Final Project Report

Project Title: Predicting plant growth stages with environmental and management data using power bi

Team Member: Bathula Manish Reddy

1. Introduction

1.1 Project Overview

This project focuses on analyzing various environmental and agricultural factors that influence plant growth milestones. By integrating data such as soil type, sunlight hours, humidity levels, temperature, water frequency, and fertilizer types, we aim to uncover patterns and insights that can support better farming decisions. A visually interactive dashboard has been developed to allow users—such as farmers, agronomists, and researchers—to monitor and compare the impact of different growing conditions on plant development.

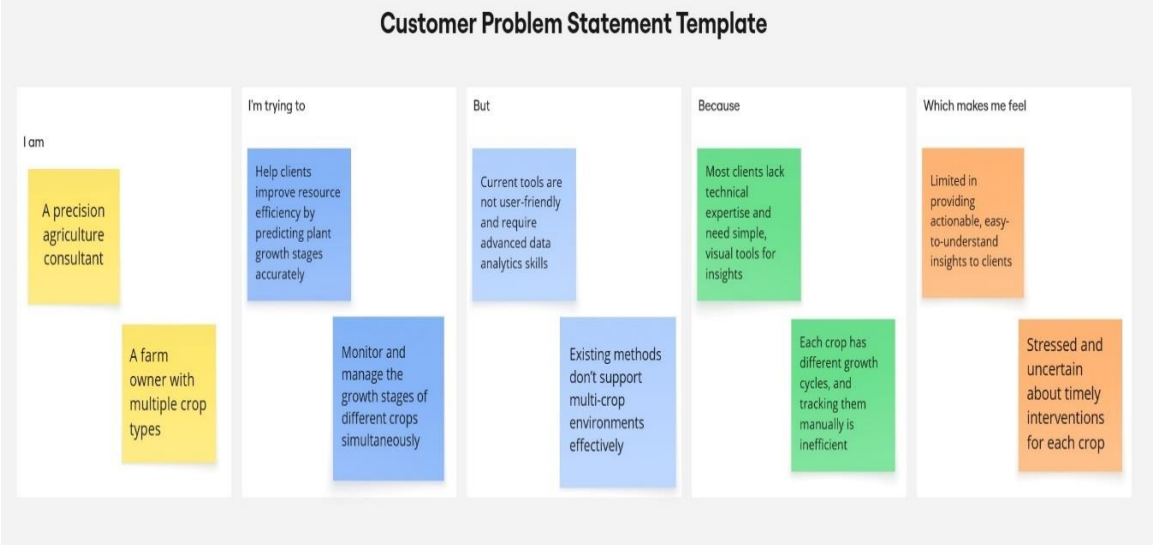
1.2 Objectives

- To collect and preprocess plant growth data for meaningful analysis.
- To visualize relationships between environmental factors and growth outcomes.
- To identify which factors (e.g., humidity, soil, water schedule) contribute most to plant growth.
- To build an interactive dashboard that enables easy interpretation of key metrics and trends.
- To support data-driven decision-making in agriculture for improved crop performance.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

Problem Statement (PS)	I am (Farmer)	I'm trying to	But	Because	Which makes me feel
PS-1	A precision agriculture consultant	Help clients improve resource efficiency by predicting plant growth stages accurately	Current tools are not user-friendly and require advanced data analytics skills	Most clients lack technical expertise and need simple, visual tools for insights	Limited in providing actionable, easy-to-understand insights to clients
PS-2	A farm owner with multiple crop types	Monitor and manage the growth stages of different crops simultaneously	Existing methods don't support multi-crop environments effectively	Each crop has different growth cycles, and tracking them manually is inefficient	Stressed and uncertain about timely interventions for each crop



2.2. Project Proposal (Proposed Solution)

The project will follow a modular, sprint-based development approach starting with data collection and preprocessing, followed by analytical modelling, visualization, and evaluation. The process includes:

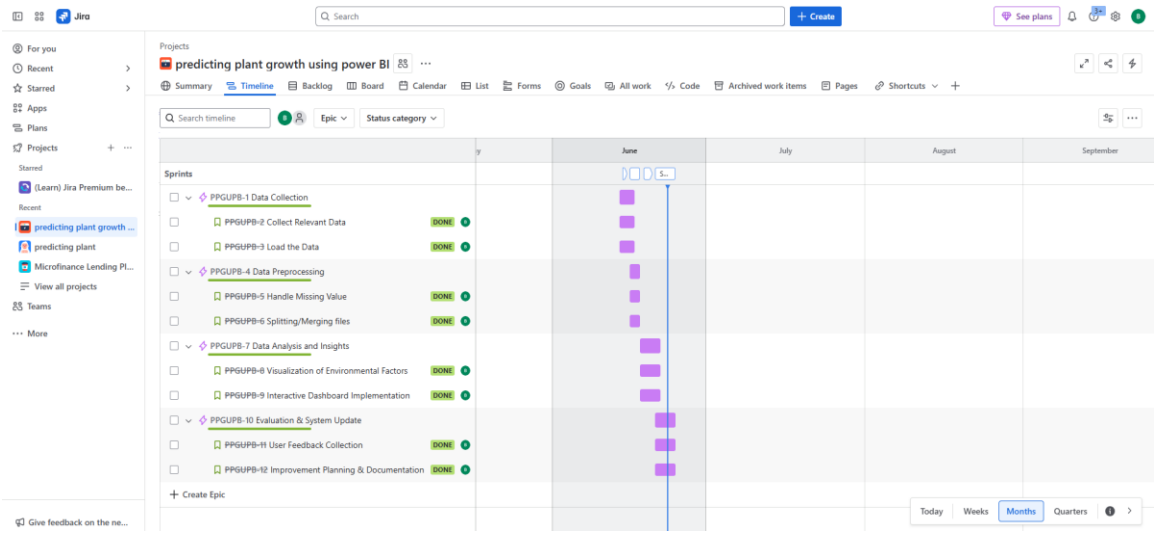
- **Data Collection & Preprocessing**: Gather data related to environmental variables and plant growth milestones from available datasets.
- **Data Analysis**: Use statistical and comparative analysis to identify patterns, correlations, and growth influencers.
- **Dashboard Development**: Create interactive dashboards using Power BI to present visual insights.
- **Feature Implementation**: Add comparison tools, alert systems, and export functionalities.
- **Evaluation & Feedback**: Conduct performance evaluation and update the system based on user feedback.

