Predicting Plant Growth Stages with Environmental and Management

1. Introduction

1.1. Project overviews

A data-driven initiative to enhance agricultural efficiency by analyzing the impact of environmental and management factors on plant growth. This project utilizes Power BI to explore, visualize, and model data related to various agricultural parameters—including soil type, sunlight exposure, humidity, temperature, water frequency, and fertilizer usage. The aim is to identify patterns and correlations that influence different plant growth stages. With accurate insights, this project seeks to bridge the knowledge gap for farmers, enabling smarter decisions, optimal resource utilization, and consistent crop performance.

1.2. Objectives

Examine environmental and input features such as soil type, sunlight hours, humidity, temperature, water frequency, and fertilizer usage to assess their individual and combined effects on plant growth.

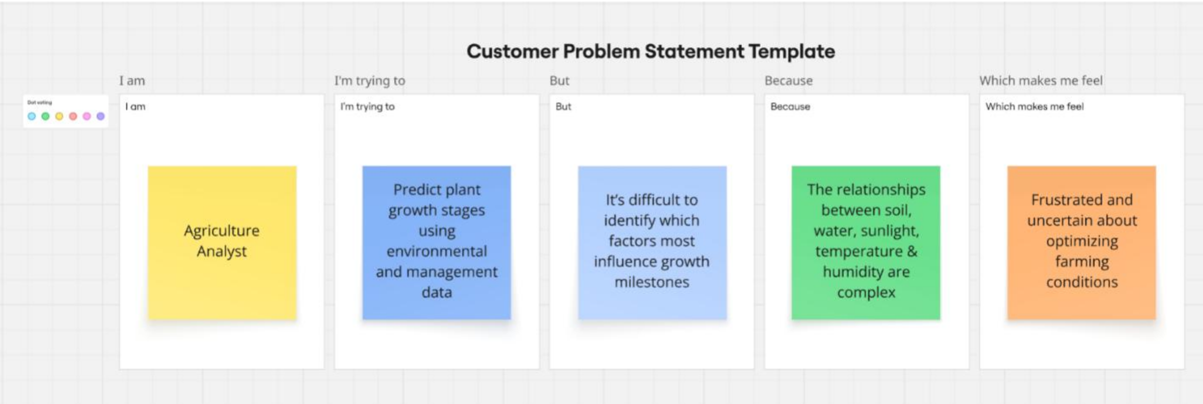
Determine the most significant variables contributing to plant development at various growth milestones.

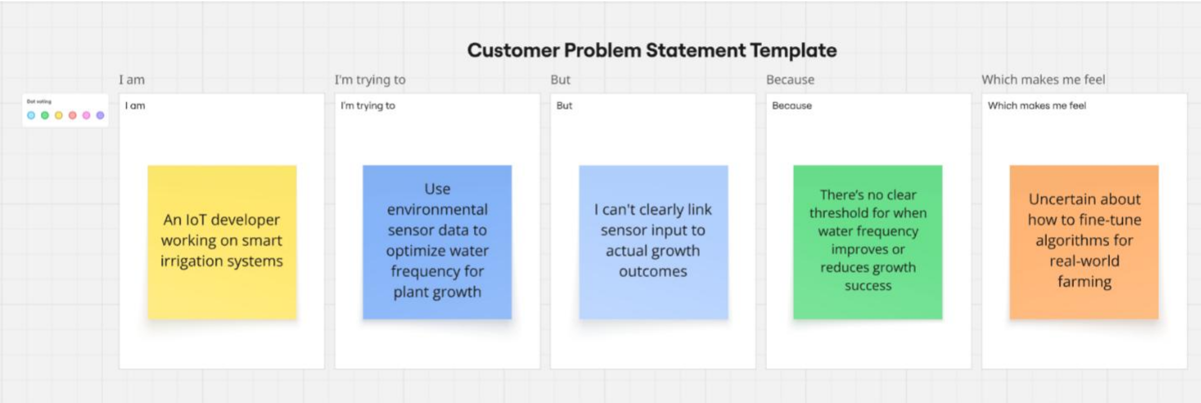
Use Power BI to visualize patterns, correlations, and growth stage classifications in an intuitive, user-friendly format for stakeholders.

Provide actionable insights and recommendations for environmental and resource conditions that promote efficient and sustainable plant growth.

