controls on plant growth performance. | I cannot clearly see which combinations of conditions (temperature, humidity, soil type) actually correlate with higher growth stages. | raw data alone is not enough—I need intuitive dashboards and predictive tools that visualize outcomes across scenarios. | uncertain about my technology’s effectiveness and less confident in communicating its value to stakeholders. |

# 3. Data Collection and Preprocessing Phase

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This project aims to predict plant growth stages by analyzing environmental and management factors using Power BI visualizations and dashboards. |
| Data Collection Plan | The data was collected from controlled agricultural experiments conducted by XYZ Company across multiple greenhouse environments. It includes detailed observations of soil type, sunlight hours, watering frequency, fertilizer type, temperature, and humidity—alongside growth milestone outcomes. |
| Raw Data Sources Identified | The dataset was curated internally and stored in tabular format (CSV) to facilitate easy import and manipulation in Power BI. External sources were not used for this phase. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | Internal dataset containing 193 records on plant growth variables (soil, sunlight, water, etc.) | <https://drive.google.com/file/d/1d6cqy-A20LSB2Q-gtbj9Y04Og6cQrJBC/view?usp=sharing> | CSV | 11 KB | Public |
| Dataset 2 | Description of the data in this source. | <https://drive.google.com/file/d/1TvuTb3aRdPTxW1S3cOwHhs4XhNkg4y9u/view?usp=sharing> | CSV | 6.5 MB | Private (with access) |

**Data Quality Report Template**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | No missing values detected in any columns. | Low | No action required; dataset completeness is confirmed. |
| Dataset | All categorical fields (e.g., Soil\_Type, Water\_Frequency) are in consistent text format. | Low | Dataset is already clean and ready; these fields will be used directly in Power BI as dimensions with filters/slicers. |
| Dataset | Column names are readable but may require renaming (e.g., replace underscores with spaces) for dashboard use. | Low | Rename column headers in Power BI visuals only for clarity (e.g., "Soil Type" instead of "Soil\_Type"). |
| Dataset | No standard labeling of binary column Growth\_Milestone (0/1 unclear to end user) | Moderate | Add a new calculated column or legend in Power BI to map 0 = "Not Achieved", 1 = "Achieved" for user clarity. |
| Dataset | Units for numerical columns (temperature in °C, humidity in %) are not explicitly labeled in the dataset. | Moderate | Add units in axis labels and tooltip descriptions in Power BI visuals to avoid misinterpretation. |

**Data Exploration and Preprocessing Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Data Overview | The dataset contains 193 records with 7 columns. It includes both environmental (sunlight hours, temperature, humidity) and management (soil type, fertilizer, water frequency) factors affecting plant growth. The target variable is Growth\_Milestone, indicating whether a plant has achieved a key growth stage. |
| Data Cleaning | Upon review, the dataset contains no missing values, no duplicate records, and all categorical values are consistent. Therefore, no additional cleaning was required. |
| Data Transformation | Performed using Power Query in Power BI:  • Applied column renaming for better readability (e.g., Soil\_Type → Soil Type)  • Filtered irrelevant rows during import (none in this case)  • Added calculated columns for mapping growth status labels (e.g., 0 = "Not Achieved", 1 = "Achieved") |
| Data Type Conversion | Verified and corrected data types:  • Sunlight\_Hours, Temperature, Humidity → Float/Decimal  • Growth\_Milestone → Whole Number  • Categorical variables retained as Text |
| Column Splitting and Merging | No splitting or merging was required. All columns were well-structured in the preprocessed file. |
| Data Modeling | Since only a single table is being used in this phase, no relationships were created. However, measures (e.g., average temperature, count of milestones) were created using DAX for deeper analysis in the dashboard. |
| Save Processed Data | The cleaned and transformed dataset is saved in Power BI Desktop (.pbix) format, using this as the source. All transformations are tracked in Power Query for transparency and reusability. |