Key Features

- ✓ Interactive Dashboards
- ✓ Growth Influencer Analysis
- ✓ Environmental Trend Tracking
- ✓ Condition-Based Recommendations

2.3. Initial Project Planning

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Member	Sprint Start Date	Sprint End Date (Planned)
Sprint-1	Data Collection	USN-1	I can collect and organize soil, water, temperature, and humidity data for further processing.	3	Medium	Manish	14 June,2025	15 June,2025
	Collect Relevant Data	USN-2	I can collect soil type, humidity, temperature, and water frequency data from various sources to ensure complete input for analysis.	2	Medium	Manish	14 June,2025	15 June,2025
	Load the Data	USN-3	I can load the collected data into a central storage or database for further processing.	1	Medium	Manish	14 June,2025	15 June,2025
Sprint-2	Data Preprocessing	USN-4	This stage focuses on cleaning and organizing the raw data to make it analysis-ready.	4	High	Manish	16 June,2025	17 June,2025
	Handle Missing Values	USN-5	I can identify and fill or remove missing values to improve data quality and consistency.	2	High	Manish	16 June,2025	17 June,2025
	Splitting/Merging Fields	USN-6	I can split combined columns (e.g., Date Time into Date and Time) or merge fields where necessary to structure the dataset properly.	2	High	Manish	16 June,2025	17 June,2025
Sprint-3	Data Analysis and Insights	USN-7	As a data analyst, I can analyse growth milestones by environmental factors to identify patterns in plant development.	8	High	Manish	18 June,2025	20 June,2025
	Visualization of Environmental Factors	USN-8	As a user, I can view graphs showing average sunlight, temperature, and humidity levels to understand growing conditions.	4	Medium	Manish	18 June,2025	20 June,2025
	Interactive Dashboard Implementation	USN-9	As a farmer, I can use a dashboard to compare plant growth across soil types and water frequencies for better planning.	4	High	Manish	18 June,2025	20 June,2025
Sprint-4	Evaluation & System Update	USN-10	As a stakeholder, I can evaluate the system's performance and suggest improvements for the next farming cycle.	8	Medium	Manish	21 June,2025	23 June,2025
	User Feedback Collection	USN-11	Gathering input from farmers, analysts, and stakeholders.	5	Medium	Manish	21 June,2025	23 June,2025
	Improvement Planning & Documentation	USN-12	Logging findings and planning improvements for future cycles.	3	Medium	Manish	21 June,2025	23 June,2025



3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Section	Description
Project Overview	This project aims to analyse the impact of environmental factors—such as soil type, humidity, temperature, water frequency, and fertilizer—on plant growth. The objective is to build insights and dashboards that help farmers make data-driven decisions to improve crop productivity.
Data Collection Plan	Data is collected through a combination of field observations, environmental sensors, and manual inputs recorded during a controlled agricultural experiment. Key variables like

	water frequency, and growth milestones were recorded for each planting cycle.
Raw Data Sources Identified	- **Soil Type Records:** Collected manually from field logs describing soil composition (e.g., loam, clay, sandy). - **Weather Data:** Temperature and humidity captured using environmental sensors or weather APIs. - **Watering Schedule:** Logged manually during the plant growth phase. - **Fertilizer Type:** Input recorded during planting based on applied treatment. - **Growth Milestone Count:** Manually observed and noted based on plant development stages.

3.1.1 Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Plant Growth Data Classification, Real-Dataset (Kaggle)	Contains plant growth data with attributes such as soil type, temperature, humidity, water frequency, fertilizer type, and growth milestones.	https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification/data	CSV	12 KB	Public

3.2. Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset	Mention the issues faced in the selected dataset.	Low/ Moderate/ High	Give the solution for that issue technically.
Plant Growth Data Classification, Real-Dataset (Kaggle)	Inconsistent text formatting in categorical columns **(Soil_Type, Water_Frequency, Fertilizer_Type)**	Moderate	Standardize text values using .str.lower().str.strip() in pandas to remove case and whitespace issues.
Plant Growth Data Classification, Real-Dataset (Kaggle)	No date or timestamp Column available	Moderate	Add a Date column during data collection to allow time-based analysis and trend visualization.
Plant Growth Data Classification, Real-Dataset (Kaggle)	**Growth_Milestone** Column contains binary values only (e.g., 0 and 1)	Low	Refine milestone tracking method to allow a range of values for better growth trend analysis.
Plant Growth Data Classification, Real-Dataset (Kaggle)	Missing engineered features such as environmental impact scores	Low	Create new derived columns like **Env_Index = (Humidity * Temp) / 100** for richer analysis.

Plant Growth Data Classification, Real-Dataset (Kaggle)	Categorical columns not ready for machine learning model input	Low	Apply Label Encoding or One-Hot Encoding to categorical fields during preprocessing.

3.3. Data Exploration and Preprocessing

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.