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	Agriculture Analyst	Predict plant growth stages using environmental and management data	It's difficult to identify which factors most influence growth milestones	The relationships between soil, water, sunlight, temperature & humidity are complex	Frustrated and uncertain about optimizing farming conditions
PS-2	An IoT developer working on smart irrigation systems	Use environmental sensor data to optimize water frequency for plant growth	I can't clearly link sensor input to actual growth outcomes	There's no clear threshold for when water frequency improves or reduces growth success	Uncertain about how to fine-tune algorithms for real-world farming

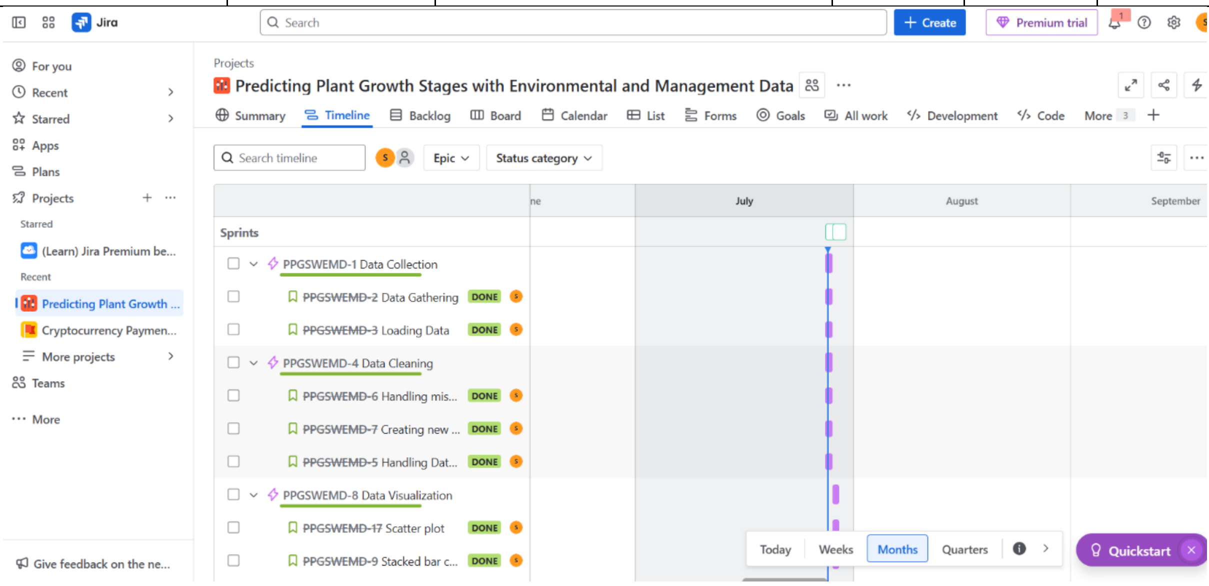
2.2. Project Proposal (Proposed Solution)

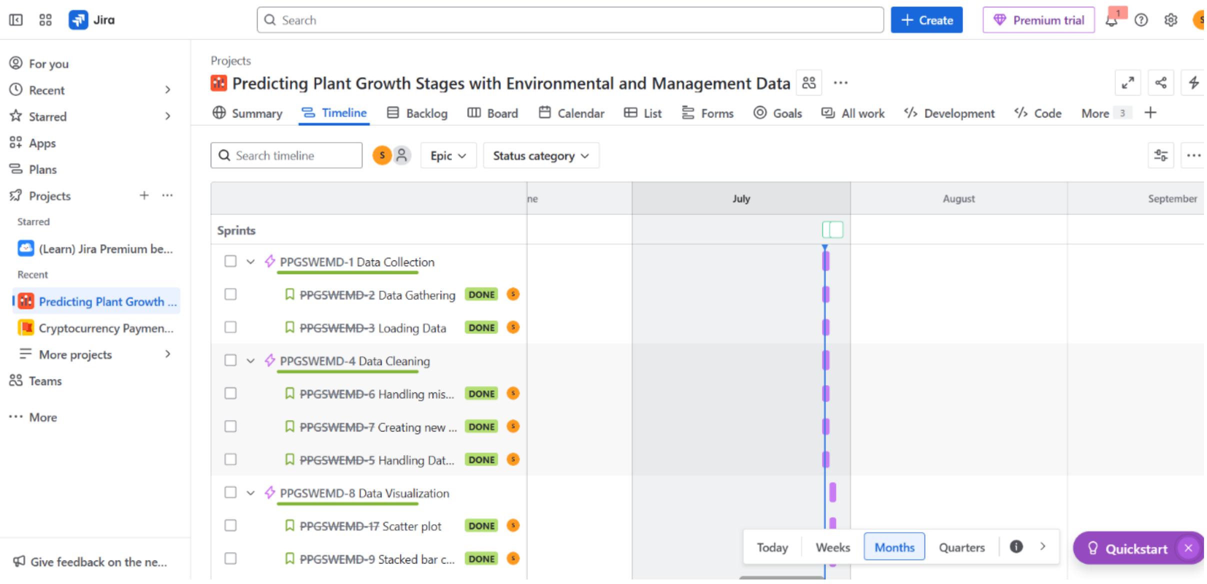
Project Overview	
Objective	To analyze environmental and agricultural data to identify key factors influencing plant growth and recommend optimal conditions for enhanced growth performance.
Scope	This project involves analyzing a dataset with multiple features such as soil type, sunlight hours, humidity, temperature, and fertilizer usage to classify growth milestones. The analysis is performed using data visualization and statistical modeling in Power BI.
Problem Statement	
Description	Farmers often lack precise insights into how environmental and input variables such as soil, water frequency, sunlight, and fertilizer impact plant development. This leads to suboptimal resource allocation and inconsistent yields.
Impact	By identifying and understanding the variables that most affect plant growth, this project can support data-driven decision-making in agriculture, increasing crop yield, reducing input waste, and improving sustainability.
Proposed Solution	
Approach	Use data cleaning and visualization techniques in Power BI to derive insights from the dataset. Categorize and analyze temperature, humidity, and water patterns to predict plant growth stages.
Key Features	Use of Power BI for rich data visualization, Feature mapping for growth stage classification, Identification of optimal ranges for temperature and humidity.