# 4. Data Visualization

**Business Questions and Visualisation**

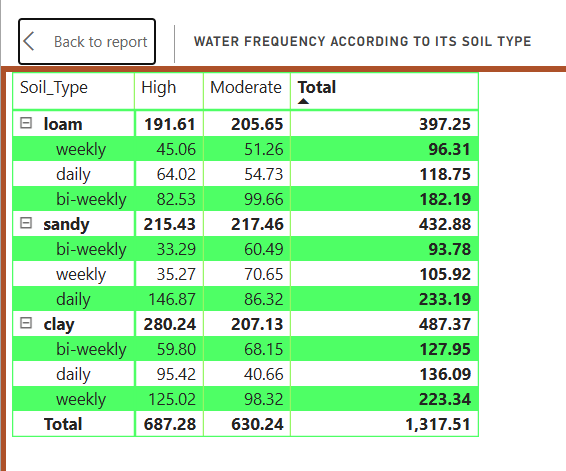
The process involves defining specific business questions to guide the creation of meaningful and actionable visualizations in Power BI. Well-framed questions help in identifying key metrics, selecting relevant data, and building visualisation that provide insights.

**1. What is the distribution of water frequency according to soil type?**

Visualization: Matrix table

Insight: Loam soil with high water frequency results in the highest total water frequency, indicating loam may benefit from more frequent watering compared to sandy and clay soils.

Screenshot:



**2. What is the average temperature across different temperature range descriptions?**

Visualization: Waterfall chart

Insight: Moderate temperature ranges exhibit the highest average temperature increase, suggesting these conditions are most favorable for plant growth.

Screenshot: A screenshot of a graph

AI-generated content may be incorrect.**3. How does temperature description impact plant growth?**

Visualization: Key influences

Insight:

Cold temperature ranges reduce plant growth, highlighting their negative effect.

Warm temperature ranges enhance growth, proving their positive contribution.

Screenshot:

A screenshot of a computer

AI-generated content may be incorrect.

**4. What is the growth milestone count by fertilizer type?**

Visualization: Donut Chart

Insight: Chemical fertilizers account for the highest growth milestone count, indicating higher effectiveness compared to organic or no fertilizer.

Screenshot:

A screenshot of a graph

AI-generated content may be incorrect.

**5. What is the average humidity based on humidity level description?**

Visualization: Ribbon chart

Insight: Humid conditions result in the highest average humidity, emphasizing the importance of adequate moisture for optimal plant growth.

Screenshot:

A graph of a number of bars

AI-generated content may be incorrect.

**6. How many growth milestones were achieved across different soil types?**

Visualization: Decomposition Tree

Insight: Loam soil leads with the highest count of growth milestones, suggesting its strong suitability for cultivation.

Screenshot:

A screenshot of a computer

AI-generated content may be incorrect.

**7. What is the average sunlight received by each soil type?**

Visualization: Bar chart

Insight: Sandy soil receives the most sunlight on average, potentially explaining its growth performance under high light conditions.

Screenshot:

A green pie chart with numbers and a green circle

AI-generated content may be incorrect.

**8. What is the growth milestone percentage achieved by each water frequency type?**

Visualization: Tree map

Insight: Daily watering results in the highest percentage of growth milestones, reinforcing the importance of frequent watering.

Screenshot:

A screenshot of a computer screen

AI-generated content may be incorrect.

**9. How does humidity level description relate to growth milestones?**

Visualization: Area chart

Insight: Humid conditions lead to the highest growth milestone count, proving the positive effect of controlled humidity.

Screenshot:

A green graph with numbers

AI-generated content may be incorrect.

# 5. Dashboard

**Dashboard Design**

Creating an effective dashboard is essential for converting raw data into actionable insights. The Power BI dashboard developed for the project *“Predicting Plant Growth Stages with Environmental and Management Data”* is a clean, interactive, and visually intuitive representation that supports agricultural analysis and decision-making.

This dashboard follows best design practices, including:

* **Clear and Intuitive Layout** – Visuals are well-organized with logical grouping of related insights (e.g., water frequency, temperature, humidity, growth milestones).
* **Appropriate Visualizations** – Matrix tables, pie charts, bar charts, waterfall visuals, and key influencer charts are used where relevant.
* **Consistent Color Theme** – Earthy tones aligned with agriculture (browns, greens) enhance readability and aesthetic appeal.
* **Interactive Filters and Slicers** – Enable exploration of soil type, temperature range, and water frequency.
* **Custom Visuals** – Waterfall and donut charts were customized for emphasis.
* **Use of Infographics and KPI logic** – Key metrics such as growth milestone count and averages are highlighted for quick understanding.