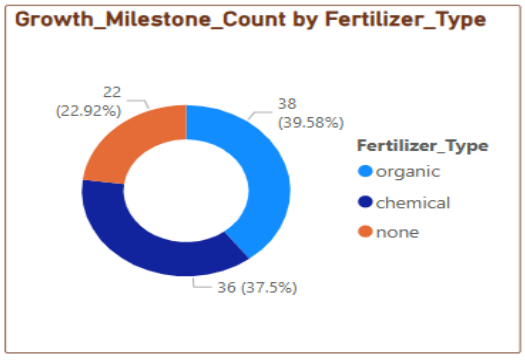
4. Data Visualization

4.1. Framing Business Questions

I. Which fertilizer type resulted in the most growth milestones?

a. *Visualization*: Doughnut Chart of Growth Milestone by Fertilizer Type

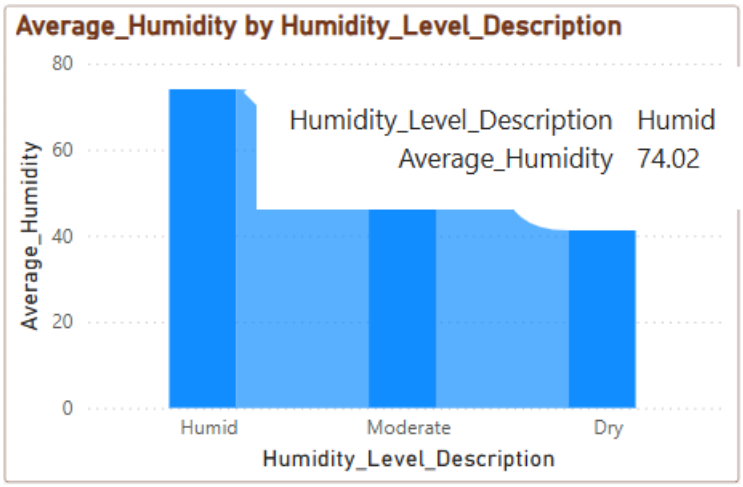
b. *Screenshot of visualization*



II. Which humidity level had the highest average humidity?

a. *Visualization*: Bar Chart of Average Humidity by Humidity Level

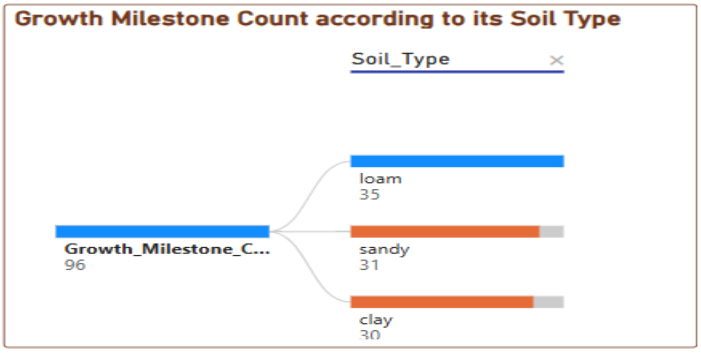
b. *Screenshot of visualisation*



III. Which soil type led to the highest growth milestones?

a. *Visualization*: Decomposition tree of Growth Milestone Count by Soil Type.

