Final Project Report

Project Title: Predicting Plant Growth Stages with Environmental and Management Data using Power Bi

Team Member: Chirag Gupta

1. Introduction

1.1 Project Overview

The project focuses on analyzing critical environmental and input variables that influence plant growth and building an interactive visualization platform to support precision agriculture. It integrates data collected from various sources, including sensors, manual field observations, and historical records, to understand how factors such as soil conditions, temperature, humidity, irrigation, and fertilizer usage impact crop performance. By combining advanced analytics with interactive dashboards, the system empowers farmers and stakeholders to visualize patterns, identify optimal growing conditions, and adopt data-driven cultivation strategies. This approach bridges the gap between data insights and practical field operations, enabling improved productivity, efficient resource management, and sustainable agricultural practices.

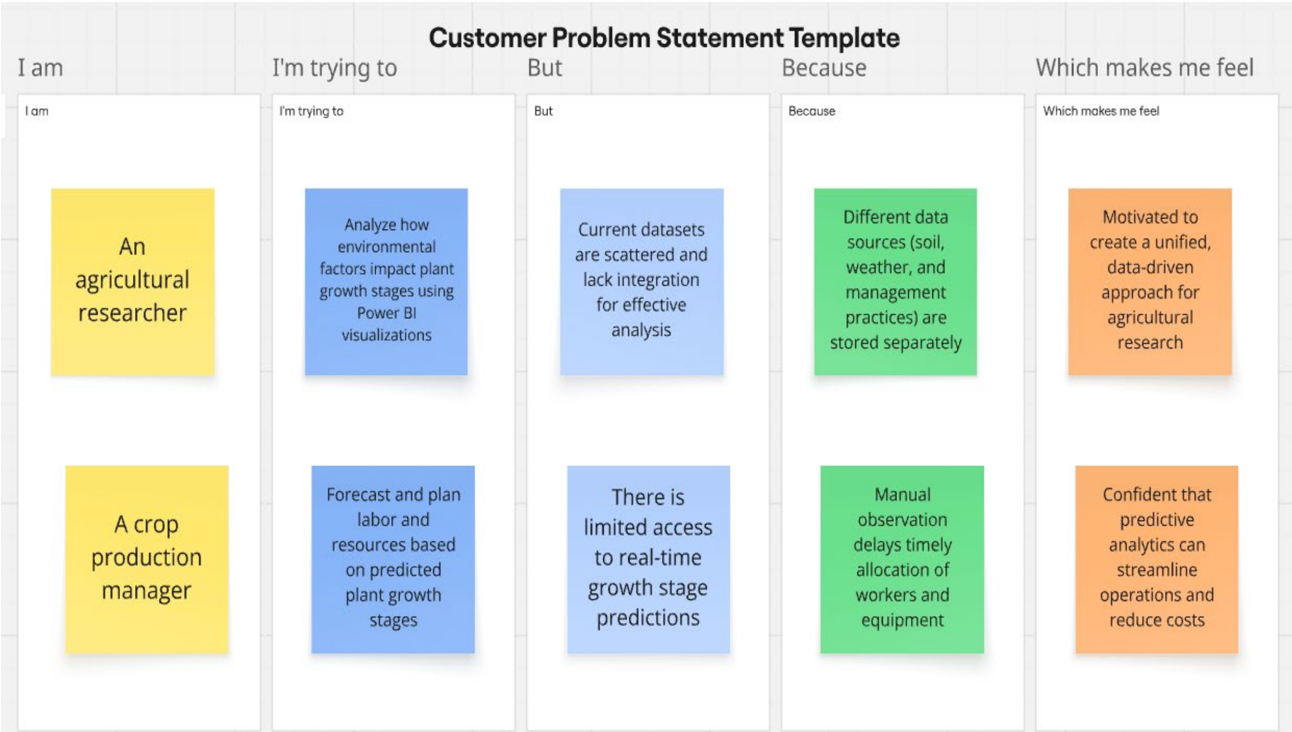
1.2 Objectives

- To analyze the impact of key environmental and input variables on plant growth.
- To establish a framework for efficient data collection, integration, and preprocessing.
- To identify critical factors such as soil properties, irrigation frequency, and fertilizer usage that most influence crop productivity.
- To design and develop an interactive visualization dashboard for exploring key insights and trends.
- To enable data-driven agricultural decision-making aimed at improving crop yield and sustainability.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	An agricultural researcher	Analyze how environmental factors impact plant growth stages using Power BI visualizations	Current datasets are scattered and lack integration for effective analysis	Different data sources (soil, weather, and management practices) are stored separately	Motivated to create a unified, data-driven approach for agricultural research
PS-2	A crop production manager	Forecast and plan labor and resources based on predicted plant growth stages	There is limited access to real-time growth stage predictions	Manual observation delays timely allocation of workers and equipment	Confident that predictive analytics can streamline operations and reduce costs



2.2. Project Proposal (Proposed Solution)

The project will adopt a modular, sprint-based development methodology, progressing through phases of data acquisition, processing, analysis, visualization, and evaluation. The development process includes:

- **Data Collection & Preprocessing:** Acquire relevant datasets on environmental parameters and plant growth milestones, followed by cleaning and structuring for analysis.
- **Data Analysis:** Apply statistical and comparative techniques to uncover patterns, correlations, and key growth influencers.
- **Dashboard Development:** Design interactive dashboards using Power BI to deliver clear and actionable visual insights.
- **Feature Enhancement:** Integrate advanced features such as comparison tools, alert notifications, and data export options.
- **Evaluation & Refinement:** Perform system performance assessment and incorporate user feedback for continuous improvement.

Key Features

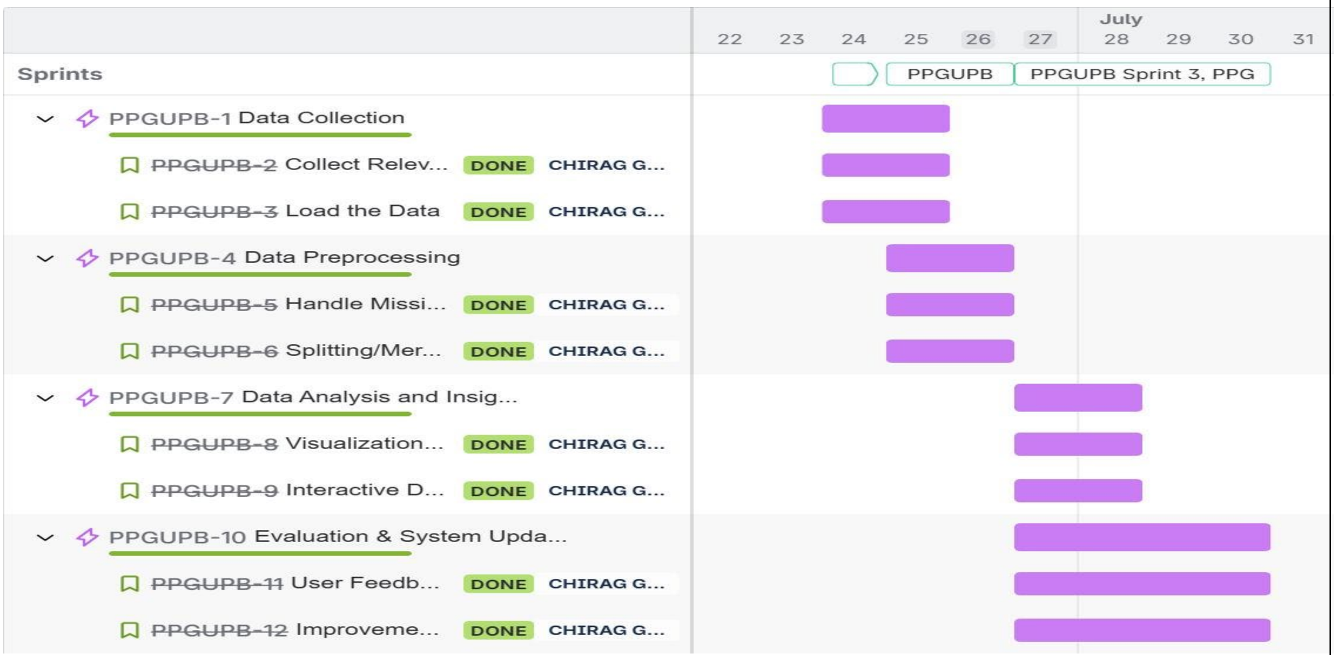
- Interactive and user-friendly dashboards
- Analysis of key growth influencers
- Tracking of environmental trends