2.3. Initial Project Planning

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Data Collection	PPGSWEMD-2	Data gathering	3	High	Sukriti
Sprint-1	Data Collection	PPGSWEMD-3	Loading Data	1	Low	Sukriti
Sprint-1	Data Cleaning	PPGSWEMD-5	Handling Data type	2	Medium	Sukriti
Sprint-1	Data Cleaning	PPGSWEMD-6	Handling missing values	3	Medium	Sukriti
Sprint-1	Data Cleaning	PPGSWEMD-7	Creating new fields	3	Medium	Sukriti
Sprint-2	Data Visualization	PPGSWEMD-9	Stacked bar chart	3	Medium	Sukriti
Sprint-2	Data Visualization	PPGSWEMD-10	Column chart	3	Medium	Sukriti
Sprint-2	Data Visualization	PPGSWEMD-11	Line chart	3	Medium	Sukriti
Sprint-2	Data Visualization	PPGSWEMD-12	Pie chart	3	Medium	Sukriti
Sprint-2	Data Visualization	PPGSWEMD-17	Scatter plot	3	Medium	Sukriti
Sprint-2	Dashboard	PPGSWEMD-15	Create dashboard	5	High	Sukriti

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-2	Report	PPGSWEMD -16	Create report	5	High	Sukriti





3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

Section	Description
Project Overview	To analyze environmental and agricultural data to identify key factors influencing plant growth and recommend optimal conditions for enhanced growth performance
Data Collection Plan	Smart Bridge Dataset, Kaggle Dataset
Raw Data Sources Identified	https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification https://www.kaggle.com/datasets/aastik1844/plant-growth-stage-dataset

Source Name	Description	Location/URL	Format	Size	Access Permissions
Smart Bridge	Plant growth classification data based on environmental features such as light intensity, soil moisture, temperature, and pH levels.	https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification	CSV	11.2 KB	Public

Kaggle Dataset	This dataset contains observations of various growth stages of a subject (e.g., plants, animals, or cells), along with environmental and physiological factors influencing development. It includes features such as time-based measurements, nutrient levels, environmental conditions, and past growth trends.	https://www.kaggle.com/datasets/aastik1844/plant-growth-stage-dataset	CSV	674.4 MB	Public
…	…	…	…	…	…

3.2. Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
Smart Bridge	Missing values or incomplete records. e.g. missing growth stage labels	Moderate	Impute missing continuous features
Smart Bridge	Outliers/extreme values. Some plant measurements (e.g. height, leaf count) may be erroneous.	Low to Moderate	Detect outliers. Clip or remove if erroneous; otherwise normalize.
Smart Bridge	Data format heterogeneity. Mixed types, free-text entries, or inconsistent units across records.	Low	Standardize units; convert categorical text to codes; ensure consistent schema

3.3. Data Exploration and Preprocessing

Section	Description
Data Overview	Name of Dataset -Plant Growth Data Classification Size of Dataset - 11.49 kB 7 Columns – 3 String, 3 Decimal, 1 Integer 193 rows
Data Cleaning	Grouping (Fields like Humidity, Temperature, and Water_Frequency were transformed into categorical bins (e.g., Humidity_Range, Temperature_Range, Water_Frequency_Total)), Duplicate Handling (duplicate removal have been applied), Derived Columns (Columns such as Growth_Milestone_Percentage, Water_Frequency_Total, and

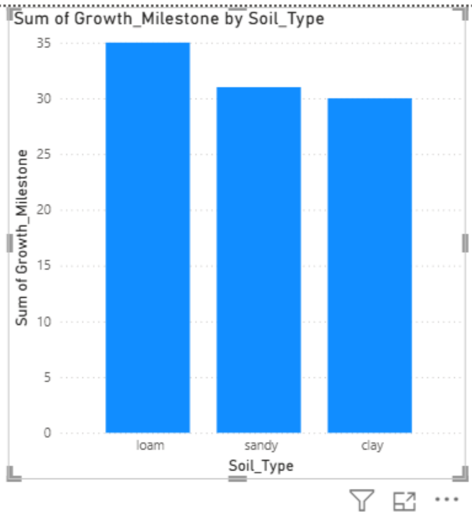
	Humidity_Range suggest transformations or calculated columns from original numerical values.), Outlier Detection (Temperature_Range)
Data Transformation	Filtering (values like “High,” “Low,” “Moderate”), sorting (Power BI often auto-sorts previewed data.) and creating calculated columns. (Fields like Water_Frequency_Total, Growth_Milestone_Percentage, and Humidity_Range are clearly derived from base columns.)
Data Type Conversion	Categorical → Numeric Water_Frequency (text or categorical) Water_Frequency_Numeric (numeric version of the same)
Column Splitting and Merging	Derived (Humidity_Range, Water_Frequency_Total, Temperature_Range)
Data Modeling	Measures are created (e.g. Average_Humidity, Average_Sunlight_hours).
Save Processed Data	plant_growth_data

