Predicting plant growth stages with environmental and management data using Power Bi

1. Introductiom

1.1 Project Overview

This project was created during my virtual internship to learn and apply Power BI for data analysis. The project is based on analyzing a comprehensive dataset containing key environmental and management factors such as soil type, sunlight hours, water frequency, fertilizer type, temperature, and humidity. By leveraging this data, the company aims to predict the growth milestones of plants, which are crucial for understanding the conditions that promote optimal growth. This project will involve the creation of interactive dashboards and predictive models to uncover patterns and insights that can inform and improve agricultural practices and greenhouse management.

1.2 Objectives

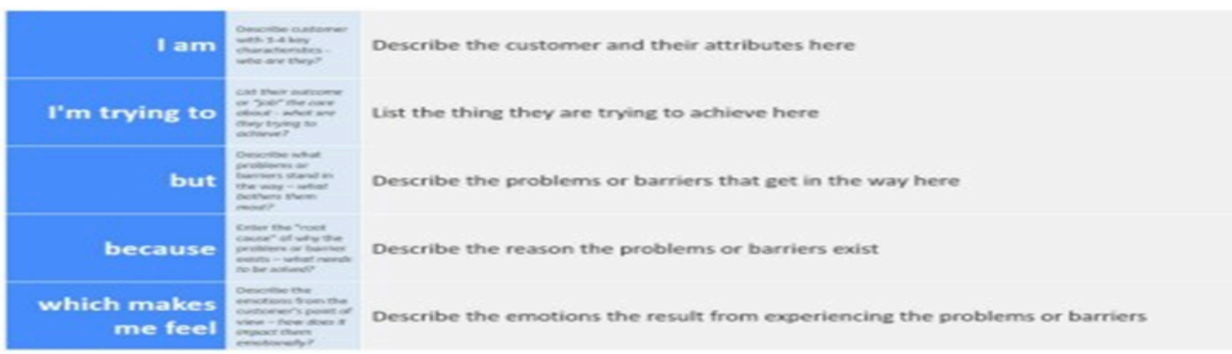
- To clean and prepare the plant_growth_data dataset for analysis.
- To understand trends in plants growth according to Fertilizer type, Soil type, Temperature, Sunlight hours and Humidity.
- To create charts and maps using Power BI to show insights from the data.
- To create charts and maps using Power BI to show insights from the data.
- To practice using Power BI tools like Power Query, DAX, and visuals.

2. Project Initialization and Planning Phase

2.1Define Problem Statement

Crop Growth is influenced by environmental conditions such as temperature, soil moisture, and sunlight, as well as management factors like irrigation and fertilization. Predicting plant growth stages is essential for optimizing resource allocation, reducing crop loss, and increasing yields. However, traditional plant growth prediction models rely on static datasets and lack real-time adaptability. Power BI can be used to create interactive dashboards that integrate environmental data, visualize growth patterns, and predict optimal farming decisions.

The Customer Problem Statement template helps us focus on what matters to create experiences people will love. A well-articulated customer problem statement allows us and our team to find the ideal solution for our customers' challenges. Throughout the process, we'll also be able to empathize with our customers, which helps us better understand how they perceive our product or service.



Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	A greenhouse manager at ABC Greenhouses responsible for ensuring consistent plant growth across locations.	Achieve uniform and optimal plant growth by standardizing successful growing	I face inconsistent plant growth due to varying soil types, sunlight exposure, and watering schedules.	The lack of data-driven insights has made it difficult to identify the best combination of growing conditions.	Frustrated and concerned about unpredictable yields and the overall health and productivity of the plants.
PS-2	An organic crop manager at GreenEarth Farms focused on delivering consistent yields across different farm plots.	Achieve reliable and optimal crop growth by identifying and applying the best organic farming	I encounter inconsistent crop growth rates due to unclear combinations of soil, fertilizer, watering, and climate.	There hasn't been a structured, data-driven approach to pinpoint the precise conditions that maximize growth.	Concerned and uncertain about my ability to provide consistent organic yields, affecting farm sustainability.