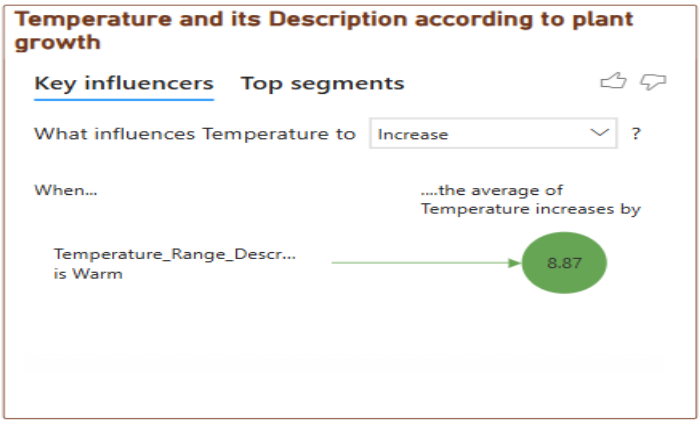
b. *Screenshot of visualisation*



IV. What condition causes temperature to increase most?

a. *Visualization*: Key Influencer for Temperature Increase

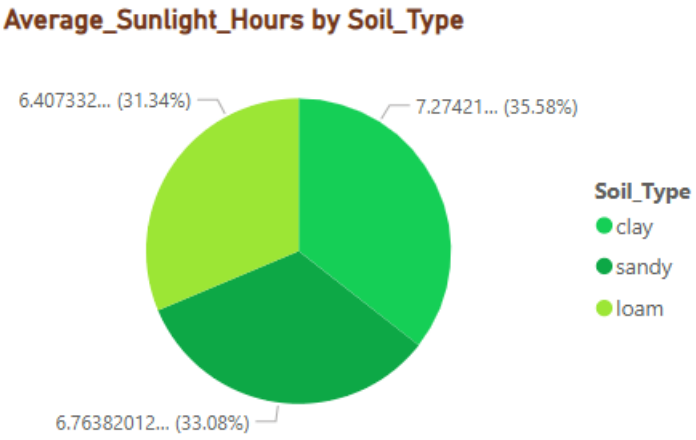
b. *Screenshot of visualisation*



V. Which soil type receives the most sunlight on average?

a. *Visualization*: Pie Chart of Sunlight Hours by Soil Type

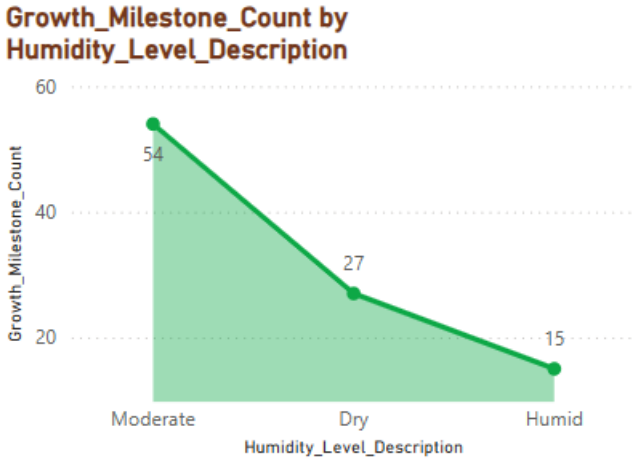
b. *Screenshot of visualisation*



VI. Which soil type receives the most sunlight on average?

a. *Visualization*: Line Chart of Growth Milestone Count by Humidity Level

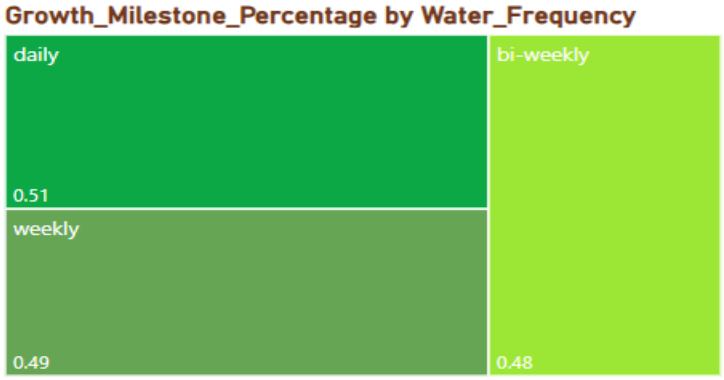
b. *Screenshot of visualisation*



VII. Which water frequency had the highest growth milestone percentage?

a. *Visualization*: Tree map of Growth Milestone % by Water Frequency

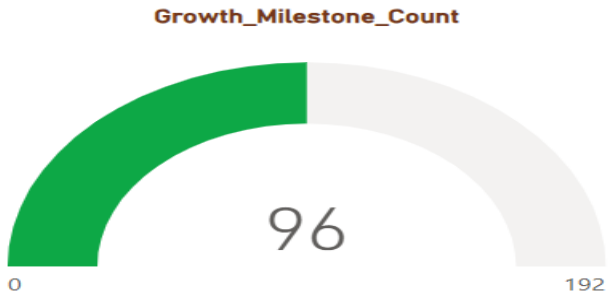
b. *Screenshot of visualisation*



VIII. What is the total growth milestone count?

a. *Visualization*: Gauge Chart of Total Growth Milestone Count

b. *Screenshot of visualisation*



4.2. Developing Visualizations

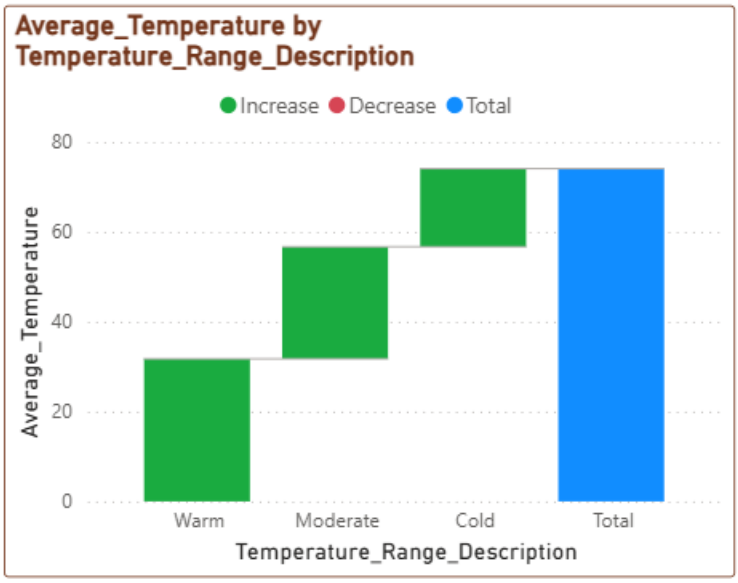
Activity 1.1: Water Frequency According to Its Soil Type

Water Frequency according to its soil Type

Soil_Type	High	Moderate	**Total**
⊟ **loam**	**191.61**	**205.65**	**397.25**
weekly	45.06	51.26	**96.31**
daily	64.02	54.73	**118.75**
bi-weekly	82.53	99.66	**182.19**
⊟ **sandy**	**215.43**	**217.46**	**432.88**
bi-weekly	33.29	60.49	**93.78**
weekly	35.27	70.65	**105.92**
daily	146.87	86.32	**233.19**
⊟ **clay**	**280.24**	**207.13**	**487.37**
bi-weekly	59.80	68.15	**127.95**
daily	95.42	40.66	**136.09**
weekly	125.02	98.32	**223.34**
Total	**687.28**	**630.24**	**1,317.51**

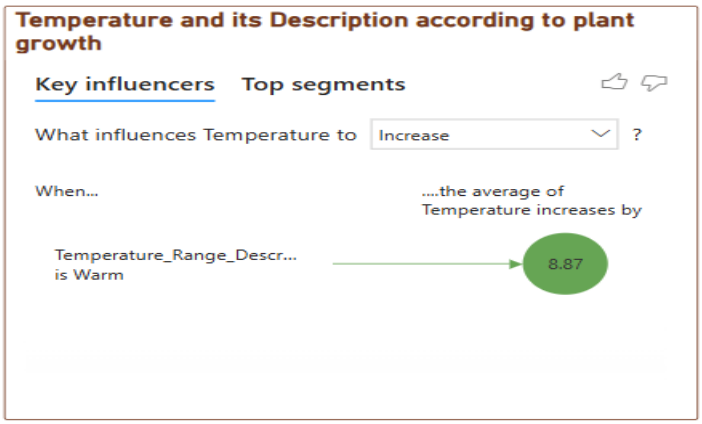
Insight: Loam soil with high water frequency results in the highest total water frequency, indicating that loam soil may require or benefit from more frequent watering compared to sandy and clay soils.

Activity 1.2: Average Temperature by Temperature Range Description



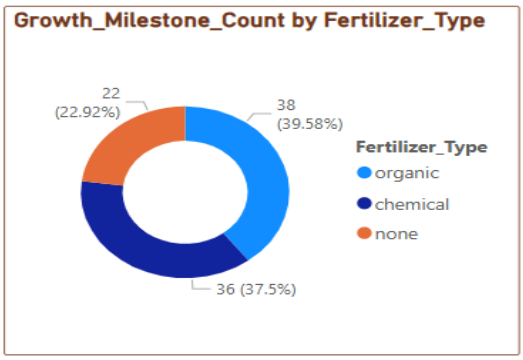
Insight: Moderate temperature ranges have the highest average temperature increase, suggesting that plants in this range experience the most growth.

Activity 1.3: Temperature and Its Description According to Plant Growth



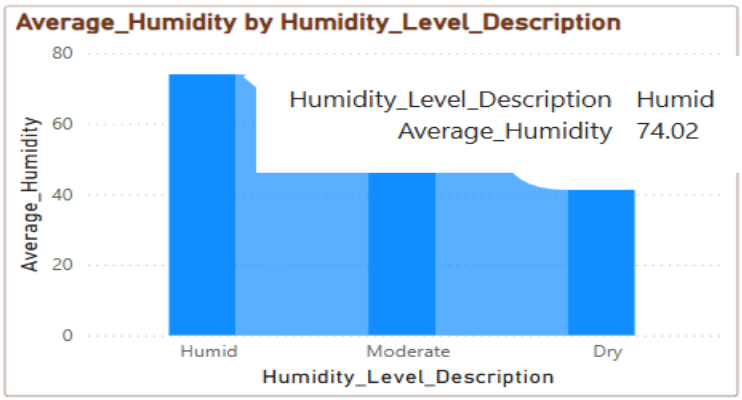
Insight: Cold temperature range decreases plant growth, highlighting the negative impact of cold temperatures on plant development.

Activity 1.4: Growth Milestone Count by Fertilizer Type



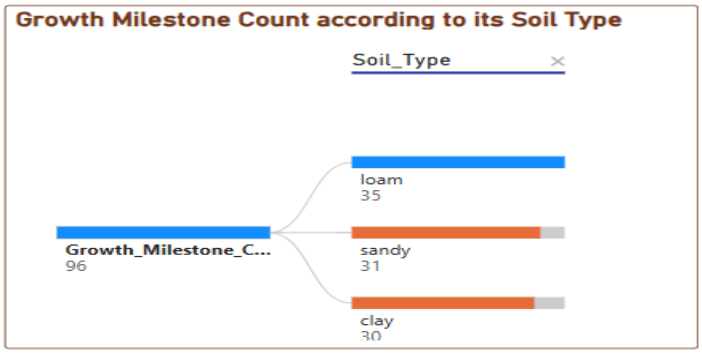
Insight: Chemical fertilizers account for the highest growth milestone count, indicating their effectiveness in promoting plant growth milestones compared to organic fertilizers or no fertilizer.