- Condition-based cultivation 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	Chirag	23 July,2025	24 July,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	Chirag	23 July,2025	24 July,2025
	Load the Data	USN-3	I can load the collected data into a central storage or database for further processing.	1	Medium	Chirag	23 July,2025	24 July,2025
Sprint-2	Data Preprocessing	USN-4	This stage focuses on cleaning and organizing the raw data to make it analysis-ready.	4	High	Chirag	25 July,2025	26 July,2025
	Handle Missing Values	USN-5	I can identify and fill or remove missing values to improve data quality and consistency.	2	High	Chirag	25 July,2025	26 July,2025
	Splitting/Merging Fields	USN-6	I can split combined columns (e.g., DateTime into Date and Time) or merge fields where necessary to structure the dataset properly.	2	High	Chirag	25 July,2025	26 July,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	Chirag	27 July,2025	28 July,2025
	Visualization of Environmental Factors	USN-8	As a user, I can view graphs showing average sunlight, temperature, and humidity levels to understand growing	4	Medium	Chirag	27 July,2025	28 July,2025

			conditions.					
	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	Chirag	27 July,2025	28 July,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	Chirag	29 July,2025	30 July,2025
	User Feedback Collection	USN-11	Gathering input from farmers, analysts, and stakeholders.	5	Medium	Chirag	29 July,2025	30 July,2025
	Improvement Planning & Documentation	USN-12	Logging findings and planning improvements for future cycles.	3	Medium	Chirag	29 July,2025	30 July,2025



3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Section	Description

Project Overview	This project focuses on analyzing how environmental factors—including soil type, humidity, temperature, irrigation frequency, and fertilizer usage—affect plant growth. Its goal is to develop actionable insights and interactive dashboards that empower farmers to make data-driven decisions and enhance crop productivity.
Data Collection Plan	Data was gathered through a combination of field observations, environmental sensors, and manual entries from a controlled agricultural experiment. For each planting cycle, key parameters such as temperature, humidity, soil type, irrigation frequency, and growth milestones were systematically recorded.
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.

3.1.1 Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Plant Growth Data Classification,	Contains plant growth data with attributes such as soil	https://www.kaggle.com/datasets/g	CSV	12 KB	Public

Real-Dataset (Kaggle)	type, temperature, humidity, water frequency, fertilizer type, and growth milestones.	orororororo23/plant-growth-data-classification/data			
Indoor Plant Health and Growth Dataset	This dataset contains 1,000 entries with 17 features related to indoor plant health, growth metrics etc.	www.kaggle.com/datasets/souvikrana17/indoorplant-health-and-growth-dataset	CSV file, fully compatible with Python, R, Excel, etc	120.32 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 Columnavailable	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** Columncontains 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	Low input	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_data.csv** or loaded into Power BI for dashboard creation.