2.2 Project Proposal

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	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI
Scope	Enhanced prediction of plant growth stages, Improved decision making via interactive analytics and visualizations
Problem Statement	

Description	Crop Growth is influenced by environmental conditions such as temperature, soil moisture, and sunlight, as well as management factors like irrigation and fertilization. Predicting plant growth stages is essential for optimizing resource allocation, reducing crop loss, and increasing yields. However, traditional plant growth prediction models rely on static datasets and lack real-time adaptability. Power BI can be used to create interactive dashboards that integrate environmental data, visualize growth patterns, and predict optimal farming decisions.
Impact	**Social:** Enhancing plant growth prediction helps farmers optimize resources, increase crop yields, and reduce environmental impact, supporting sustainable agriculture. **Business:** This solution can generate revenue through precision agriculture consulting, AI-powered farm management systems and partnerships with agritech startups.
Proposed Solution	
Approach	Data Collection & Integration, Data Preparation & Modeling, Predictive Modeling, Integration with Power BI, Interactive Dashboards & Decision Support, Iterative Improvement
Key Features	It replaces static, inflexible models with a data-driven, real-time, and interactive system that can adapt as conditions change, provide evidence-based recommendations, and continually improve with data. Power BI acts as the hub for visualization, decision support, and stakeholder engagement throughout the agricultural growing season

Resource Requirements

Software		
Frameworks	Python frameworks	Basic Python
Libraries	Additional libraries	NIL
Development Environment	IDE, version control	Microsoft Power BI, Excel
Data		

Data	Source, size, format	https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification, 193rows and 7 columns, csv

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	- Minimum CPU: 1.4 GHz x64 processor (dual-core minimum). Recommended: 2.0 GHz or faster multi-core processor. - GPU: Not strictly required but modern GPU helps with rendering visuals smoothly.
Memory	RAM specifications	- Minimum: 1-2 GB RAM for basic tasks. - Recommended: 4 GB minimum, with 16 GB or more preferred for heavy data and complex reports.
Storage	Disk space for data, models, and logs	- Minimum disk space: 1-5 GB free space for installation and data storage. - Recommend SSD for faster load and query performance. - Large datasets/models require more storage accordingly.

2.3 Initial Project Planning

Product Backlog, Sprint Schedule, and Estimation

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

Sprint	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Members**	**Sprint Start Date**	**Sprint End Date**
Sprint-1	Data Collection & Extraction	PPGSWEMD-2	Downloading The Dataset	1	High	Abhinav Tomar	23 July 2025	30 July 2025
Sprint-1	Prepare The Data for Visualization	PPGSWEMD-4	Prepare The Data for Visualizatio	3	High	Abhinav Tomar	23 July 2025	30 July 2025
Sprint-1	Data Visualization	PPGSWEMD-6	Plant Growth Classification	5	High	Abhinav Tomar	23 July 2025	30 July 2025
Sprint-2	Dashboard	PPGSWEMD-8	Responsive And Design Of Dashboard	3	Medium	Abhinav Tomar	31 July 2025	7 August 2025
Sprint-2	Design Of Report	PPGSWEMD-10	Report	3	High	Abhinav Tomar	31 July 2025	7 August 2025
Sprint-2	Performance Testing	PPGSWEMD-12	Utilization Of DAX Expressions	3	Medium	Abhinav Tomar	31 July 2025	7 August 2025
Sprint-2		PPGSWEMD-13	No of Visualizations/Graphs	2	Low	Abhinav Tomar	31 July 2025	7 August 2025

3. Data Collection and Preprocessing Phase

3.1 Data Collection Plan and Raw Data Sources Identified

Elevate our data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan

Section	Description
Project Overview	The project titled "Predicting Plant Growth Stages with Environmental and Management Data Using Power BI" focuses on leveraging data analytics and
Sources	Kaggle.
Raw Data Sources Identified	The "plant_growth_data.csv" dataset is designed to support the prediction and analysis of plant growth stages based on a combination of environmental and management factors. The data helps in understanding how factors like soil, sunlight, watering, fertilizer, temperature, and humidity interact to influence plant growth.