**Major Outcomes from the Dashboard:**

1. **Loam Soil with High Water Frequency Is Most Common:**  
   The matrix table reveals that loam soil, particularly with daily and moderate watering, leads to the highest total water frequency (397.25), suggesting it may require more frequent watering.
2. **Moderate Temperature Range Boosts Growth:**  
   The waterfall chart shows that moderate temperature ranges contribute the most to average temperature, supporting optimal plant growth.
3. **Warm Temperature Is the Strongest Positive Influencer:**  
   The key influencer visual highlights that warm temperature increases the average temperature significantly by 8.57°C, favoring growth conditions.
4. **Chemical Fertilizers Are Most Effective:**  
   A donut chart illustrates that chemical fertilizer usage leads to the highest number of growth milestones (38 out of 96), suggesting greater effectiveness.
5. **Humidity Level Strongly Impacts Growth:**  
   Both bar charts and milestone counts show that humid conditions lead to higher average humidity and the most growth milestones, underlining the importance of moisture.
6. **Loam Soil Supports Maximum Growth Milestones:**  
   The horizontal bar chart shows that loam soil accounts for the highest number of plant growth milestones (35 out of 96), making it the most conducive soil type.
7. **Sandy Soil Receives Maximum Sunlight:**  
   Sandy soil is associated with the highest average sunlight hours (~7.2 hrs), which could impact growth performance under adequate light.
8. **Daily Watering Has Highest Success Rate:**  
   The growth milestone percentage analysis shows daily watering is associated with the highest rate of plant growth milestones, emphasizing the importance of consistent watering schedules.

# 6. Report

**Report**

A report is a structured document designed to communicate data analysis results clearly and comprehensively. In this project, the Power BI report presents deep insights into plant growth patterns influenced by environmental and management variables such as sunlight, humidity, water frequency, temperature, and soil type.

The design process involved connecting the dataset, building interactive visualizations, applying consistent formatting, and organizing all components logically for an intuitive flow. This report caters to researchers, agronomists, and decision-makers who need actionable insights for agricultural planning and optimization.

**Observations and Insights from the Report**

**1. Average Conditions for Plant Growth**

The report includes three key performance indicators:

* Average Sunlight Hours: 6.83
* Average Humidity: 58.10
* Average Temperature: 25.08

These averages reflect a moderately warm and humid environment with sufficient daily sunlight exposure, which is favorable for most plant species.

**2. Sunlight Distribution by Soil Type**

The pie chart reveals that:

* Clay soil receives the highest average sunlight at 7.27 hours.
* Sandy soil follows at 6.76 hours.
* Loam soil receives the least at 6.41 hours.

This suggests that the location or structure of clay soil zones may naturally expose them to more sunlight.

**3. Growth Milestone Count by Humidity Level**

The line chart shows that:

* Moderate humidity results in the highest growth milestone count at 54.
* Dry conditions follow with 27.
* Humid environments lead to the least milestones at 15.

This implies that excess humidity may hinder growth, while moderate moisture levels are optimal.

**4. Water Frequency and Growth Milestone Percentage**

The tree map visual shows:

* Daily watering produces the highest percentage of growth milestones.
* Weekly watering has moderate success.
* Bi-weekly watering yields the lowest percentage.

This confirms that consistent watering is a critical factor in achieving growth stages.

**5. Textual Report Highlights**

The textual box on the report highlights:

* Humid conditions had the highest average humidity at 74.02, which is 79.53 percent higher than dry conditions at 41.23.
* Temperature range descriptions impact growth. Cold conditions caused the greatest decrease in temperature by 9.70.
* Clay soil not only receives the most sunlight but may also provide better conditions for certain crops.
* Moderate humidity had the highest milestone count, which was 260 percent higher than humid conditions.

**6. Overall Growth Milestone Count**

The gauge chart shows a total of 96 growth milestones, reinforcing the dataset’s robust size and diverse growing conditions.

# 7. Performance Testing

# 7.1 Utilization of DAX Expressions

# DAX (Data Analysis Expressions) is the core formula language used in Power BI to perform custom calculations and derive insights from data. It allows for creating dynamic measures and columns to support advanced analytics. In this project, DAX was employed extensively to compute metrics related to plant growth stages, environmental conditions, and category mapping. Below are the key DAX implementations used:

# Activity 2.1: Average Humidity as a *Measure*

# A DAX measure was created to calculate the overall average humidity across all records, providing a high-level environmental overview.