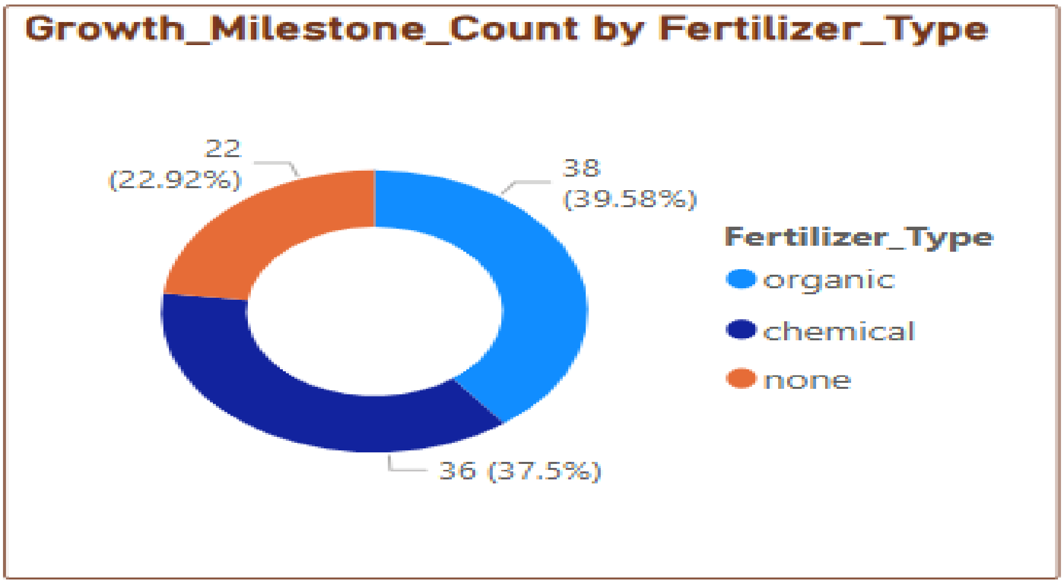
4. Data Visualization

4.1. Framing Business Questions

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

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

b. *Screenshot of visualization*

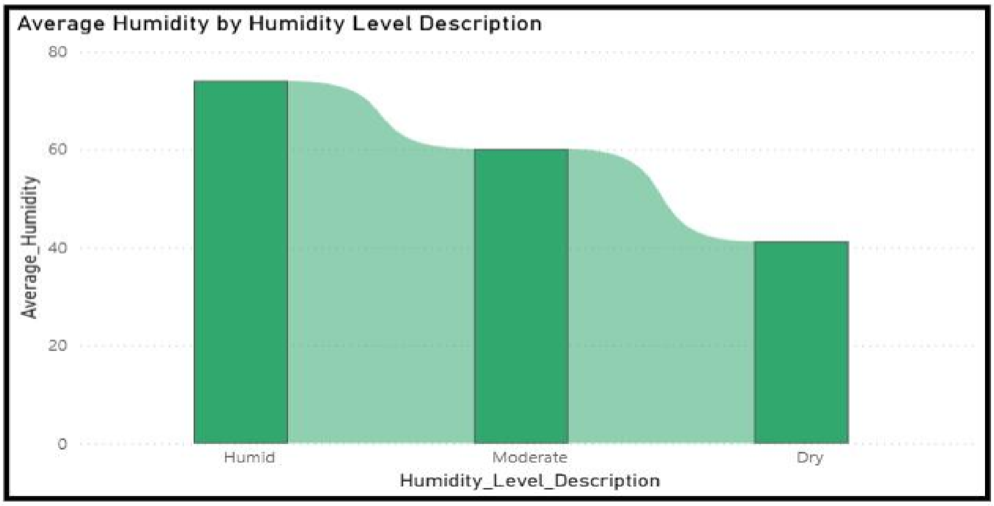


2. Which humidity level had the highest average humidity?

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

b. *Screenshot of visualization*

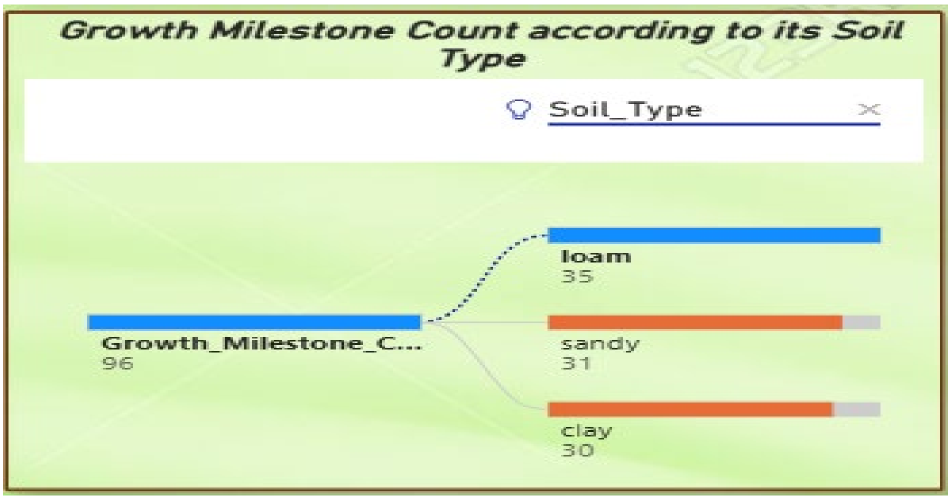
Insights : Humid conditions lead to the highest average humidity, emphasizing the importance of maintaining high humidity levels for optimal plant growth.



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

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

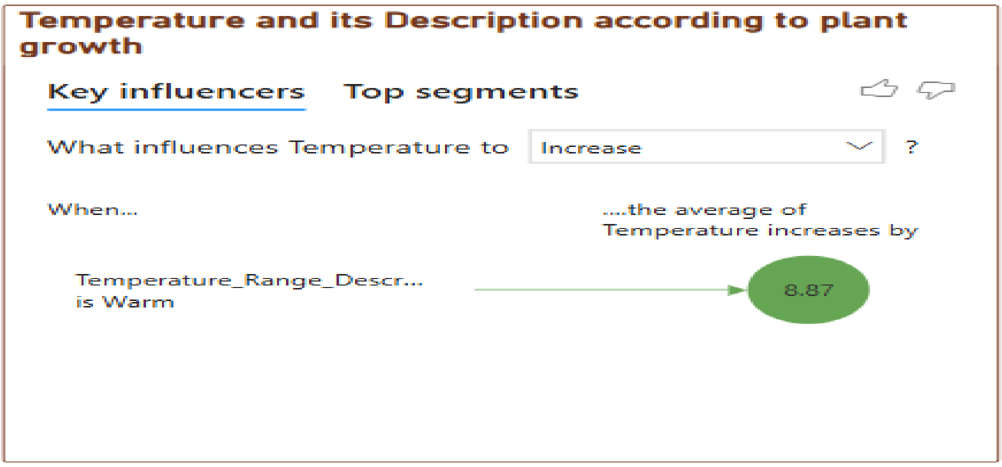
b. *Screenshot of visualization*



4. What condition causes temperature to increase most?

a. *Visualization*: Key Influencer for Temperature Increase

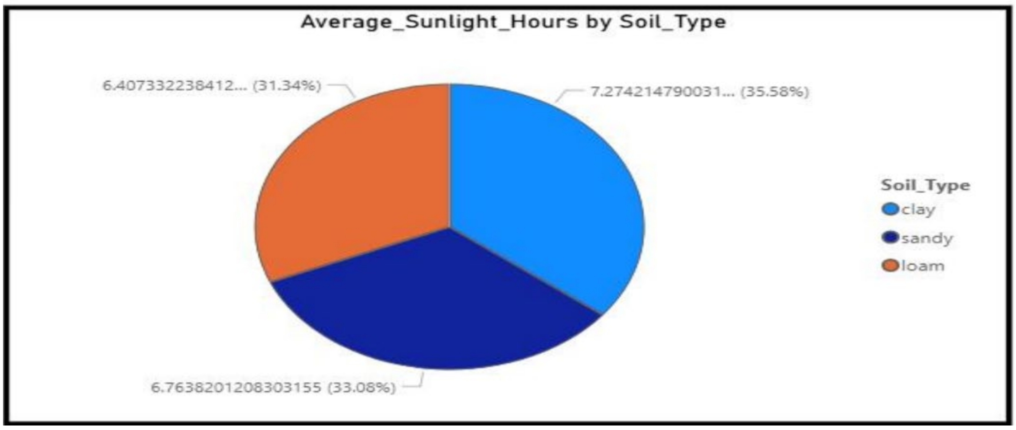
b. *Screenshot of visualization*



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

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

b. *Screenshot of visualisation*



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

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

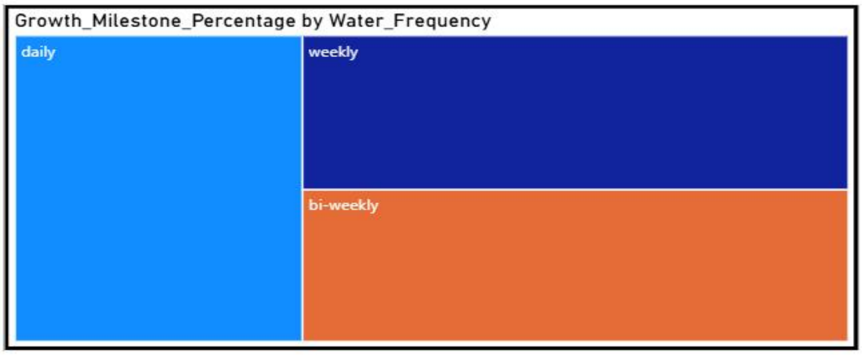
b. *Screenshot of visualization*

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

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

b. *Screenshot of visualization*

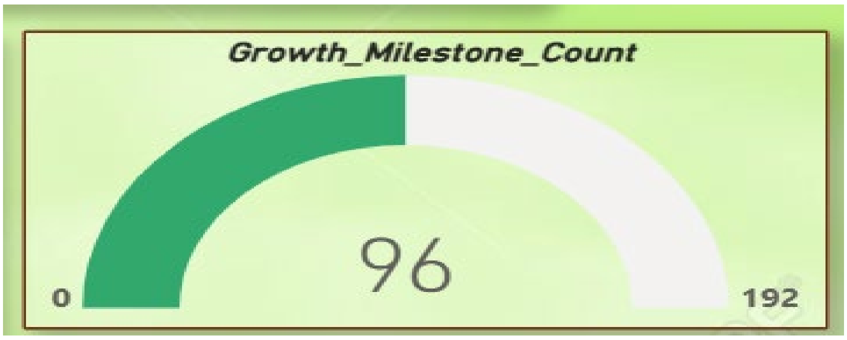
Insights : Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.