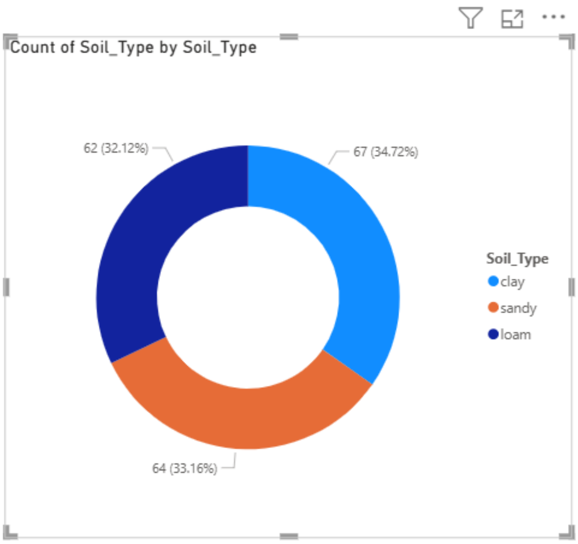
4. Data Visualization

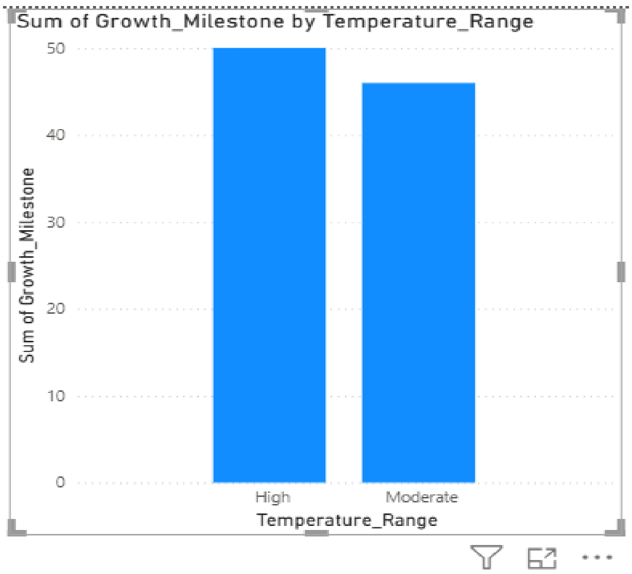
4.1. Framing Business Questions

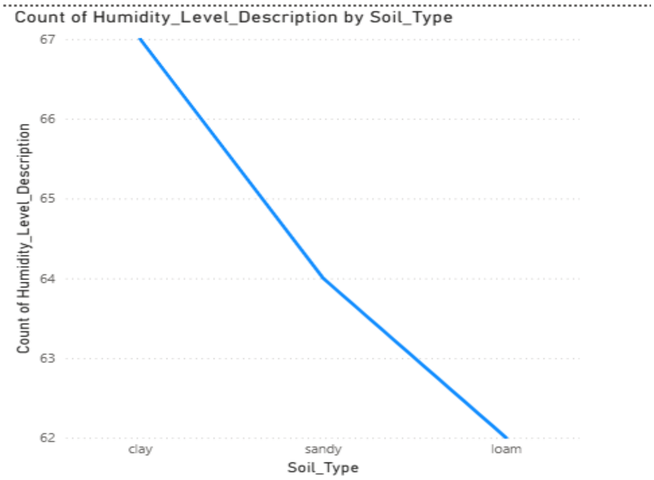
1. Which soil type is most prevalent in the dataset?
2. How does soil type affect plant growth milestone?
3. Which temperature range is most favorable for plant growth?
4. How does humidity level vary across soil types?
5. What is the impact of sunlight hours on different soil types?
6. How do fertilizer types influence plant growth across soil types?
7. What is the Growth stage distribution under different environmental conditions?
8. Which combination of environmental factors results in the fastest growth?