Activity 1.5: Average Humidity by Humidity Level Description



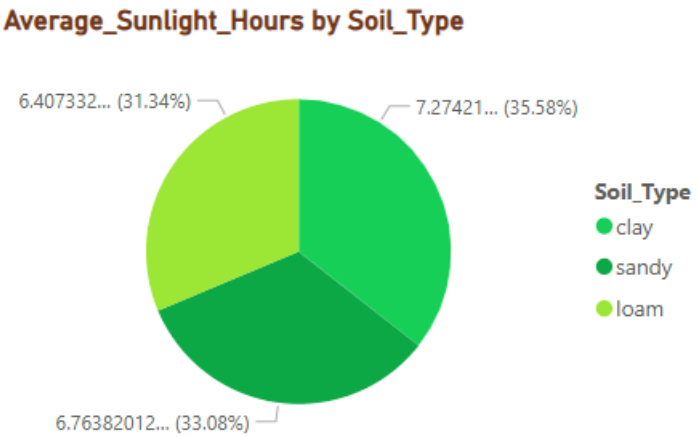
Insights: Humid conditions lead to the highest average humidity, emphasizing the importance of maintaining high humidity levels for optimal plant growth.

Activity 1.6: Growth Milestone Count According to Its Soil Type



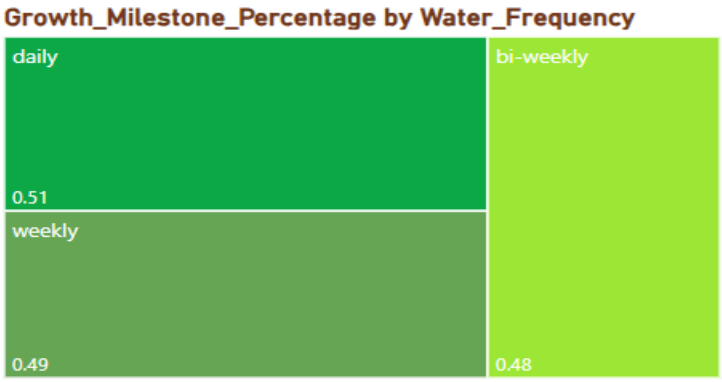
Insights: Loam soil has the highest growth milestone count, suggesting it is the most conducive soil type for achieving plant growth milestones.

Activity 1.7: Average Sunlight Hours by Soil Type



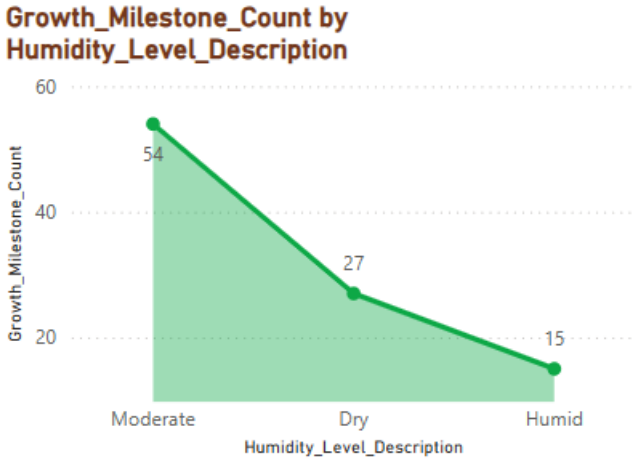
Insights: Sandy soil receives the highest average sunlight hours, which may contribute to its growth performance under sufficient light conditions.

Activity 1.8: Growth Milestone Percentage by Water Frequency



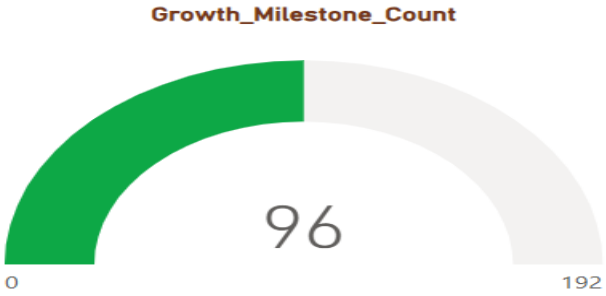
Insights: Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.

Activity 1.9: Growth Milestone Count by Humidity Level Description



Insights: Humid conditions lead to the highest growth milestone count, reinforcing the importance of maintaining adequate humidity levels for optimal growth

Activity 1.10: Growth Milestone Count



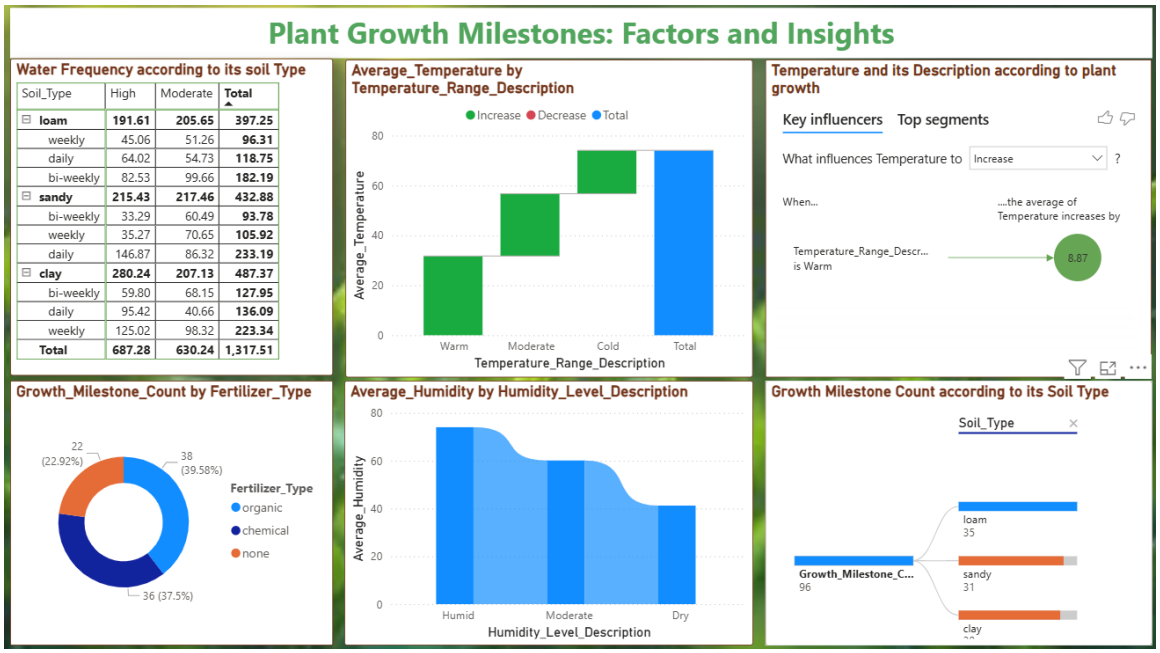
Insights: The total growth milestone count is 96, with the highest contributions coming from environments with optimal humidity and temperature conditions.