8. What is the total growth milestone count?

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

b. *Screenshot of visualization*



4.2. Developing Visualizations

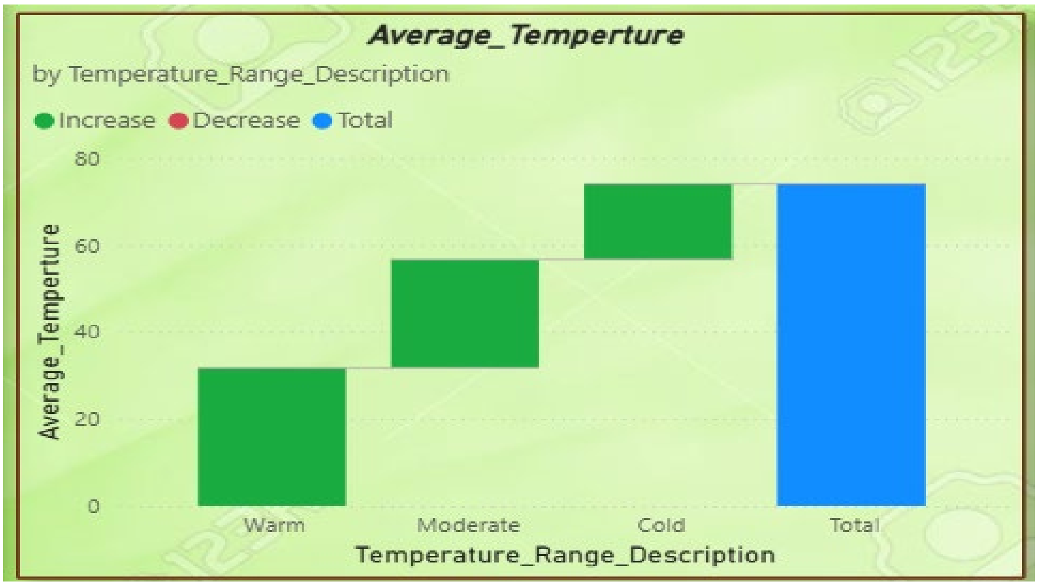
Activity 1.1: Water Frequency According to Its Soil Type

Water Frequency according to its soil Type

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

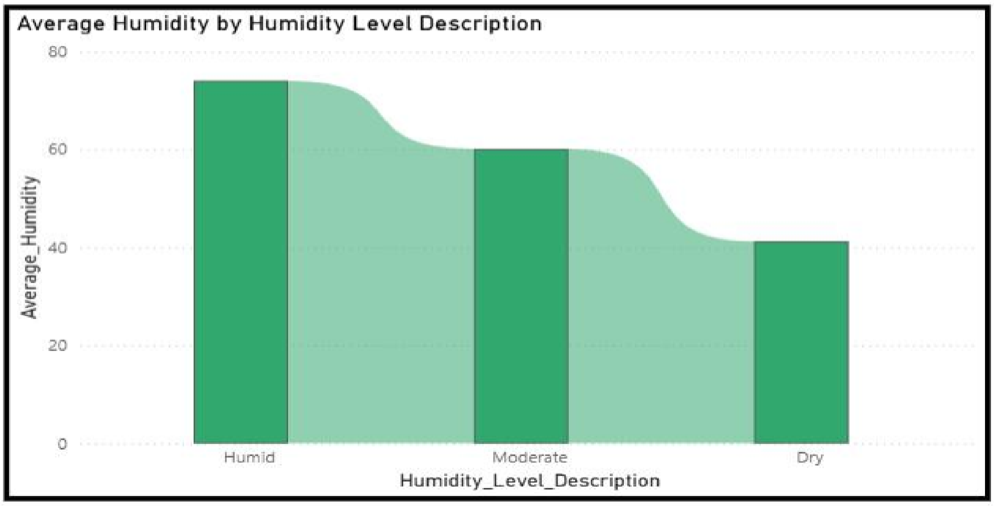
Activity 1.2: Average Temperature by Temperature Range Description



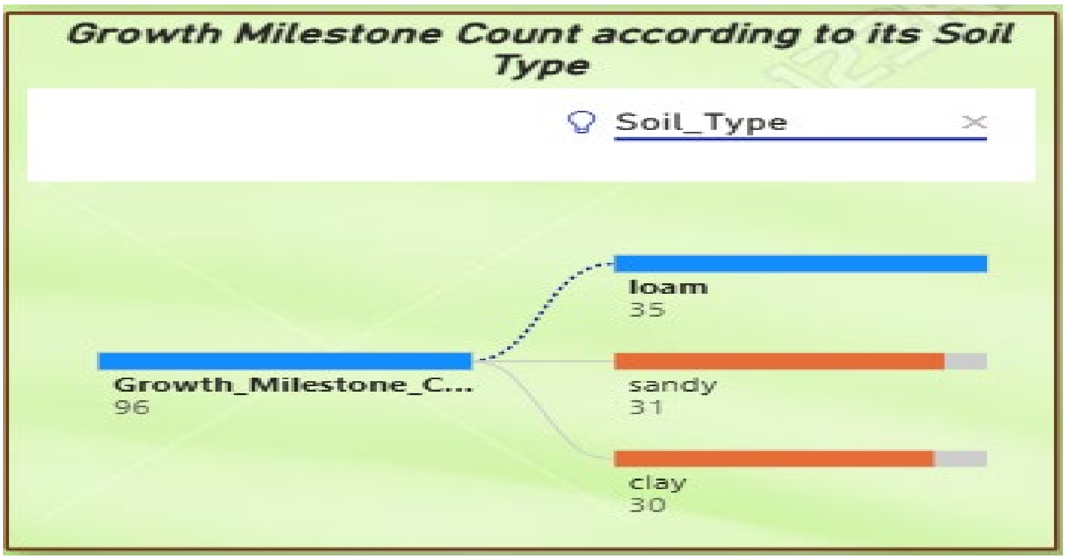
Insight: Moderate temperature ranges have the highest average temperature increase, suggesting that plants in this range experience the most growth.

Activity 1.3: 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.