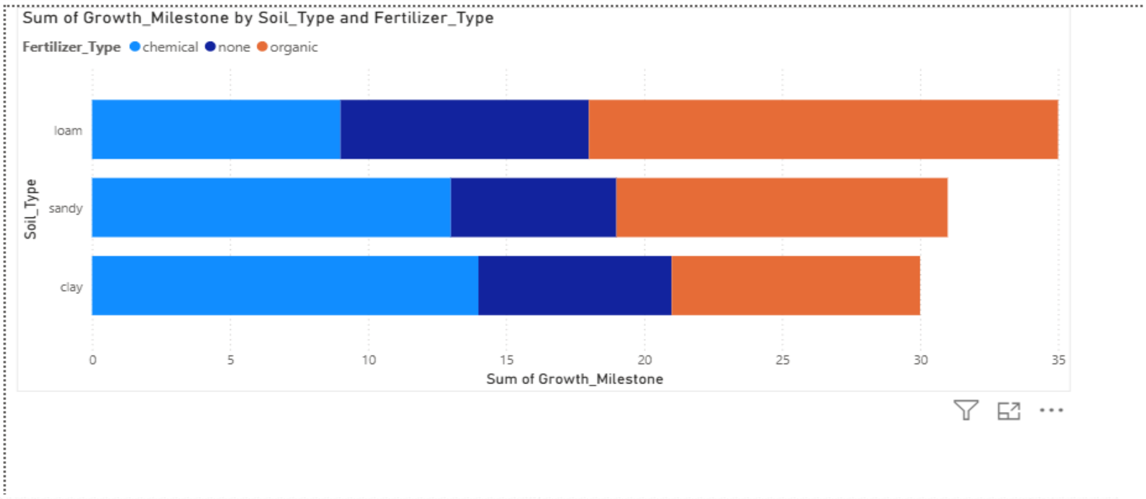
4.2. Developing Visualizations

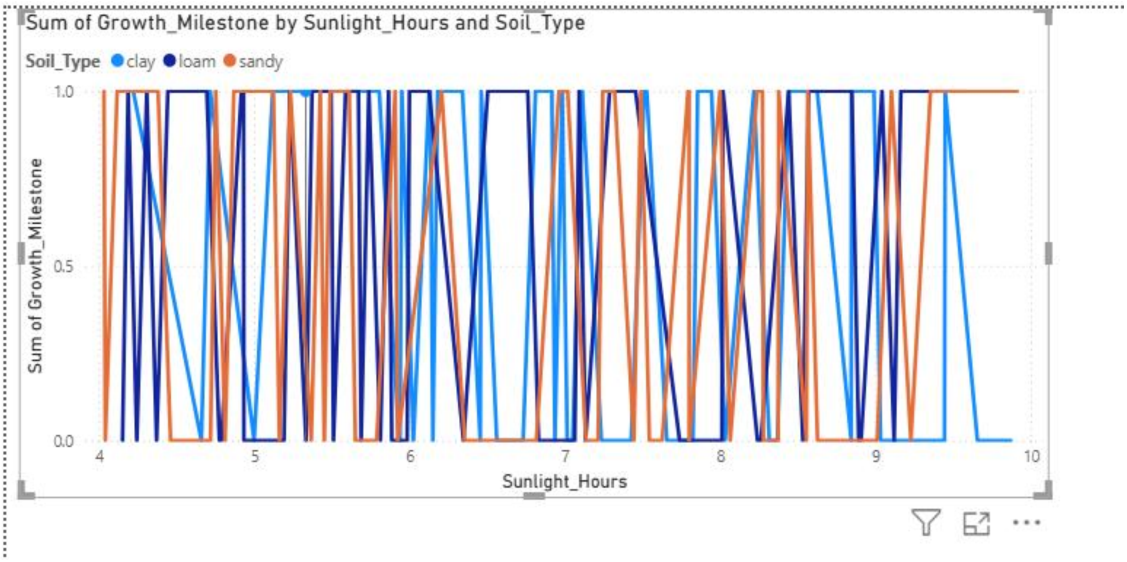


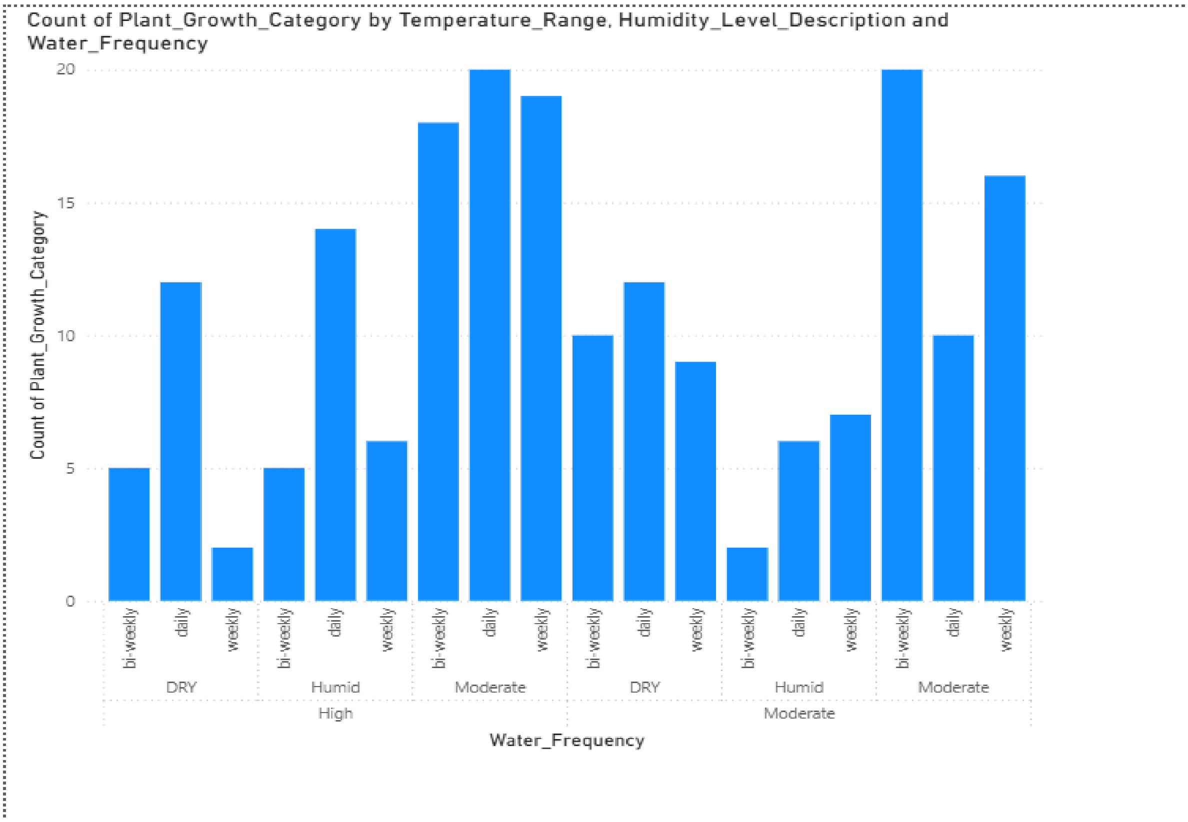