**Avg\_Humidity = AVERAGE(plant\_growth\_data[Humidity])**

# Activity 2.2: Average Sunlight Hours as a *Measure*

# This measure computes the mean value of sunlight hours to identify light exposure conditions that support plant growth.

**Avg\_Sunlight\_Hours = AVERAGE(plant\_growth\_data[Sunlight\_Hours])**

# Activity 2.3: Average Temperature as a *Measure*

# A measure was created using DAX to find the average temperature, helping to analyze thermal conditions influencing plant development. Avg\_Temperature = AVERAGE(plant\_growth\_data[Temperature])

# Activity 2.4: Growth Milestone Count as a *Measure*

# DAX was used to count the total number of growth milestones achieved across the dataset.

**Growth\_Milestone\_Count = COUNTROWS(FILTER(plant\_growth\_data,plant\_growth\_data[Growth\_Milestone]=1))**

# Activity 2.5: Growth Milestone Percentage as a *Measure*

# This measure calculates the percentage of records that achieved a growth milestone (1) relative to the total number of entries.

**Growth\_Milestone\_Percentage = DIVIDE([Growth\_Milestone\_Count],COUNTROWS(plant\_growth\_data),0)**

# Activity 2.6: Water Frequency Numeric as a *New Column*

# A calculated column was introduced to assign numeric values to watering frequencies (e.g., daily = 3, weekly = 2, bi-weekly = 1) for comparative analysis.

**Water\_Freq\_Nums = SWITCH([Water\_Frequency],"daily",1,"bi-weekly",2,"weekly",3,BLANK())**

# Activity 2.7: Temperature Range as a *New Column*

# This column categorizes temperature values into ranges (e.g., cold, moderate, warm) to support grouping in visualizations.

**Temp\_Range = SWITCH(TRUE(),[Temperature]<15,"Low", [Temperature]>=15 && [Temperature]<25,"Moderate",[Temperature]>=25, "High")**

# Activity 2.8: Humidity Range as a *New Column*

# A new column to classify humidity levels into ranges such as dry, moderate, and humid for better segmentation in visuals.

**Humidity\_Range = SWITCH(TRUE(),[Humidity]<40,"Low",[Humidity]>=40 && [Humidity]<60,"Moderate",[Humidity]>=60,"High")**

# Activity 2.9: Humidity Level Description as a *New Column*

# Based on humidity values, this column assigns a textual descriptor (e.g., “Moderate Humidity”), improving readability.

**Humidity\_Level\_Desc = SWITCH(TRUE(),[Humidity]<30,"Very Dry", [Humidity]>=30 && [Humidity]<50,"Dry", [Humidity]>=50 && [Humidity]<70,"Moderate", [Humidity]>=70 && [Humidity]<90,"Humid", [Humidity]>=90,"Very Humid")**

# Activity 2.10: Temperature Range Description as a *New Column*

# This derived column enhances understanding by providing descriptive labels to numerical temperature bands.

**Temperature\_Range\_Description = SWITCH(TRUE(),[Temperature] < 10, "Very Cold",[Temperature] >= 10 && [Temperature] < 20, "Cold",**

**[Temperature] >= 20 && [Temperature] < 30, "Moderate",**

**[Temperature] >= 30 && [Temperature] < 40, "Warm",**

**[Temperature] >= 40, "Hot")**

# Activity 2.11: Growth Milestone Description as a *New Column*

# A column translating binary milestone indicators (0 or 1) into clear descriptions like “Achieved” or “Not Achieved”.

**Growth\_Milestone\_Description = SWITCH([Growth\_Milestone],**

**0, "Early Stage",**

**1, "Mature Stage",**

**"Unknown Stage")**

# Activity 2.12: Plant Growth Category as a *New Column*

# An advanced classification column that groups entries into high, medium, or low growth categories based on multiple environmental factors.

**Plant\_Growth\_Category = SWITCH([Growth\_Milestone], 0, "Initial Growth", 1, "Advanced Growth", "Uncategorized")**

**7.3 Number of Visualizations**

A total of **11 visualizations** were developed using Power BI. These visual elements have been consistently used across the **Data Visualization**, **Dashboard**, and **Final Report** sections to extract insights from the dataset. Each chart addresses a specific business question related to plant growth and environmental conditions.