Raw Data Sources

Source Name	Description	Location/URL	Format	Size	Access Permissions

plant_growth_data.csv	The "plant_growth_data.csv " dataset contains environmental and management factors influencing plant growth, such as soil type, sunlight hours, watering frequency, fertilizer type, temperature, and humidity .	https://www.kaggle.com/datasets/gororororo23/plant-growth-data-classification	CSV	11.49 KB	Public

3.2 Data Quality Report

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
plant_growth_data.csv	No missing values in any column	Low	No action needed; dataset has complete records for all columns.
plant_growth_data.csv	Categorical value consistency	Low	All categorical fields ('Soil_Type', 'Water_Frequency',

			'Fertilizer_Type') contain only expected unique values. No issue.
plant_growth_data.csv	Data granularity/representativeness	Moderate	Ensure enough diverse records for each combination of environmental/management factors; if not, augment dataset or collect more samples.
plant_growth_data.csv	Value ranges for environmental features	High	Double-check that all values for sunlight hours, humidity, and temperature are within realistic ranges for actual agricultural scenarios.

Data Exploration and Preprocessing

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	**Description**
Data Overview	The "Plant Growth Data Classification" dataset is designed to support the prediction and analysis of plant growth stages based on a combination of environmental and management factors. Each row in the dataset represents the condition of a particular plant at a point in time, along with the observed growth milestone it has reached. The data helps in understanding how factors like soil, sunlight, watering, fertilizer, temperature, and humidity interact to influence plant growth.
Data Cleaning	Handle missing values, duplicates, and correct errors.
Data Transformation	We used **Power Query** in Power BI for all preprocessing steps. Here's what we did:

- **New Columns Created:**

Water_Frequency_Numeric: A calculated column added to classify water frequency into numerics:

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

Temperature Range: A calculated column added to classify the temperature into different ranges:

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

Humidity Range: A calculated column added to classify the humidity into different ranges:

```
Humidity_Range = SWITCH(TRUE(),[Humidity]<15,"Low",[Humidity]>=15 && [Humidity]<25,"Moderate",[Humidity]>=25,"High")
```

Humidity_Level_Description: A calculated column added to classify the humidity into different humid levels:

```
Humidity_Level_Decsription = 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")
```

Temperature_Range_Description: A calculated column added to classify temperature into different temperature levels:

```
Temperatur_Range_Decsription = SWITCH(TRUE(),[Temperature]<10,"Very Cold",[Temperature]>=10 && [Temperature]<20,"Cold",[Temperature]>=20 && [Temperature]<30,"Moderate",
[Temperature]>=30 && [Temperature]<40,"Warm",[Temperature]>=40,"Hot")
```

Growth_Milestone_Description: A calculated column added to classify Growth_Milestone into different levels:

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

Plant_Growth_Category: A calculated column added to classify Growth_Milestone into plant growth category:

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

<table>
<tr><td></td><td>

- **New Measures Created:**

Average_Humidity: Calculate a measure for average of humidity.

```
Average_Humidity = AVERAGE(plant_growth_data[Humidity])
```

Average_Sunlight_Hours: Calculate a measure for average of Sunlight_Hours.

```
Average_Sunlight_Hours = AVERAGE(plant_growth_data[Sunlight_Hours])
```

Average_Temperature: Calculate a measure for average of Temperature.

```
Average_Temperature = AVERAGE(plant_growth_data[Temperature])
```

Growth_Milestone_Count: Calculate the rows that have Growth_Milestone=1.