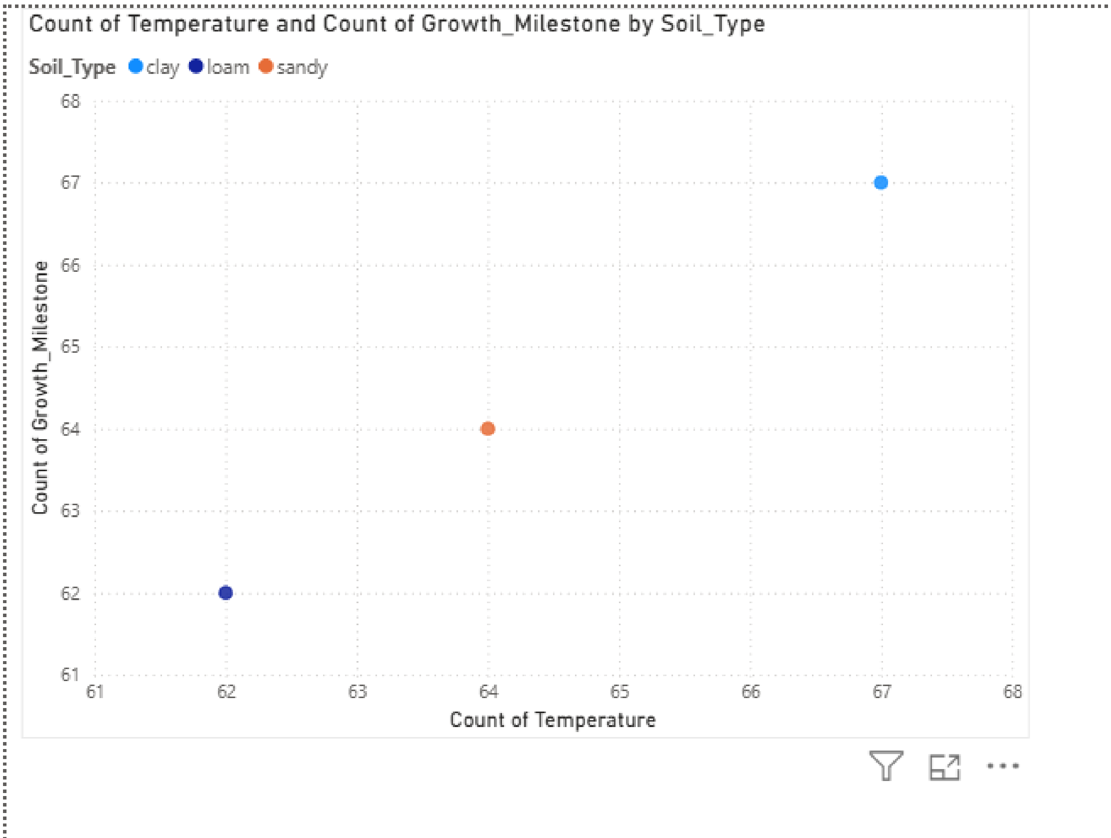



Count of Humidity_Level_Description by Soil_Type
67
66
65
64
63
62
Count of Humidity_Level_Description
clay
sandy
loam
Soil_Type