Activity 1.11: Average Humidity, Average Temperature, Average Sunlight Hours

Insights: The data indicates a temperate and comfortable climate with an average temperature of 25.08°C, moderate humidity of 58.10%, and a reasonable amount of sunlight averaging 6.83 hours per day.

5. Dashboard

5.1. Dashboard Design File



Note: Highlight the major outcomes in form of bullet points

Sample:

*Here are **six** potential outcomes from the dashboard image provided:*

i. Loam Soil Supports Highest Plant Growth Milestones

The **Growth Milestone Count by Soil Type** visual shows that **loam soil** accounts for the highest number of growth milestones (35), followed by sandy (31) and clay (30). This indicates that loam provides a well-balanced texture and optimal conditions for plant development compared to other soil types.

ii. Warm Temperatures Significantly Boost Plant Growth

From both the **Average Temperature by Temperature Range Description** and the **Key Influencers visual**, it is clear that **warm temperature** conditions result in a significant **increase in plant growth**, with an average temperature increase of **8.87** units. This suggests that managing environmental temperature can be critical for maximizing yield.

iii. Chemical Fertilizers Yield Slightly Better Results

The **Growth Milestone Count by Fertilizer Type** donut chart shows that **chemical fertilizers** contribute to the highest milestone count (38), followed closely by **organic fertilizers** (36). This outcome highlights the effectiveness of fertilization—particularly chemical and organic types—in promoting plant growth.

iv. Clay Soil Requires the Most Frequent Watering

Based on the **Water Frequency according to its Soil Type** table, **clay soil** has the highest total water requirement (487.37), particularly in the **daily and weekly categories**. This implies that while clay supports growth, it demands more water due to its dense structure and slower drainage.

v. Humid Conditions Favor Higher Growth Potential

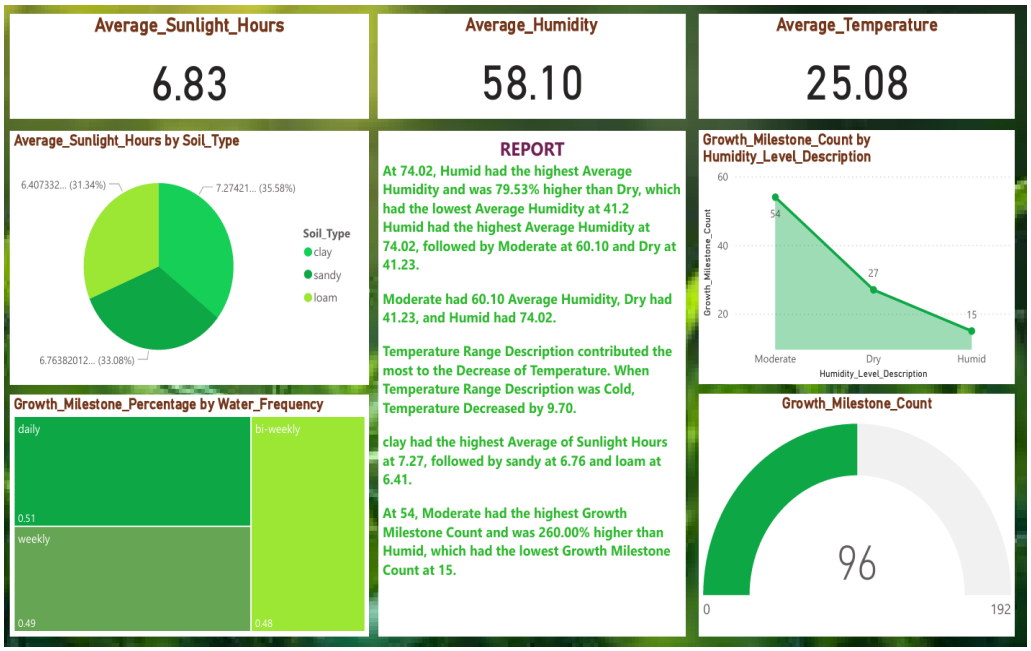
The **Average Humidity by Humidity Level Description** chart shows that **humid conditions** have the highest average humidity (around 75%), which is ideal for plant growth. Plants in humid environments are less prone to water stress, supporting healthier development and more milestones.

vi. Interactive Analysis Enables Deeper Insights

With visuals like the **decomposition tree** and **key influencer analysis**, users can interactively explore how each variable—such as soil type or temperature—affects outcomes. These features enable more personalized insights and real-time decision-making for optimizing plant growth conditions.

6. Report

6.1. Story Design File



Observations drawn from reports in Power BI can provide valuable insights into business performance and trends.

i. Moderate Humidity Drives the Highest Growth Milestones

The Growth Milestone Count by Humidity Level Description chart reveals that:

- **Moderate humidity** conditions lead to the highest number of plant growth milestones (**54**),
- Followed by **Dry** (27), and **Humid** (15).

This suggests that while high humidity may be good for moisture retention, moderate humidity strikes a better balance for optimal plant development.

ii. High Humidity Doesn't Guarantee Higher Growth

Despite Humid conditions having the **highest average humidity** of **74.02%**, they resulted in the **lowest growth milestone count** (15), as noted in the report. In contrast, **Moderate** humidity, with an average of **60.10%**, yielded the most growth.

- This highlights that extremely high humidity may hinder growth, possibly due to over-saturation or fungal risk.

iii. Cold Temperatures Significantly Lower Growth Potential

The report notes that when the Temperature Range Description was Cold, the **temperature decreased by 9.70 units**, which negatively affected plant growth. This reinforces the trend that colder climates are not suitable for optimal growth, emphasizing the need for warm-to-moderate temperatures.

iv. Clay Soil Receives the Most Sunlight

According to the pie chart and report:

- **Clay soil** had the **highest average sunlight hours** (7.27),
- Followed by **sandy** (6.76),
- And **loam** with the least (6.41).

This implies that clay-based plots receive better sun exposure, which could be beneficial if temperature and water conditions are well managed.

v. Balanced Water Frequency Leads to Better Growth

The Growth Milestone Percentage by Water Frequency treemap shows:

- **Daily watering** leads the chart (0.51 or 51%),
- Followed closely by **weekly** (0.49),
- **Bi-weekly** watering has the lowest milestone share (0.48).

This means frequent and consistent watering schedules support healthier growth, while less frequent watering may not provide enough moisture.

vi. Overall Growth Milestone Count Stands at 96

The gauge chart at the bottom-right shows a **total growth milestone count of 96** out of a possible maximum (192). This helps measure overall progress and suggests that there is potential to double the plant growth performance through improved environmental management and input strategies.

vii. Average Metrics Snapshot

From the KPI tiles at the top:

- **Average Sunlight Hours**: 6.83 hours
- **Average Humidity**: 58.10%
- **Average Temperature**: 25.08°C

These average conditions provide a baseline reference for evaluating whether current plant environments are within an optimal range or need adjustment.

