Project Title:
Predicting Plant Growth Stages with Environmental and Management Data Using Power BI

Internship / Program:
SmartBridge Data Analytics Virtual Internship

Submitted By:
Darshika Grover
(Email Id - darshikagrover2005@gmail.com**)**

Date of Submission:
30th July 2025

1. Introduction

1.1 Project Overview

This project, "Predicting Plant Growth Stages with Environmental and Management Data Using Power BI", focuses on analyzing how environmental and management factors influence plant growth.
Key parameters like sunlight hours, humidity, temperature, soil type, water frequency, and fertilizer type are considered to classify plant growth milestones.
Using Power BI, the project transforms raw data into interactive dashboards and visual reports, helping users easily interpret complex environmental data and identify growth patterns.

1.2 Objectives

The primary objectives of this project are:

1. To analyze environmental and management data affecting plant growth.
2. To classify plant growth stages using visual and interactive dashboards.
3. To provide clear insights for improving plant health and yield.
4. To simplify raw data analysis for non-technical users.

2. Project Initialization and Planning Phase

2.1 Define Problem Statement

Predicting Plant Growth Stages with Environmental and Management Data Using Power BI

Environmental and management data such as temperature, humidity, water frequency, and sunlight play a critical role in plant growth. However, when this data is unstructured, it becomes difficult for students, researchers, or agricultural planners to analyze it effectively.

The main challenge is to understand how these variables interact and influence different growth stages of plants. Without visual tools or a structured approach, interpreting the data can be time-consuming, inconsistent, and prone to error.

This project addresses the problem by using Power BI to transform raw environmental and management data into clear, interactive dashboards. These visuals help predict and identify growth stages more intuitively, making insights accessible even to users without a technical background.

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	a **student** exploring plant growth patterns	identify how temperature, humidity, and water frequency affect plant growth classification	raw data is hard to interpret and lacks clear visual patterns	there are too many parameters to analyze manually	confused about which factors affect growth most
PS-2	a **student** learning data visualization using Power BI	create an interactive dashboard that helps others understand plant growth under different environmental conditions	traditional charts (like Excel) don't offer interactivity or clear insight	I need tools to model, slice, and filter complex data for better understanding	limited in my ability to draw accurate or useful conclusions

2.2 Project Proposal (Proposed Solution)

This project proposal outlines a solution to address a specific problem. With a clear objective, defined scope, and a concise problem statement, the proposed solution details the approach, key features, and resource requirements, including hardware, software, and personnel.

Project Overview	
Objective	To build an interactive Power BI dashboard that classifies plant growth stages based on environmental factors such as temperature, humidity, and water frequency using structured data visualizations.
Scope	This project covers data import, transformation, and dashboard creation using Power BI. It includes cleaning the dataset, generating calculated fields, and designing visuals that help users interpret plant growth patterns effectively. The solution is limited to data analytics and visualization — it does not cover predictive modelling.
Problem Statement	
Description	Raw plant growth data often contains multiple environmental variables, making it difficult to understand how each factor affects growth. Without visual aid, it's hard to extract meaningful insights from the data.

Impact	Creating a Power BI dashboard simplifies the interpretation of such data. It supports learners and analysts in quickly identifying how different growing conditions influence plant development, improving decision-making in agricultural research or learning environments.
Proposed Solution	
Approach	● Import the dataset into Power BI ● Clean and transform the data using Power Query ● Create calculated columns (e.g., temperature ranges, growth stages) ● Design bar charts, line graphs, and slicers ● Create a dashboard summarizing key insights visually
Key Features	● Interactive dashboard with filters and slicers ● Categorization of growth stages by temperature and water input ● Easy navigation and summary card display ● Visual comparison of growth under different humidity levels

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	Standard laptop/PC with i5 or equivalent CPU