-

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

Growth_Milestone_Percentage: Calaculate the percentage of Growth_Milestone_Count from all rows.

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

</td></tr>
<tr><td>Data Type Conversion</td><td>There is no need of data type conversion as all are already assigned correct datatype.</td></tr>
<tr><td>Column Splitting and Merging</td><td>There is no split or merge used in dataset.</td></tr>
<tr><td>Save Processed Data</td><td>Save the cleaned and processed data for future use.</td></tr>
</table>

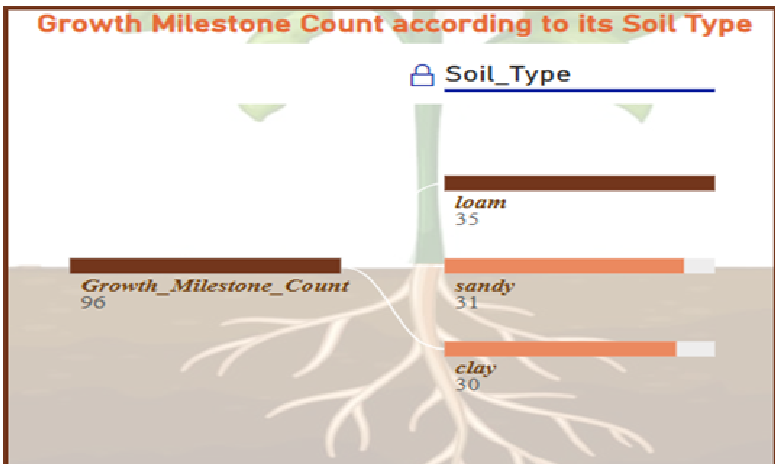
4. Data Visualization

Visualization development refers to the process of creating graphical representations of data to facilitate understanding, analysis, and decision-making. The goal is to transform complex datasets into visual formats that are easy to interpret, make informed decisions.

Business Questions and Visualisation

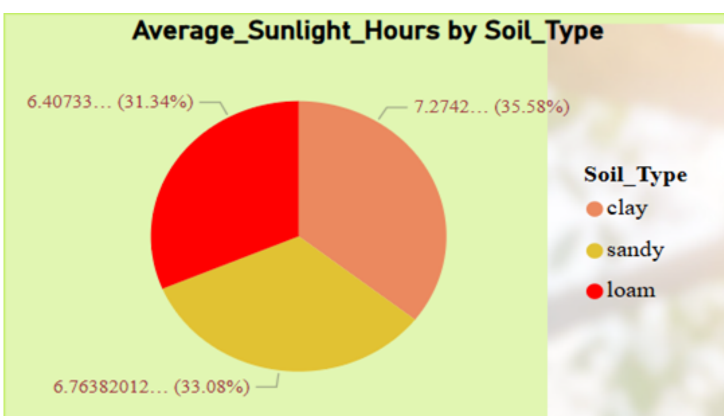
Which Soil Type contributes to the Highest Plant Growth Milestones?

Visualization: Bar chart showing "Growth Milestone Count according to its soil type".



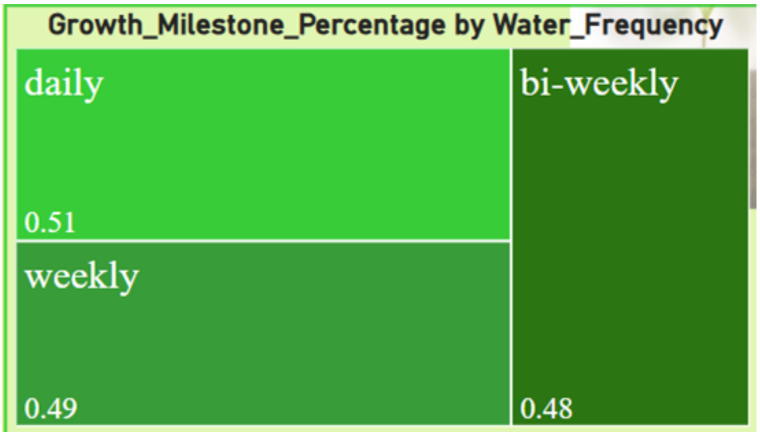
What is the distribution of average sunlight hours for each soil type?