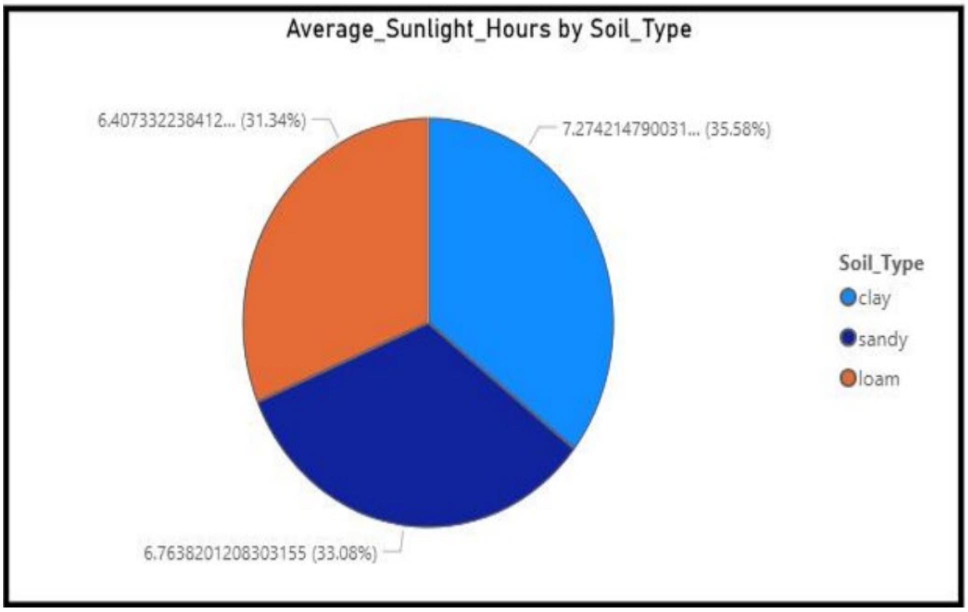


Activity 1.4: Growth Milestone Count According to Its Soil Type



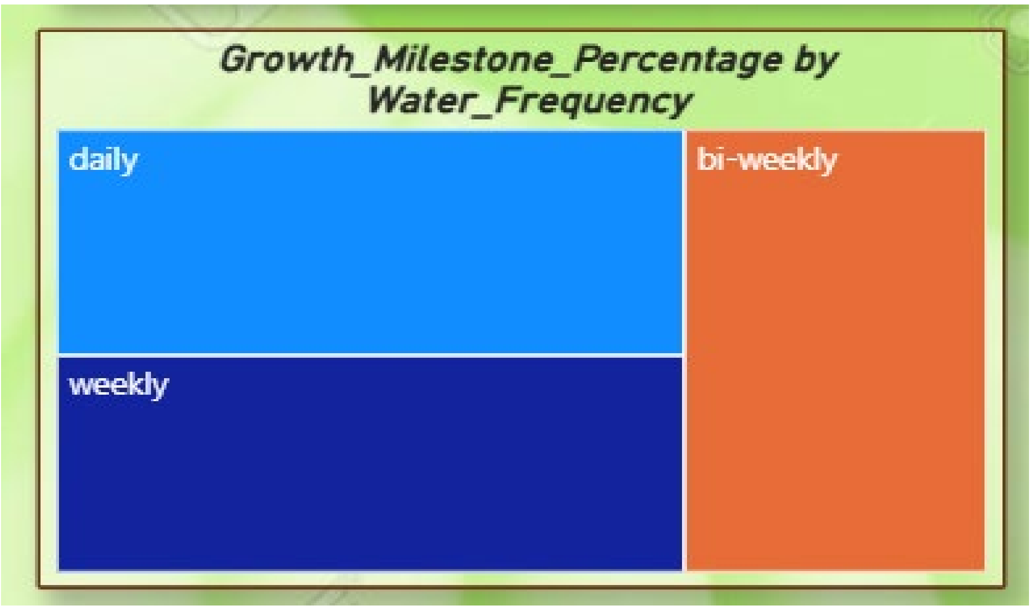
Insights:
Loam soil has the highest growth milestone count, suggesting it is the most conducive soil type for achieving plant growth milestones.

Activity 1.5: 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 1.6: Growth Milestone Percentage by Water Frequency

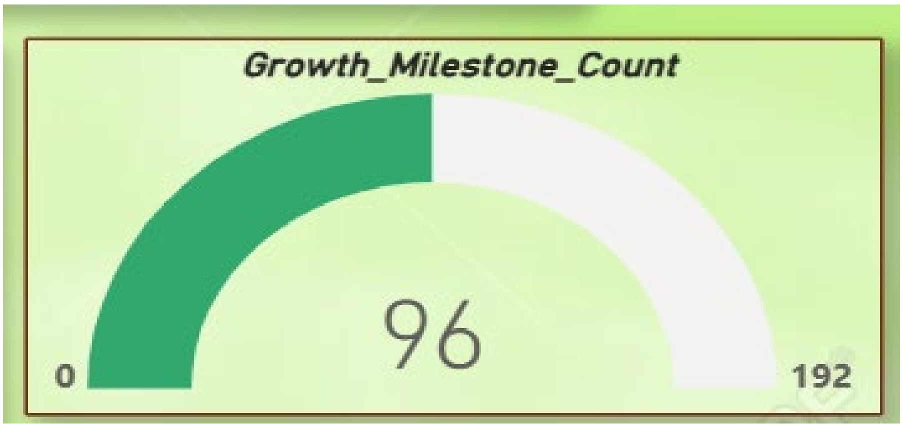


Insights: Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.

Activity 1.7: 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.

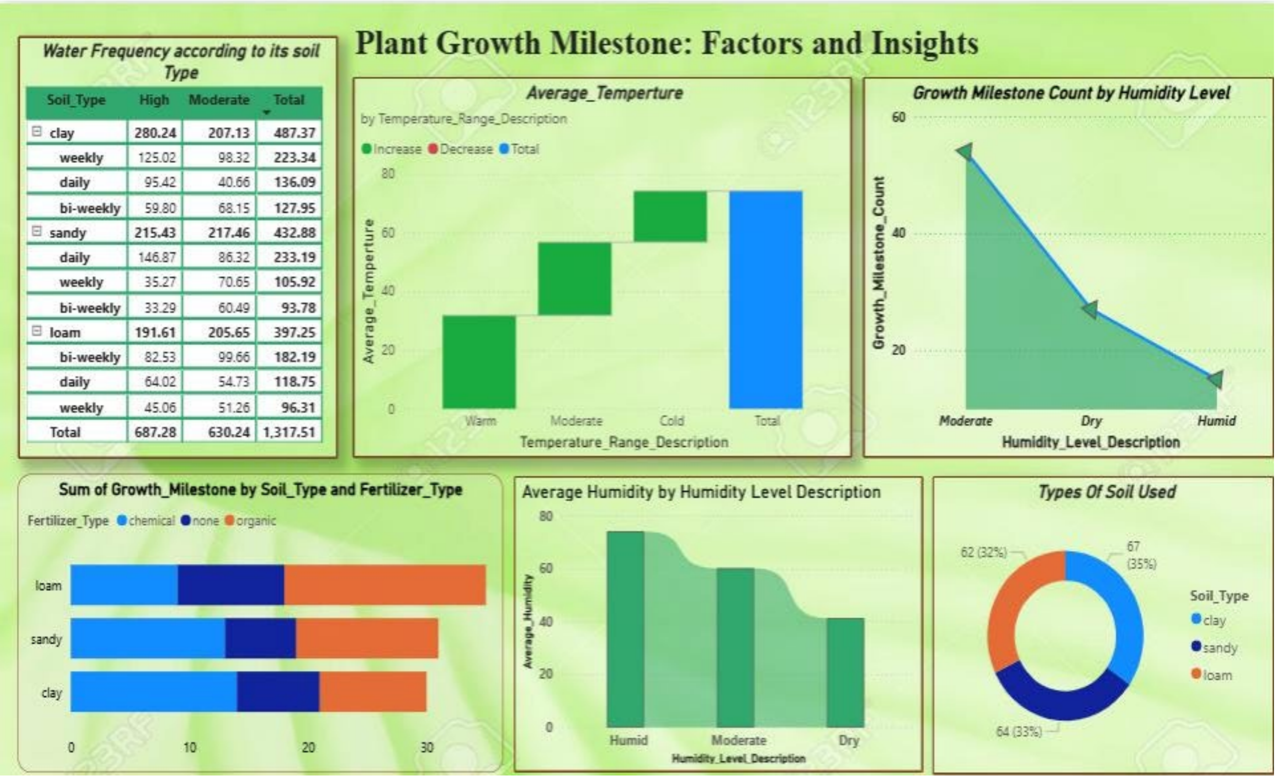
Activity 1.8: Growth Milestone Count



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

5. Dashboard

5.1. Dashboard Design File



Note: Highlight the major outcomes in form of bullet points

Here are six potential insights derived from the given dashboard image:

i. Clay Soil Consumes the Highest Water Volume

The **Water Frequency by Soil Type** table shows that clay soil requires the highest total water (487.37 units), followed by sandy (433.34 units) and loam (397.25 units). Clay soil’s dense structure likely leads to slower drainage, necessitating more frequent watering, especially under high humidity conditions.