Memory	RAM specifications	8 GB RAM
Storage	Disk space for data, models, and logs	1 GB minimum for dataset and Power BI file
Software		
Frameworks	Power BI frameworks	Power Query for ETL, DAX for logic
Libraries	Additional libraries	Built-in Power BI DAX functions
Development Environment	IDE, version control	Power BI Desktop
Data		
Data	Source, size, format	CSV format Kaggle dataset

2.3 Initial Project Planning

Use the below template to create a product backlog and sprint schedule

Sprint	Functional Requirement (Epic)	User Sto	User Story / Task	Story Points	Priority	Team Members	Sprint Start Date	Sprint End Date
Sprint 1	Data Loading and Preparation	USN-1	As a user, I want to load the plant growth dataset into Power BI and clean missing or invalid data.	2	High	Darshika Grover	19/07/2025	20/07/2025
Sprint 1	Data Transformation	USN-2	As a user, I want to create calculated columns for categorizing temperature ranges and growth stages.	2	High	Darshika Grover	20/07/2025	21/07/2025
Sprint 2	Visualization - Temperature Impact	USN-3	As a user, I want to build a graph to analyze average temperature based on growth classification.	2	High	Darshika Grover	22/07/2025	23/07/2025

Sprint 2	Visualization - Environmental Factors	USN-4	As a user, I want to visualize the impact of water frequency and humidity on plant growth.	3	Medium	Darshika Grover	24/07/2025	25/07/2025
Sprint 3	Dashboard Design	USN-5	As a user, I want to create an interactive dashboard that displays key metrics with slicers and filters.	3	High	Darshika Grover	26/07/2025	27/07/2025
Sprint 3	Insight Generation and Summary	USN-6	As a user, I want to summarize findings from the dashboard in a textbox or insights card.	1	Medium	Darshika Grover	27/07/2025	27/07/2025
Sprint 3	Review and Export	USN-7		1	High	Darshika Grover	28/07/2025	30/07/2025

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan and Raw Data Sources Identified

Data Collection Plan Template

Section	Description
Project Overview	This project aims to visualize and predict plant growth stages using Power BI. The dataset includes environmental and management factors like temperature, humidity, sunlight, and water frequency. Interactive dashboards will be used to analyze how these variables influence plant development.
Data Collection Plan	The data was sourced from a public dataset available on Kaggle. It includes structured entries for key growth indicators and environmental conditions. This data supports the development of dashboards that reveal patterns across various growth stages.
Raw Data Sources Identified	Kaggle-hosted dataset containing ~10,000 rows on plant growth and environmental conditions. The dataset includes numerical and categorical variables used for classification.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Plant Growth Dataset	Contains temperature, humidity, water frequency, sunlight, and growth level classification for plants	https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification	CSV	~10MB	Public (Free Access)

3.2 Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing values in humidity and water frequency columns	Moderate	Rows with missing values were removed using Power Query in Power BI
Kaggle Dataset	Duplicate records found for plant entries	High	Duplicates were removed using the “Remove Duplicates” feature in Power BI
Kaggle Dataset	Inconsistent data types in the temperature column	Moderate	Converted temperature column to decimal format; removed non-numeric entries
Kaggle Dataset	Categorical labels for growth level are not standardized	Low	Categorical values were cleaned using calculated columns in DAX