Visualization: Pie chart entitled "Average_Sunlight_Hours by Soil_Type"



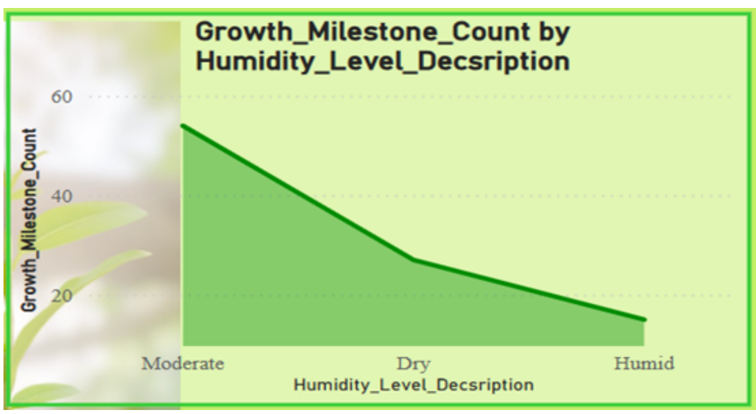
What watering frequency maximizes growth milestone achievement?

Visualization:Stacked square chart named "Growth_Milestone_Percentage by Water_Frequency".



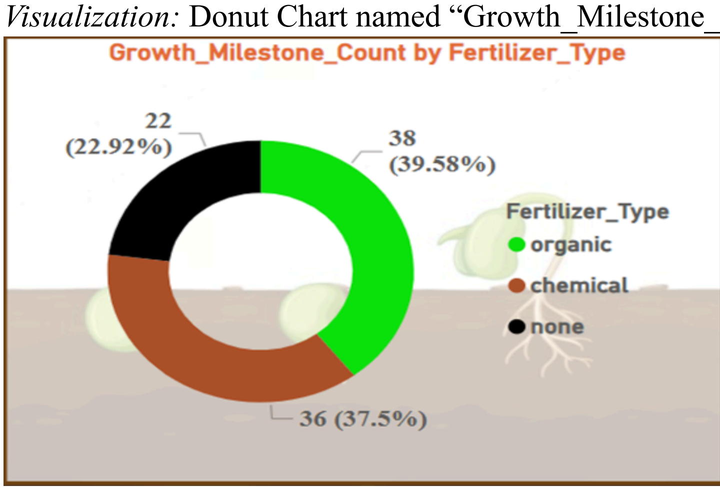
What Humidity ranges optimize plant growth?

- *Visualization:* Line Chart for Growth_Milestone_Count by Humidity_Level_Description



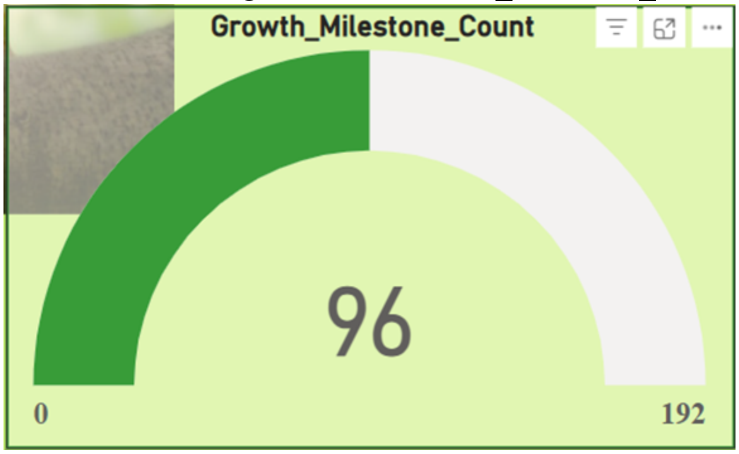
How does Fertilizer Type affect plant growth milestone?