ii. Moderate Humidity Levels Drive Maximum Growth Milestones

The **Growth Milestone Count by Humidity Level** chart indicates that moderate humidity levels support the highest growth milestones (~55), followed by dry (~28) and humid (~15) conditions. This suggests that excessive or insufficient humidity negatively impacts overall plant growth performance.

iii. Loam Soil Performs Best with Organic Fertilizers

From the **Sum of Growth Milestones by Soil Type and Fertilizer Type** visual, loam soil paired with organic fertilizers shows the highest milestone count. This implies loam's balanced soil texture and organic fertilizers' long-term nutrient release offer optimal conditions for plant development.

iv. Average Humidity Peaks in Humid Environments

The **Average Humidity by Humidity Level Description** chart confirms that humid areas maintain the highest average humidity (~75%), while dry regions show the lowest (~45%). This helps identify environmental zones where humidity regulation strategies are essential for better crop outcomes.

v. Soil Type Usage Is Well Distributed

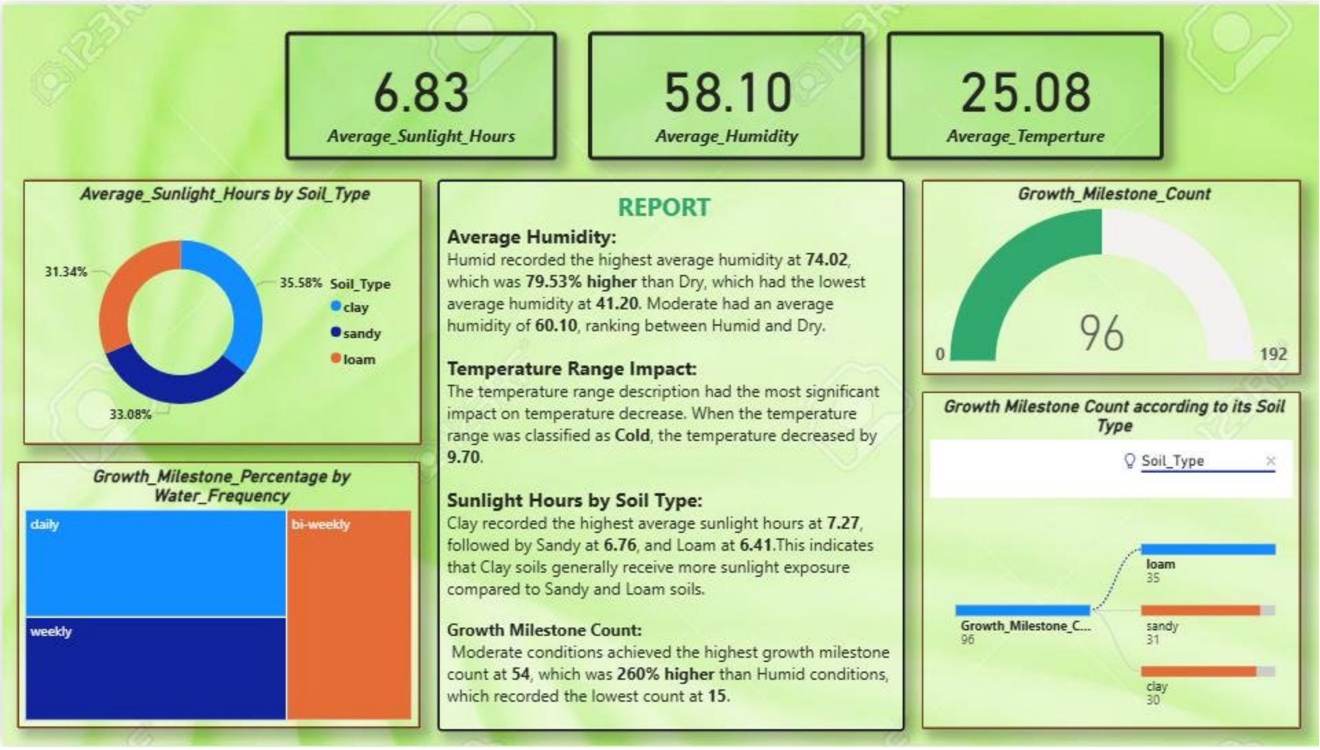
The **Types of Soil Used** donut chart shows relatively even distribution among soil types: clay (35%), sandy (33%), and loam (32%). This indicates a diverse experimental setup and suggests that findings apply across a broad range of soil conditions rather than being biased toward one type.

vi. Temperature Variability Impacts Growth Outcomes

The **Average Temperature by Temperature Range** waterfall chart highlights significant increases in plant growth in warmer climates compared to colder conditions. Warmer environments correlate with better growth milestones, emphasizing the importance of temperature management in agricultural planning.

6. Report

6.1. Story Design File



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

i. Clay Soil Requires the Highest Water Volume

The **Water Frequency by Soil Type** table shows:

- Clay soil consumes the highest total water volume of **487.37 units** (280.24 for high humidity and 207.13 for moderate humidity).
- Sandy soil follows with **433.34 units** (215.43 high, 217.46 moderate).
- Loam soil uses the least water at **397.25 units** (191.61 high, 205.65 moderate).
  This highlights clay's dense structure, which demands frequent watering, especially under **daily watering schedules (146.87 units)**.

ii. Moderate Humidity Delivers Maximum Growth Milestones

From the **Growth Milestone Count by Humidity Level Description** chart:

- Moderate humidity achieves **55 milestones**.
- Dry conditions result in **28 milestones**.
- Humid conditions have the lowest count at **15 milestones**. This demonstrates that balanced moisture conditions are more conducive to plant growth compared to extremely humid or dry conditions.

iii. Loam Soil with Organic Fertilizer Shows Highest Performance

The **Growth Milestone by Soil Type and Fertilizer Type** visual indicates:

- Loam soil paired with organic fertilizers produces the highest growth milestones (**~30 milestones**).
- Sandy soil with chemical fertilizers ranks second with **~25 milestones**.
- Clay soil records fewer milestones (**~20 milestones**) despite higher water consumption.
  This suggests that soil structure combined with sustainable fertilizer choices plays a critical role in growth outcomes.

iv. Humid Conditions Record the Highest Average Humidity

The **Average Humidity by Humidity Level Description** chart reveals:

- Humid conditions: **~75% humidity**
- Moderate conditions: **~60% humidity**
- Dry conditions: **~45% humidity**
  Although humidity levels are highest under humid conditions, the milestone count is lowest, indicating excess moisture can inhibit optimal growth.

v. Soil Type Distribution is Fairly Balanced

The **Types of Soil Used** donut chart indicates an even distribution:

- Clay soil: **67 instances (35%)**

- Sandy soil: **64 instances (33%)**
- Loam soil: **62 instances (32%)**
  This ensures that conclusions are not biased toward a single soil type and are applicable across varied soil conditions.

vi. Warm Temperature Conditions Enhance Plant Growth

The **Average Temperature by Temperature Range** waterfall chart shows:

- Warm conditions increase temperature impact by **~8.87 units**, leading to better growth outcomes.
- Cold conditions reduce temperature significantly, decreasing plant productivity.
  This emphasizes the importance of maintaining optimal temperature ranges for sustainable growth.