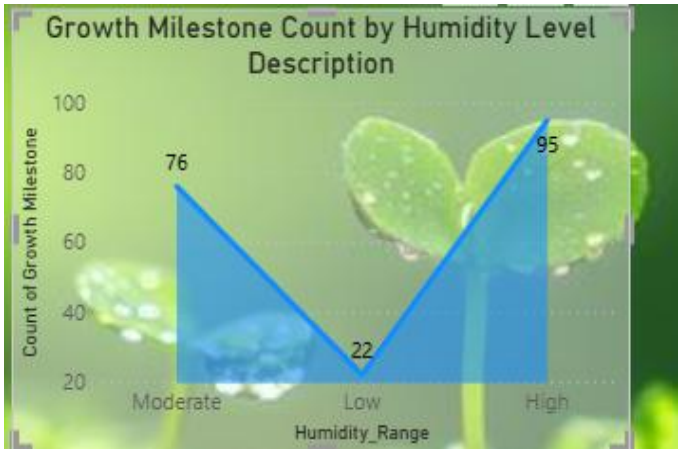
3.3 Data Exploration and Preprocessing

Section	Description
Data Overview	The dataset contains plant growth records along with environmental parameters like temperature, humidity, and water frequency. The goal is to analyze and classify plant growth stages using visual tools.
Data Cleaning	Missing values were removed using filters in Power Query. Duplicates were removed based on plant ID and timestamp.
Data Transformation	Power Query was used to filter data, sort by temperature, and create calculated columns for categorized temperature and growth stages.
Data Type Conversion	Data types for temperature, humidity, and frequency columns were converted to decimal. Date columns were set to
Column Splitting and Merging	No splitting or merging was needed as the dataset did not contain any DateTime or composite columns.
Data Modelling	A single-table model was used. Measures were created using DAX for average temperature, humidity, and growth counts.
Save Processed Data	The cleaned dataset was saved within Power BI as a PBIX file and also exported as Excel for backup.

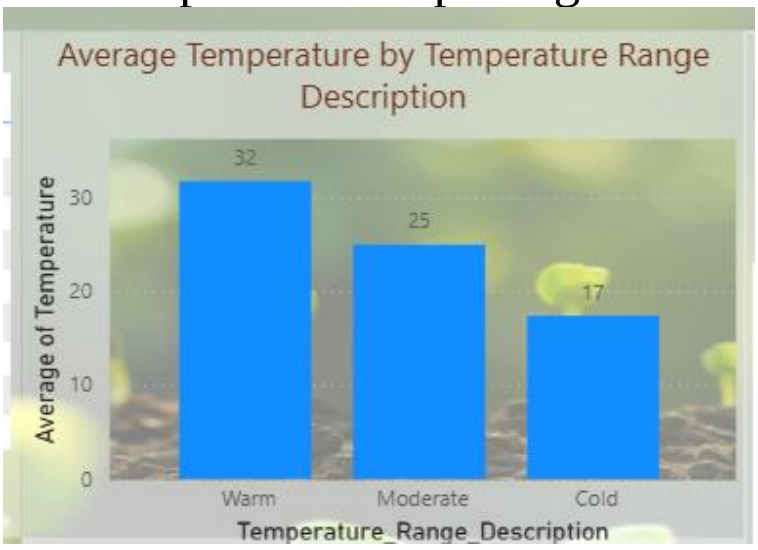
4. Data Visualization

4.1 Framing Business Questions

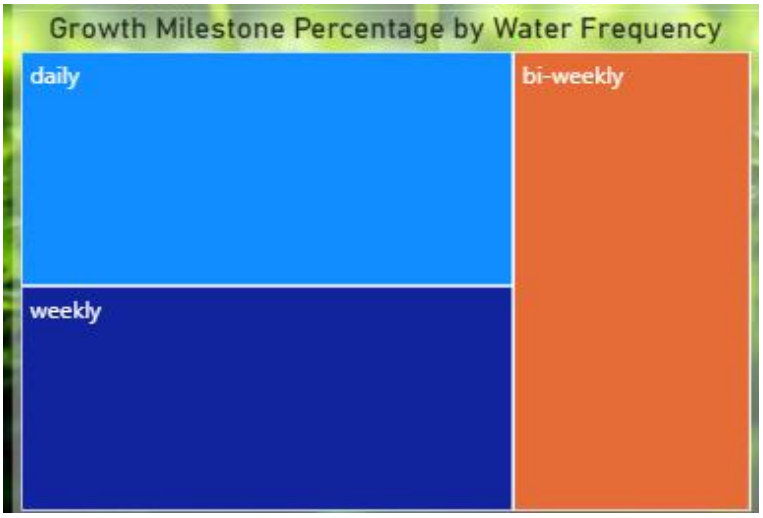
1. **What is the distribution of plant growth stages?**
    - *Visualization*: Bar chart showing the count of each plant growth stage.