- *Visualization:* Donut Chart named "Growth_Milestone_Count by Fertilizer_Type".



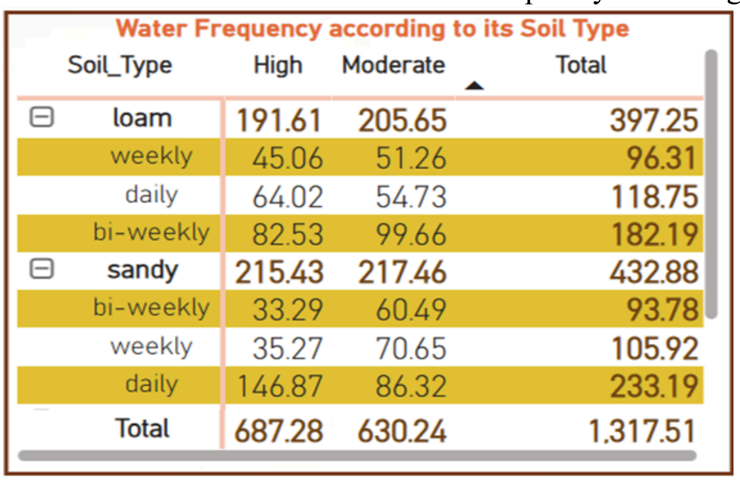
What is the overall growth milestone count across the description?

- *Visualization:* Gauge entitled "Growth_Milestone_Count"



Which combination of water frequency and soil type leads to the highest aggregated values?

- *Visualization:* Matrix named "Water frequency according to its Soil Type.

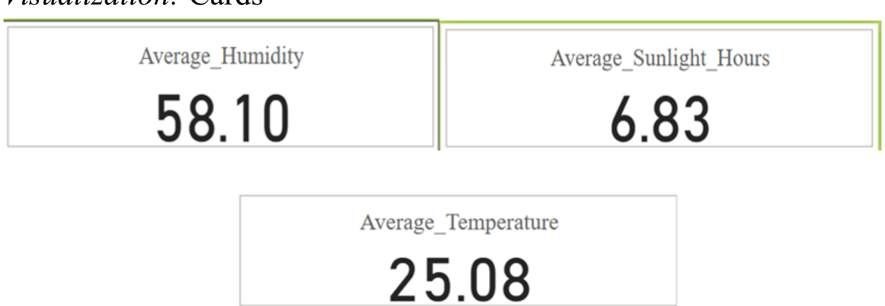


Water Frequency according to its Soil Type

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
Total	687.28	630.24	1,317.51

What is average temperature, average humidity and average sunlight hours?

- *Visualization:* Cards



5. Dashboard

Creating an effective dashboard involves thoughtful design to ensure that the presented information is clear, relevant, and easily understandable for the intended audience. Here are some key principles and best practices for dashboard design

Activity 1: Interactive and visually appealing dashboards

Creating interactive and visually appealing dashboards involves a combination of thoughtful design, effective use of visual elements, and the incorporation of interactive features. Here are some tips to help you design dashboards that are both visually appealing and engaging for users so take care of below points