The charts included are:

1. **Water Frequency According to Its Soil Type**
2. **Average Temperature by Temperature Range Description**
3. **Temperature and Its Description According to Plant Growth**
4. **Growth Milestone Count by Fertilizer Type**
5. **Average Humidity by Humidity Level Description**
6. **Growth Milestone Count According to Its Soil Type**
7. **Average Sunlight Hours by Soil Type**
8. **Growth Milestone Percentage by Water Frequency**
9. **Growth Milestone Count by Humidity Level Description**
10. **Growth Milestone Count**
11. **Average Humidity, Average Temperature, Average Sunlight Hours**

These visualizations are integrated in:

* **Section 4: Data Visualization** – Business questions and insights
* **Section 5: Dashboard** – Interactive visual analytics summary
* **Section 6: Report** – Storytelling and final analytical insights

# 8. Conclusion/Observation

The project demonstrated how plant growth can be predicted using key environmental and resource management variables. Through Power BI, the data was transformed into interactive visuals that support data-driven agricultural decisions.

# 9. Future Scope

Future enhancements may include integrating real-time sensor data from IoT devices, expanding dataset sources, and using machine learning models within Power BI for predictive alerts.

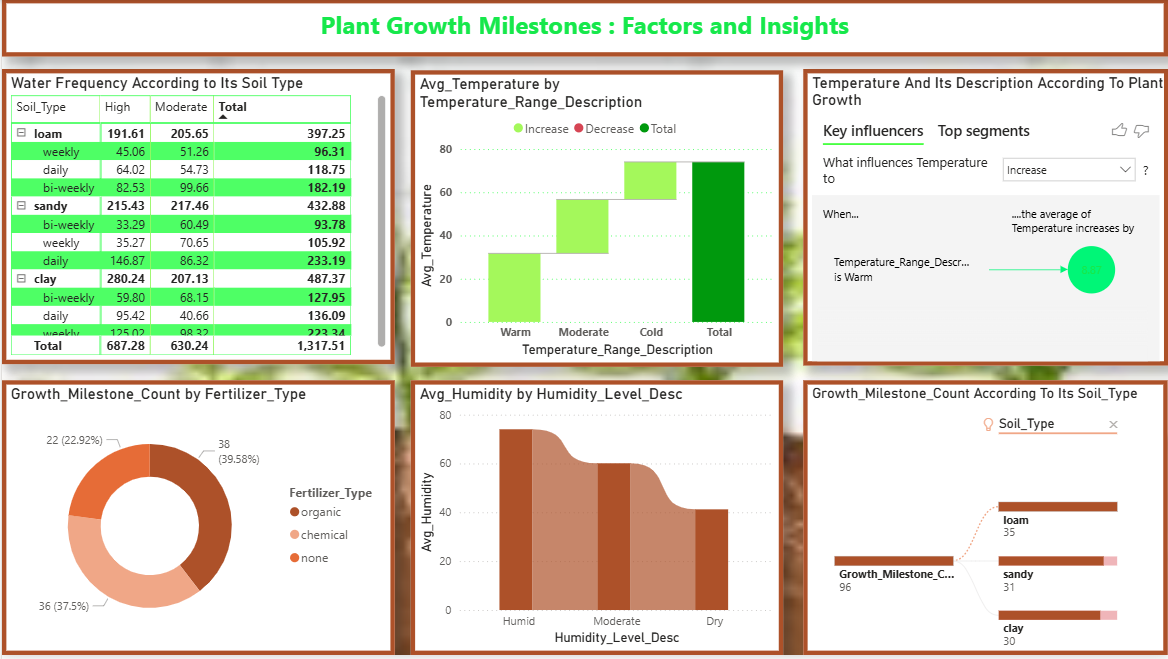
# 10. Appendix

## GitHub & Project Demo Link

GitHub: <https://github.com/ARULCHANDRU/SmartBridge_Data_Analytics_With_Power_BI>  
Demo: https://yourprojectdemo.com

## Visualization Screenshots

Below is the visual output created in Power BI for plant growth prediction based on environmental and management data.



## Dashboard Screenshot

This is the Power BI dashboard designed to summarize key metrics and insights visually.