Sum of Growth_Milestone by Soil_Type and Fertilizer_Type
Fertilizer_Type chemical none organic
loam
sandy
clay
Soil_Type
0
5
10
15
20
25
30
35
Sum of Growth_Milestone

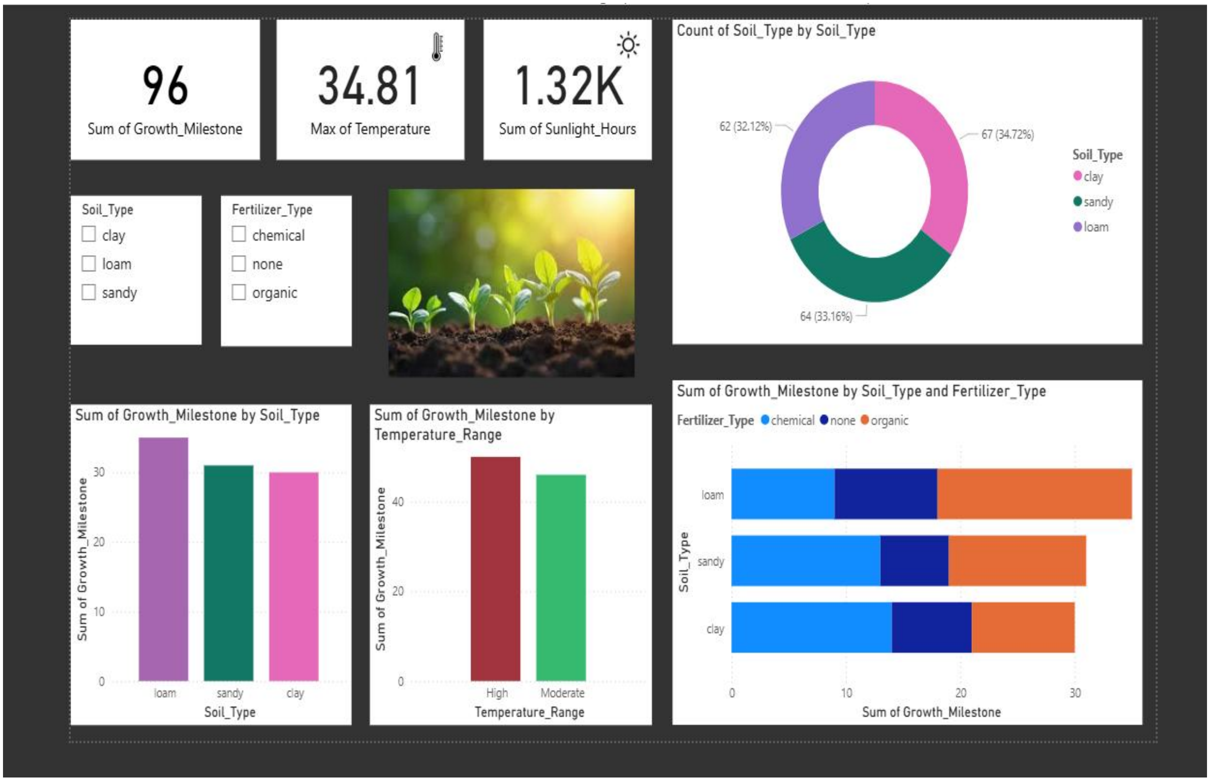

Sum of Growth_Milestone by Sunlight_Hours and Soil_Type
Soil_Type clay loam sandy
1.0
0.5
0.0
Sum of Growth_Milestone
4
5
6
7
8
9
10
Sunlight_Hours


Count of Plant_Growth_Category by Temperature_Range, Humidity_Level_Description and Water_Frequency
Count of Plant_Growth_Category
20
15
10
5
0
bi-weekly
daily
weekly
DRY
Humid
Moderate
High
Moderate
Water_Frequency


Count of Temperature and Count of Growth_Milestone by Soil_Type
Soil_Type clay loam sandy
Count of Growth_Milestone
68
67
66
65
64
63
62
61
Count of Temperature

5. Dashboard

5.1. Dashboard Design File



Major Insights from Dashboard:

1. Total Growth Milestone across all data points is 96 units.
2. Loam soil achieved the highest total growth milestone, followed closely by sandy and clay soils.
3. Maximum recorded temperature is 34.81°C.
4. Sunlight exposure totals 1.32K hours, indicating strong light availability as a growth factor.
5. Growth is higher under “High” temperature ranges compared to “Moderate”.
6. Organic fertilizer appears most effective, contributing the largest share of growth across all soil types.
7. Chemical fertilizer also contributes significantly but not as much as organic.

6. Report

6.1. Story Design File

This dashboard provides valuable insights into how environmental and agricultural management factors influence plant growth milestones. Such analysis is crucial for making informed agronomic decisions.