7. Performance Testing

7.1 Utilization of Data filters

Description:
Data filters are used to narrow down the dataset based on specific conditions or criteria. In this project, filters were applied on attributes like:

- **Soil Type**
- **Water Frequency**
- **Humidity Level**
- **Temperature Range**

These filters allow users to interact with the dashboard by selecting specific values to see customized visuals. For example, a user can filter for only *clay soil* and *daily water frequency* to analyze growth metrics under those conditions.

7.2 Number of Calculation Fields

- A total of **11 calculation fields** were created in the project.
- These include **measures** (used to calculate averages, counts, and percentages).
- Also include **new columns** (used to categorize or classify data).
- Measures are used to summarize numerical data like average temperature, humidity, and sunlight.
- New columns help in creating labels such as temperature range (Low, Moderate, High) or humidity level (Dry, Humid).

- These calculated fields improve the quality of visuals by making the data more understandable.
- They support advanced filtering, grouping, and comparative analysis in the dashboard.
- All calculations were done using **DAX formulas** in Power BI.

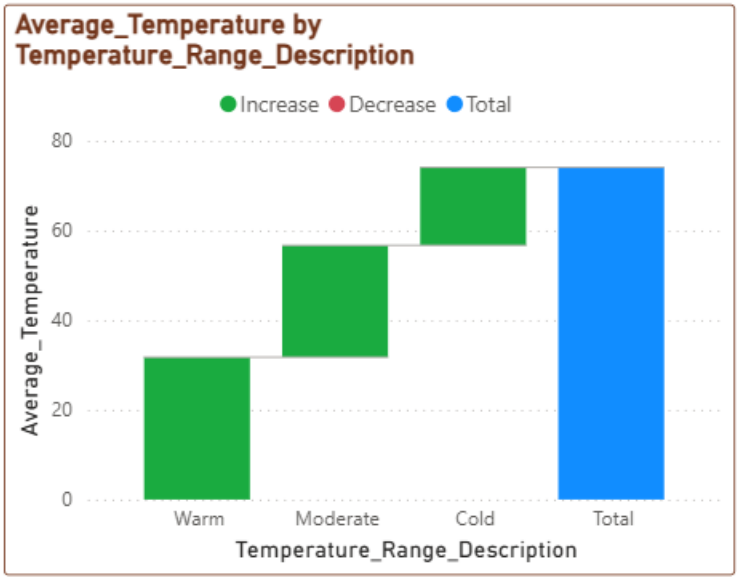
7.3 No of Visualization

I. Water Frequency According to Its Soil Type

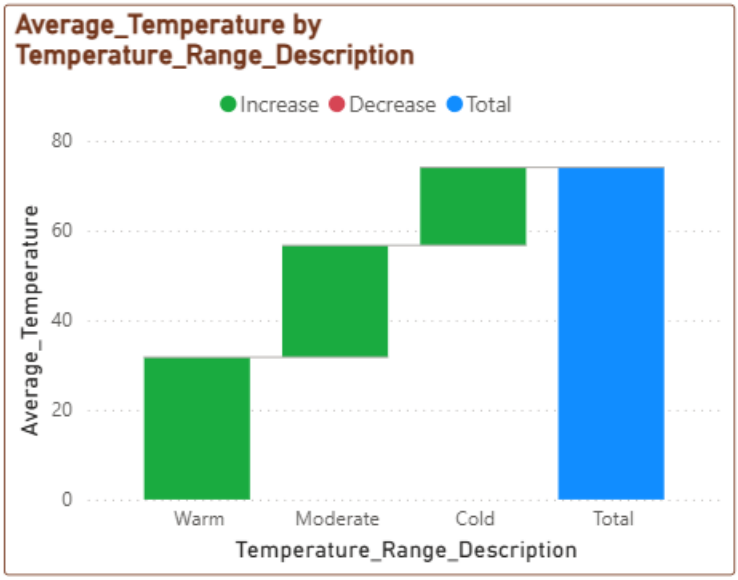
Water Frequency according to its soil Type

Soil_Type	High	Moderate	Total
⊟ **loam**	**191.61**	**205.65**	**397.25**
weekly	45.06	51.26	**96.31**
daily	64.02	54.73	**118.75**
bi-weekly	82.53	99.66	**182.19**
⊟ **sandy**	**215.43**	**217.46**	**432.88**
bi-weekly	33.29	60.49	**93.78**
weekly	35.27	70.65	**105.92**
daily	146.87	86.32	**233.19**
⊟ **clay**	**280.24**	**207.13**	**487.37**
bi-weekly	59.80	68.15	**127.95**
daily	95.42	40.66	**136.09**
weekly	125.02	98.32	**223.34**
Total	**687.28**	**630.24**	**1,317.51**

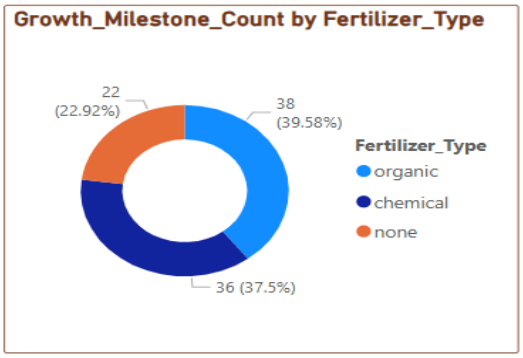
II. Average Temperature by Temperature Range Description



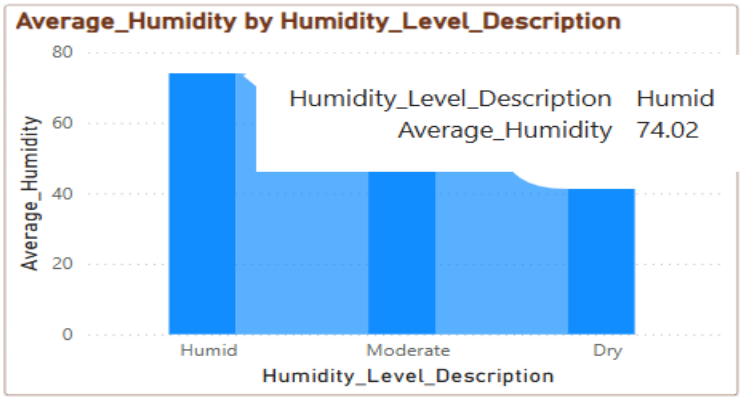
III. Temperature and Its Description According to Plant Growth