2. **How does temperature affect plant growth levels?**
    - *Visualization*: Clustered column chart or key influencer chart showing the impact of temperature on plant growth stages.

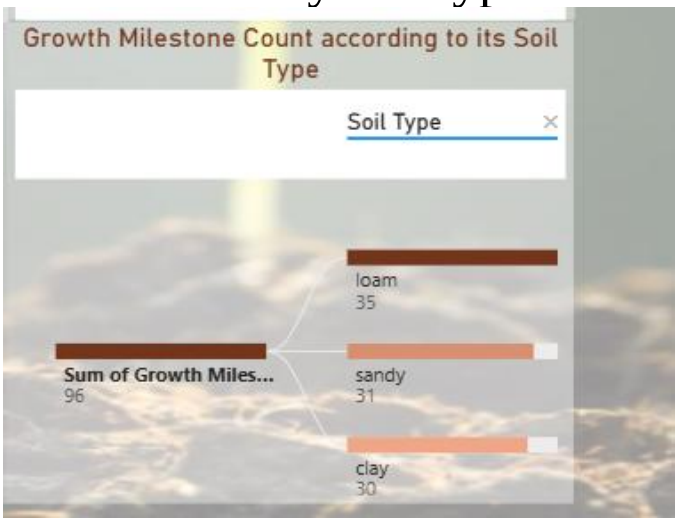


3. **How does water frequency influence plant growth milestones?**
    - *Visualization*: Treemap showing the percentage of growth milestones achieved for each water frequency.



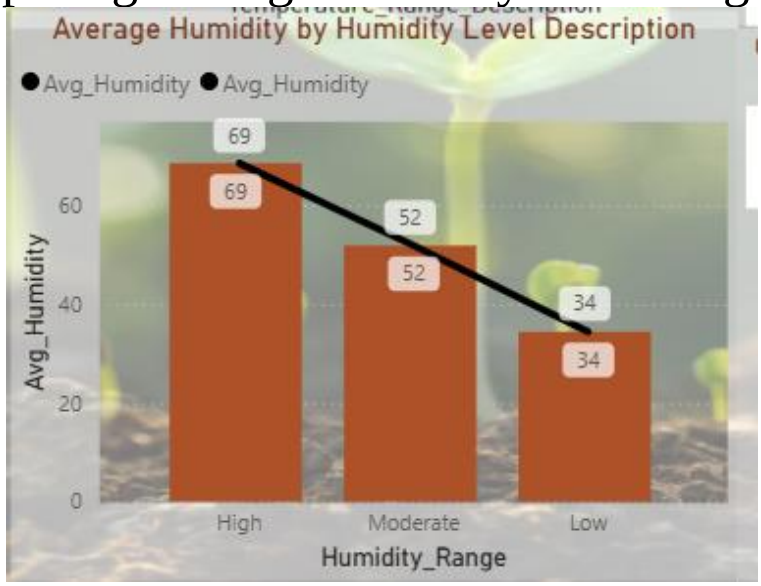
4. How does soil type affect plant growth milestones?

- *Visualization*: Horizontal bar chart showing growth milestone counts by soil type.



5. What is the average humidity across different growth stages?

- *Visualization*: Line and clustered column chart comparing average humidity for each growth stage.