1. Trends Over Time:

- The dashboard enables analysis of how plant growth milestones vary with sunlight hours over different soil types, helping identify time-based growth patterns.
- A dense trend chart showcases fluctuations in growth metrics based on varying sunlight exposure, offering clues into optimal daily light conditions.

2. Performance Comparisons:

Bar charts allow comparison of growth milestones across soil types (loam, sandy, clay), with loam showing the highest total growth, suggesting its favourable role in plant development

Comparison of growth outcomes under different water frequencies (daily, bi-weekly, weekly) reveals which irrigation strategy performs better under varying environmental conditions.

3. Environment Segmentation:

- Environmental factors like temperature range, humidity level, and water frequency are segmented to assess their individual and combined impacts on different plant growth categories.
- For instance, high humidity combined with weekly watering appears to support higher growth in moderate temperature ranges.

- Metrics like Sum of Growth Milestones (96) and Sum of Sunlight Hours (1.32K) help gauge the input-output ratio and evaluate how efficiently environmental resources are converted into growth milestones.

4.Resource Optimization:

- Metrics like Sum of Growth Milestones (96) and Sum of Sunlight Hours (1.32K) help gauge the input-output ratio and evaluate how efficiently environmental resources are converted into growth milestones.
- These KPIs can guide better fertilizer or irrigation planning.

Example:

1. Soil Type Distribution:

- The pie chart indicates that clay soil is the most prevalent type in the dataset (34.72%), followed by sandy (33.16%) and loam (32.12%). Despite loam being the least prevalent, it shows the highest sum of growth milestones — a potential insight for soil management strategy.

2. Growth Based on Watering Frequency:

- In moderate temperature and humidity, plants watered weekly consistently reach higher growth categories compared to those watered daily or bi-weekly. This suggests potential over-watering or under-watering risks.

3. Sunlight Dependency:

- The line chart indicates that growth performance is closely linked with sunlight hours, though the response differs across soil types — loam responds more consistently to higher sunlight exposure.

7. Performance Testing

7.1 Utilization of Data filters

Growth_Milestone_Count = COUNTROWS(FILTER(plant_growth_data, plant_growth_data[Growth_Milestone]=1))

Growth_Milestone_Percentage = DIVIDE([Growth_Milestone_Count], COUNTROWS(plant_growth_data), 0)

Humidity_Level_Description = 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")

Growth_Milestone_Description = SWITCH([Growth_Milestone], 0, "Early Stage", 1, "Mature Stage", "Unknown Stage")

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

Plant_Growth_Category = SWITCH([Growth_Milestone], 0, "Initial Growth", 1, "Advance Growth", "Uncategorized")

Temperature_Range = SWITCH(TRUE(), [Temperature]<15, "Low", [Temperature]>=15 && [Temperature]<25,"Moderate", [Temperature]>=25,"High")

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")

Water_Frequency_Numeric = SWITCH([Water_Frequency],"daily", 1,"bi-weekly", 2,"weekly", 3,BLANK())

7.2 No of Calculation Field

Average_Humidity = AVERAGE(plant_growth_data[Humidity])

Average_Sunlight_hours = AVERAGE(plant_growth_data[Sunlight_Hours])

Average_Temperature = AVERAGE(plant_growth_data[Temperature])

Growth_Milestone_Percentage = DIVIDE([Growth_Milestone_Count], COUNTROWS(plant_growth_data), 0)

Water_Frequency_Total = SUM(plant_growth_data[Water_Frequency_Numeric])

7.3 No of Visualization

Stacked Colum Chart
Donut Chart
Stacked Bar Chart
Line Chart
Scatter Chart

8.Conclusion/Observation

This project leverages the power of data analytics and visualization through Power BI to unlock actionable insights into plant growth dynamics. By analyzing key environmental and agricultural variables, it identifies the critical factors that influence different growth stages and offers evidence-based recommendations for optimal growing conditions. The outcomes of this analysis are expected to empower farmers and agricultural planners with practical, data-driven strategies—enhancing crop productivity, optimizing resource usage, and contributing to more sustainable farming practices. Ultimately, this project lays the foundation for smarter agriculture through informed decision-making and technological integration.

9. Future Scope

- **Integration with Real-Time IoT Sensors:**
Incorporating real-time environmental data from IoT-based soil and weather sensors can further improve the accuracy of growth predictions and allow for dynamic, real-time decision-making.

- **Expansion to Crop-Specific Models:**
Future work can focus on building customized models for specific crop types, enabling tailored recommendations based on unique growth requirements and regional factors.

- **Predictive Analytics and Forecasting:**
Advanced machine learning models can be developed to predict future growth trends and potential yield outcomes based on current and historical data.

10. GitHub & Project Demo Link

https://drive.google.com/file/d/1f6vw8x_XSCmzylM_SAsyYwOJv3XHxljZ/view?usp=drive_link

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