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.

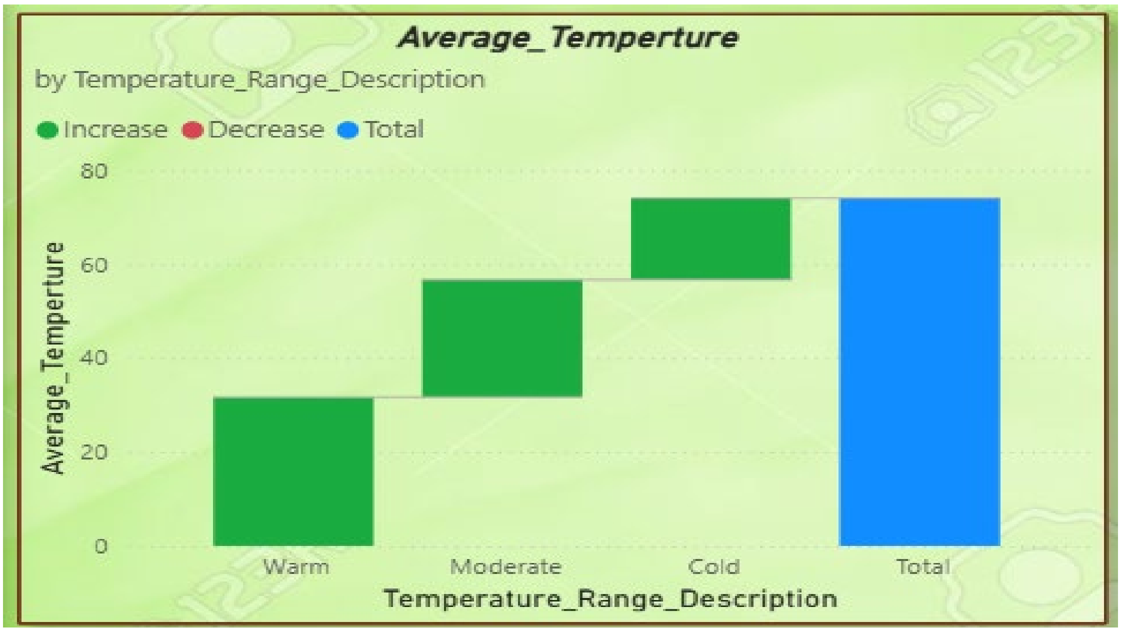
7.3 Number of Visualization

1: Water Frequency According to Its Soil Type

Water Frequency according to its soil Type

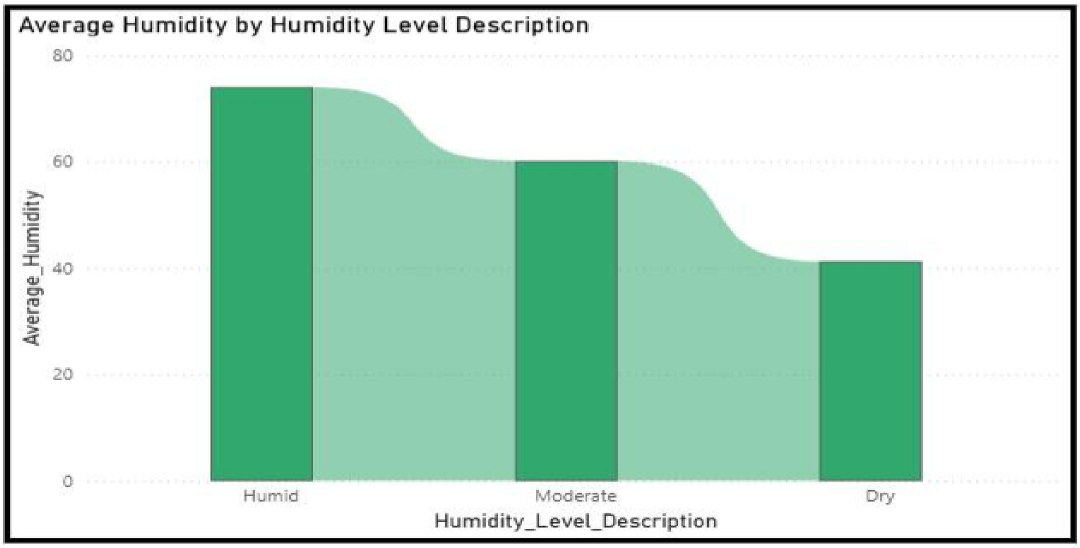
Soil_Type	High	Moderate	Total
⊟ clay	280.24	207.13	487.37
weekly	125.02	98.32	223.34
daily	95.42	40.66	136.09
bi-weekly	59.80	68.15	127.95
⊟ sandy	215.43	217.46	432.88
daily	146.87	86.32	233.19
weekly	35.27	70.65	105.92
bi-weekly	33.29	60.49	93.78
⊟ loam	191.61	205.65	397.25
bi-weekly	82.53	99.66	182.19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