6. What is the average sunlight received by plants?

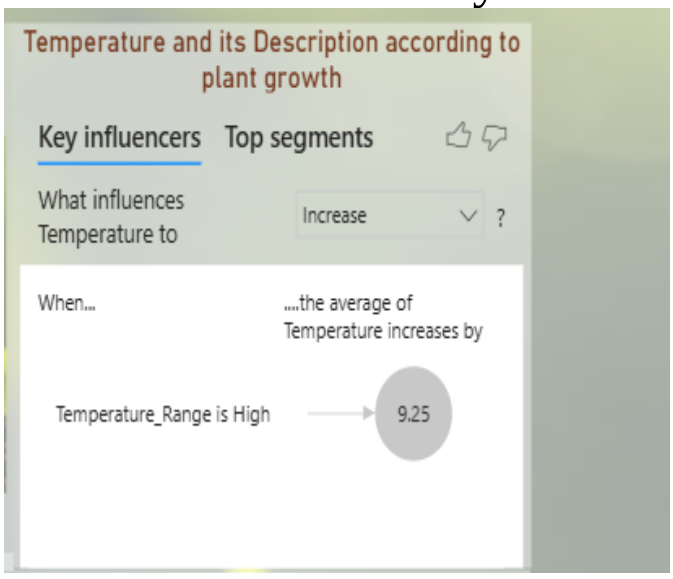
- *Visualization*: Card visual showing the average sunlight hours.

7. What is the total growth milestone count?

- *Visualization*: Gauge visual showing total growth milestone count.

8. What is the combined impact of environmental factors on plant growth?

- *Visualization*: Multi-row card or summary dashboard showing average temperature, humidity, and sunlight with slicers for interactivity.



4.2 Developing Visualizations

Visualizations were created in Power BI to convert raw environmental and management data into meaningful insights.
Key visuals included:

- Bar Charts: Growth milestones by soil type and humidity levels
- Line Chart: Average humidity across growth stages
- Treemap: Growth milestone percentage by water frequency
- KPI Cards: Average temperature, humidity, and sunlight
- Gauge Chart: Total growth milestone count
- Key Influencer Chart: Factors affecting plant growth

All visualizations were combined to form one interactive dashboard and one report page.