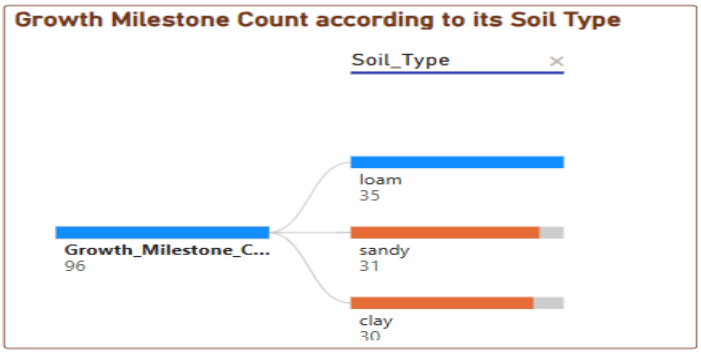
IV. Growth Milestone Count by Fertilizer Type



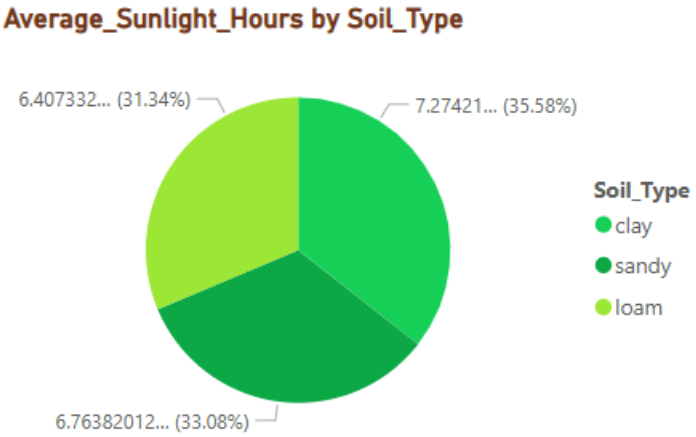
V. Average Humidity by Humidity Level Description



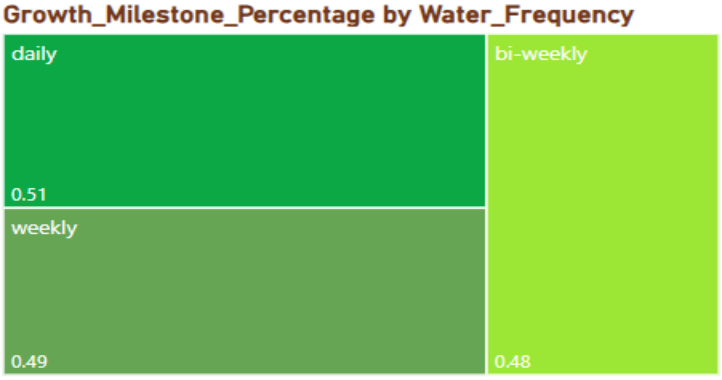
VI. Growth Milestone Count According to Its Soil Type



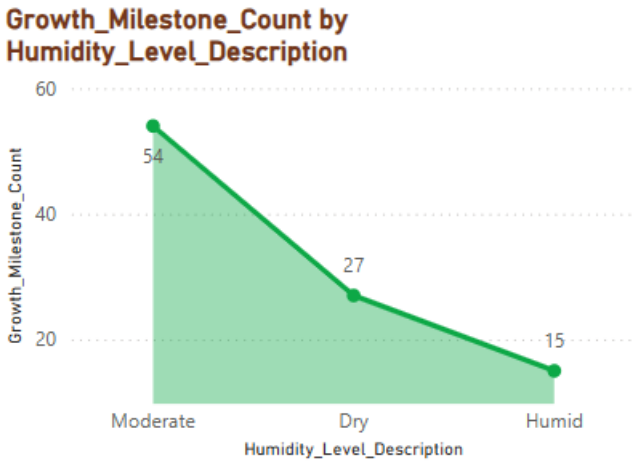
VII. Average Sunlight Hours by Soil Type



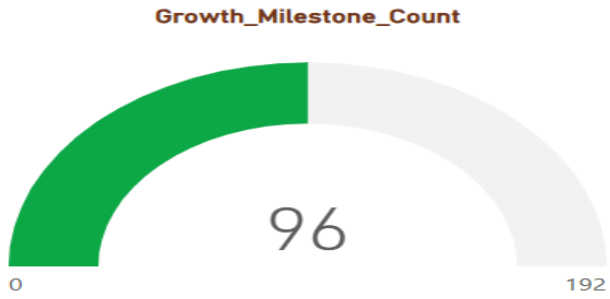
VIII. Growth Milestone Percentage by Water Frequency



IX. Growth Milestone Count by Humidity Level Description



X. Growth Milestone Count



XI. Average Humidity, Average Temperature, Average Sunlight Hours

Average_Sunlight_Hours	Average_Humidity	Average_Temperature
6.83	58.10	25.08

8. Conclusion/Observation

The plant growth analysis project successfully demonstrates how data-driven techniques can support informed agricultural decisions. By integrating environmental factors such as temperature, humidity, sunlight hours, soil type, fertilizer use, and water frequency, the project enables a deeper understanding of the conditions that influence plant growth milestones.

Using Power BI dashboards, we visualized and analyzed the impact of these factors interactively. Key findings revealed that moderate humidity and warm temperatures are highly conducive to achieving growth milestones. Furthermore, insights such as the higher growth success with organic fertilizers and daily watering schedules offer practical recommendations for farmers and agricultural planners.

The project also included important features such as filters, custom measures, calculated columns, and visual storytelling, enhancing user experience and decision-making. Overall, the solution provides a powerful framework for smart farming practices, enabling real-time monitoring, optimization, and sustainable growth strategies for future agricultural cycles.

9. Future Scope

- **Real-Time Data Integration**
  Incorporate IoT sensors for live tracking of soil and environmental conditions.
- **Predictive Analytics Using Machine Learning**
  Use historical data to forecast plant growth and yield trends.
- **Alert and Notification System**
  Notify users of unfavourable environmental changes instantly.
- **Mobile Application Support**
  Provide dashboard access on smartphones for field-level usage.
- **Integration with Satellite and Weather Data**
  Enhance insights by combining local data with satellite/weather feeds.
- **Custom Recommendations Engine**
  Offer personalized tips based on real-time and historical inputs.
- **Support for Multiple Crop Types**
  Expand analysis to cover a variety of crops and farming patterns.
- **Data Export and Sharing Enhancements**
  Allow easy export of reports in formats like Excel and PDF.

10. Appendix

Git hub Link:

https://github.com/manish6625/Predicting-Plant-Growth-Stages-with-Environmental-and-Management-Data-Using-Power-BI

Project Demo Link:

https://drive.google.com/file/d/1iNh7u3SmuQBvi1OXBuqCIZHiDB83M1_3/view?usp=sharing