2: Average Temperature by Temperature Range Description

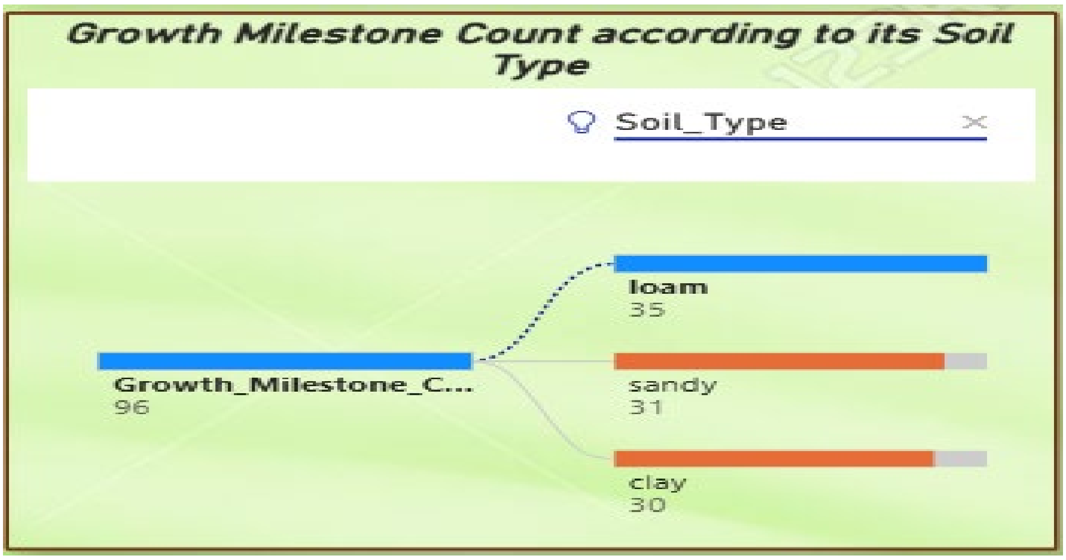


3: Average Humidity by Humidity Level Description

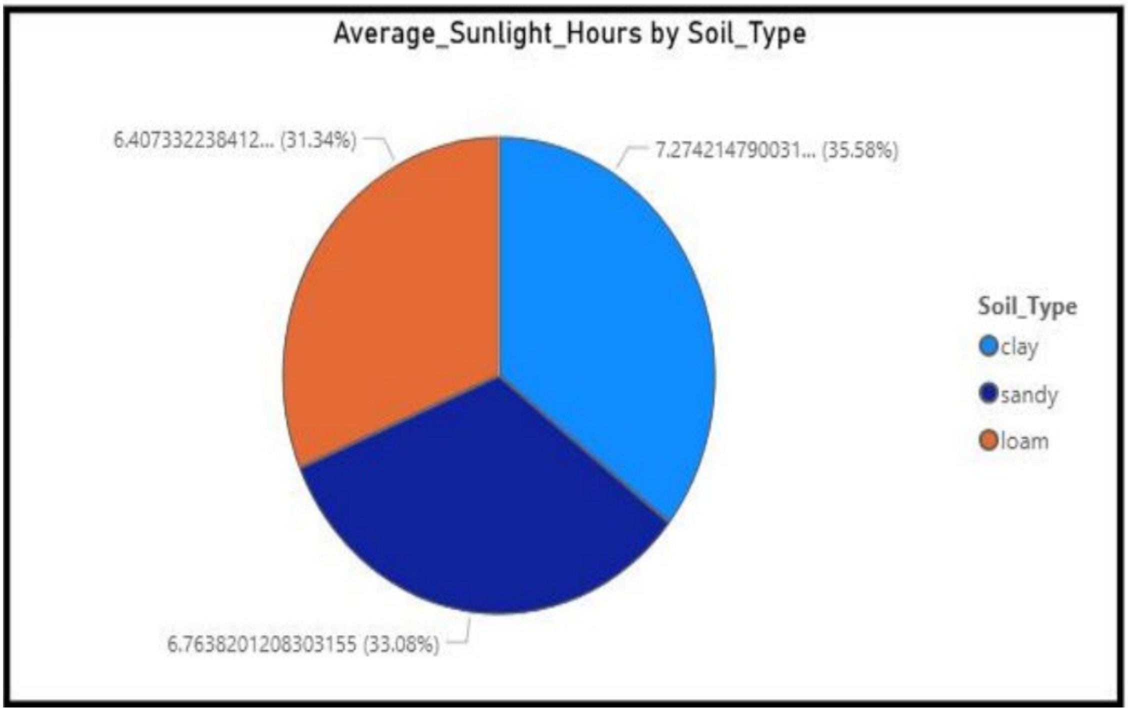
Insights : Humid conditions lead to the highest average humidity, emphasizing the importance of maintaining high humidity levels for optimal plant growth.



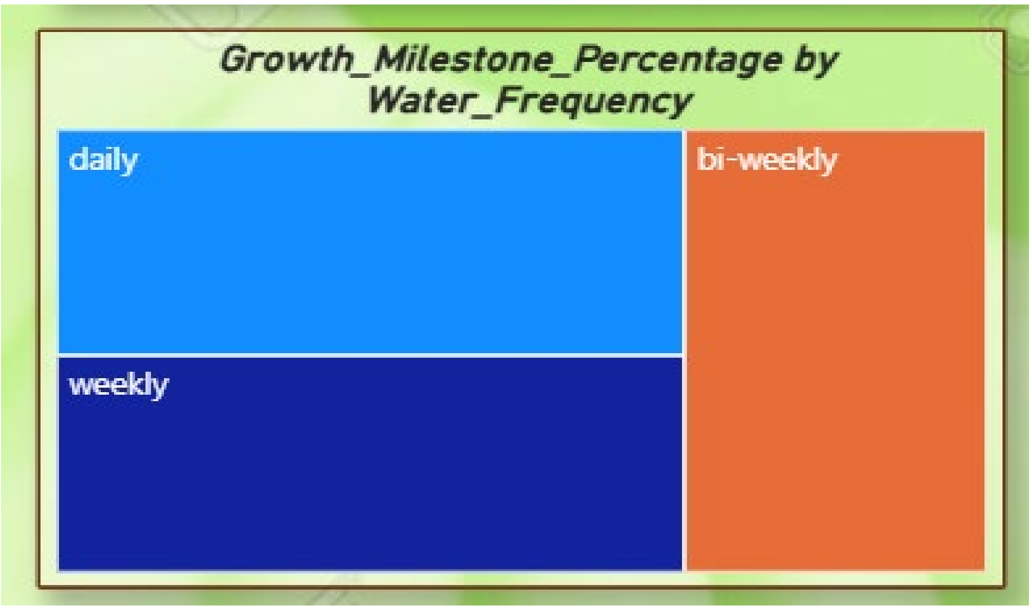
4: Growth Milestone Count According to Its Soil Type



5: Average Sunlight Hours by Soil Type

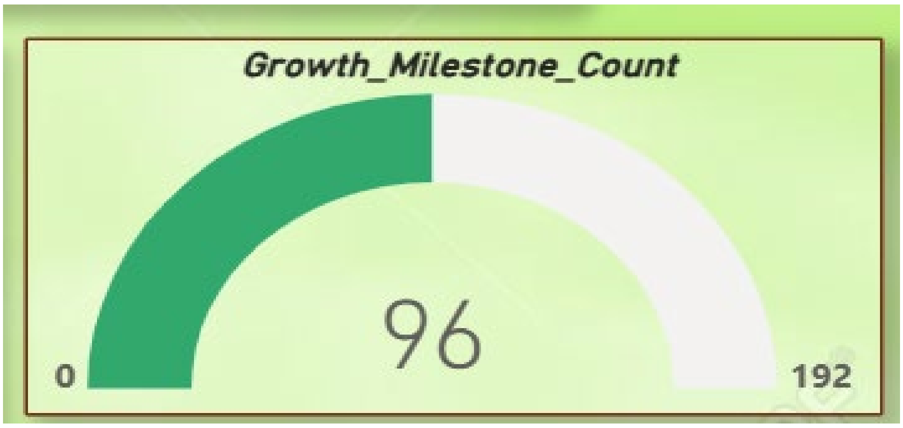


6: Growth Milestone Percentage by Water Frequency



7: Growth Milestone Count by Humidity Level Description

8: Growth Milestone Count



9. Average Humidity, Average Temperature, Average Sunlight Hours

8. Conclusion/Observation

The plant growth analysis project effectively showcases how data-driven approaches can enhance agricultural decision-making. By combining key environmental factors—including temperature, humidity, sunlight hours, soil type, fertilizer type, and water frequency—the project provides valuable insights into conditions that influence plant growth milestones.

Through interactive Power BI dashboards, the impact of these factors was analyzed and visualized, enabling deeper exploration and trend identification. The analysis highlighted that moderate humidity and warm temperature ranges significantly contribute to improved growth outcomes. Additionally, findings such as the superior performance of organic fertilizers and the benefits of daily watering offer actionable recommendations for farmers and agricultural planners.

The solution incorporates advanced features like filters, custom measures, calculated columns, and visual storytelling, enriching user experience and supporting informed decisions. Overall, this project delivers a robust framework for smart farming, empowering 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

Github Link:

https://github.com/chirag-2004/Predicting-Plant-Growth-Stages-with-Environmental-and-Management-Data-using-Power-BI

Project Demo Link:

https://drive.google.com/file/d/1sFxdWQwlFMdnEA9wOb0-xHAnBodxvVGq/view?usp=drive_link