5. Dashboard

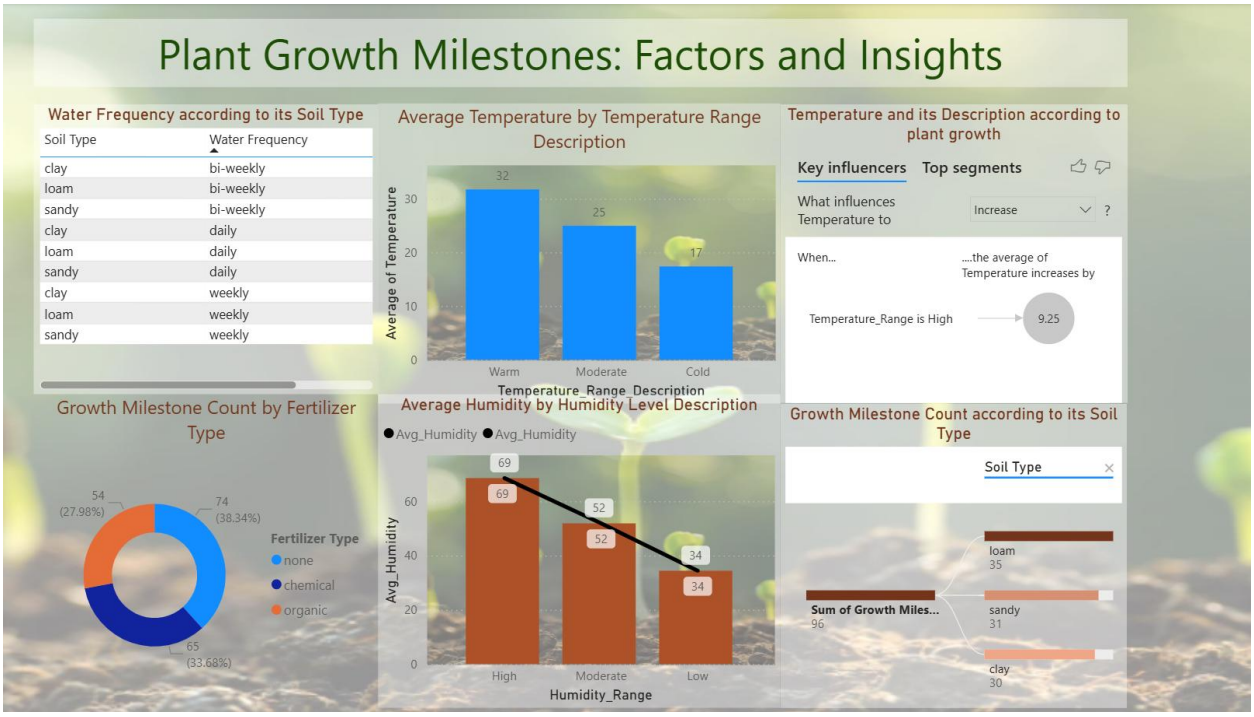
5.1 Dashboard Design File

This dashboard provides a visual analysis of environmental and management factors influencing plant growth stages. It allows users to quickly identify trends and patterns using interactive visuals and KPIs. The dashboard simplifies raw environmental data, helping in decision-making and improving understanding of conditions that lead to optimal plant growth.

Activity 1: Interactive and visually appealing dashboards

Plant Growth Milestones – Factors and Insights

This dashboard displays how soil type, water frequency, temperature, humidity, and fertilizer usage affect plant growth milestones.



The following key insights can be derived from the dashboard:

1. **Water Frequency according to its Soil Type**

- **Insight:** Shows how water frequency varies across soil types. Loamy soil with a moderate watering frequency supports better plant growth.

2. **Average Temperature by Temperature Range Description**

- **Insight:** Warm temperatures (~32°C) are linked with better plant growth, while cold temperatures (~17°C) result in lower growth.

3. **Temperature and its Description according to Plant Growth (Key Influencer)**

- **Insight:** High temperature ranges significantly increase the average growth performance.

4. **Growth Milestone Count by Fertilizer Type**

- **Insight:** Chemical fertilizers result in the highest growth milestone count, followed by organic fertilizers.

5. **Average Humidity by Humidity Level Description**

- **Insight:** High humidity (~69%) is most favourable for plant growth, while low humidity (~34%) leads to lower growth stages.

6. **Growth Milestone Count according to its Soil Type**

- **Insight:** Loamy soil supports the highest growth milestones (35), followed by sandy and clay soils.

6. Report

6.1 Story Design File

This report presents a visual analysis of environmental and management factors affecting plant growth stages. Using Power BI, raw agricultural data such as **temperature, humidity, sunlight hours, water frequency, and soil type** has been transformed into **interactive dashboards**.

These dashboards allow users to **quickly understand trends**, identify **optimal growth conditions**, and support **better decision-making for plant health and yield improvement**.

Report Description

The dashboard consists of **key KPIs, charts, and summaries** to provide both high-level insights and detailed breakdowns:

1. **KPI Cards:**
    - Average Sunlight Hours: **6.83 hours**
    - Average Humidity: **58.10%**
    - Average Temperature: **25.08°C**
2. **Visualizations:**
    - **Pie Chart:** Average Sunlight Hours by Soil Type
    - **Treemap:** Growth Milestone Percentage by Water Frequency
    - **Column & Line Chart:** Growth Milestone Count by Humidity Level Description
    - **Gauge Chart:** Total Growth Milestone Count
    - **Text Report Section:** Highlights the highest performing soil types and water frequencies.

Key Insights

1. **Clay soil** showed the highest average sunlight hours (**7.27 hours**) followed by **sandy** and **loamy soils**.

2. **Moderate humidity (58.10%)** supported the highest growth milestones.
3. **Moderate growth milestone count** was **495.45% higher than dry conditions**, emphasizing the need for adequate humidity.
4. **Daily water frequency** led to the highest growth milestone percentage, followed by **bi-weekly** and **weekly** schedules.
5. The **total growth milestone count was 96**, represented by the **gauge visual**.

7. Performance Testing

Performance testing ensures that the Power BI dashboard and report are optimized, interactive, and able to handle data exploration smoothly without lag.

7.1 Utilization of Data Filters

- Filters/Slicers Used:
    a. Soil Type
    b. Water Frequency
    c. Humidity Level Description

These filters enable interactive exploration, allowing users to focus on specific environmental conditions and see how they affect plant growth milestones.

7.2 Number of Calculated Fields

- Calculated Columns & Measures Created in DAX: 10
    - Key Measures:
        - Growth Milestone Count
        - Growth Milestone Percentage
        - Average Temperature
        - Average Humidity
        - Average Sunlight Hours

These measures and calculated fields power the KPIs and insights in the dashboard and report.

7.3 Number of Visualizations / Graphs

Total Visuals Created: 11

1. Water Frequency according to Soil Type (Table/Matrix)
2. Average Temperature by Temperature Range Description (Bar Chart)
3. Temperature and Its Description according to Plant Growth (Key Influencer Chart)
4. Growth Milestone Count by Fertilizer Type (Doughnut Chart)
5. Average Humidity by Humidity Level Description (Line & Column Chart)
6. Growth Milestone Count according to Soil Type (Horizontal Bar Chart)
7. Average Sunlight Hours by Soil Type (Pie Chart)
8. Growth Milestone Percentage by Water Frequency (Treemap)
9. Growth Milestone Count by Humidity Level Description (Line & Column Chart)
10. Growth Milestone Count (Gauge Visual)

11. KPI Cards – Average Humidity, Average Temperature, Average Sunlight Hours

8. Conclusion / Observation

The Plant Growth Stages Analysis Project successfully transformed raw environmental and management data into interactive Power BI dashboards and reports.
Key observations from the analysis:

1. Soil Type Impact: Loamy soil supports the highest growth milestones, followed by sandy and clay soils.
2. Temperature Influence: Moderate temperatures (~25–32°C) are most favourable for healthy plant growth.
3. Humidity Role: High to moderate humidity (~58–69%) results in higher growth milestone counts.
4. Water Frequency: Daily or moderate watering schedules lead to maximum growth milestones.
5. Fertilizer Impact: Chemical fertilizers contribute to the highest growth milestones.
6. KPI Insights: The total growth milestone count is 96, and average sunlight is ~7 hours/day.

Conclusion:

- The dashboard and report make plant growth analysis simpler and more accessible, even for non-technical users.

- Environmental factors such as soil type, water frequency, and humidity play a crucial role in achieving optimal plant growth.
- This project demonstrates the power of Power BI in converting raw agricultural data into actionable insights for improved decision-making and yield optimization.

9. Future Scope

The project can be extended and enhanced in the following ways:

1. Integration with Real-Time IoT Sensors:
    a. Collect live data for temperature, humidity, and soil moisture to automatically update the Power BI dashboard.
2. Predictive Analytics using Machine Learning:
    a. Develop models to predict plant growth stages and yield outcomes based on environmental conditions.
3. Expansion to Multiple Crops:
    a. Include datasets for various crop types to provide comparative insights.
4. Automated Alerts for Plant Health:
    a. Generate notifications when any environmental factor reaches critical levels that could affect growth.
5. Mobile-Friendly Dashboards:
    a. Publish dashboards in a mobile-optimized format for easy access by farmers and researchers in the field.

10.Appendix

10.1 GitHub & Project Demo Link

- GitHub Repository Link: *https://github.com/dar2005/Plant-Growth-Power-BI-Project*
- Project Demo Link (Optional): *https://drive.google.com/file/d/1h6tqxh84kAmm78nZbPIssSBAaU2dvGVz/view?usp=sharing*