Clear and Intuitive Layout

Use Appropriate Visualizations

Colour and Theming

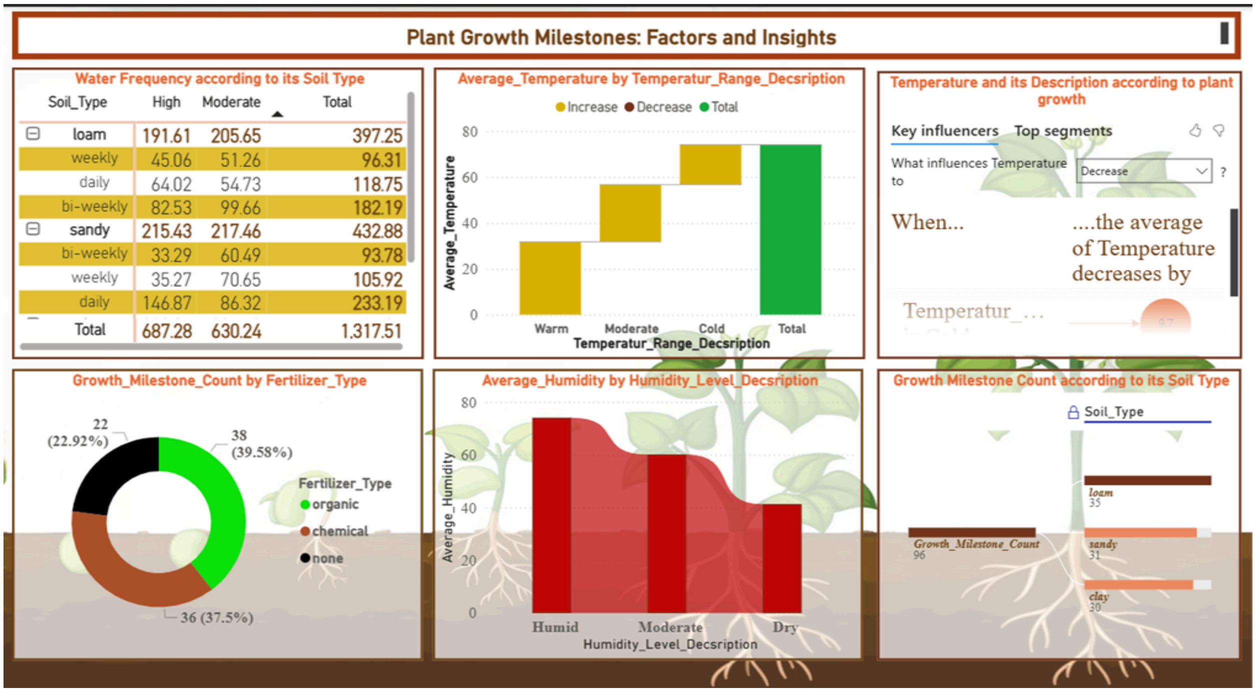
Interactive Filters and Slicers

Drill-Down Capabilities

Responsive Design

Custom Visuals and Icons

Use of Infographics



Note: Highlight the major outcomes in form of bullet points

Here are Five potential outcomes from the dashboard image provided:

Loam Soil achieves the most growth milestone: The "Growth Milestone Count according to its Soil Type" bar chart shows loam soil (96 milestones) far surpasses sandy (31) and clay (30) soil.

Organic Fertilizer Outperforms Others: The donut chart reveals the highest growth milestone count for organic fertilizer (38), followed by chemical (36), and none (22).

Daily Watering Promotes Vigorous Growth: In the "Water Frequency according to its Soil Type" table, daily watering has consistently high totals, especially with loam soil (233.19 for sandy, 233.19 for loam).

Moderate Humidity Optimizes Growth: Maintaining moderate humidity (not too dry, not too humid) is optimal for plant growth **.**

Sandy Soil Uses the Most Water: In "Water Frequency according to its Soil Type," sandy soil requires more frequent and higher total watering than loam.

6. Report

A report is a comprehensive document that provides a detailed and structured account of data analysis, findings, and insights. It is typically used for in-depth analysis, documentation, and communication of results. Reports are suitable for a diverse audience, including decision-makers, analysts, and stakeholders who need a comprehensive understanding of the data.

Designing a report in Power BI involves connecting to data sources, creating visualizations like charts and graphs, customizing their appearance and interactivity, organizing them logically on the canvas, formatting elements for consistency and clarity, and optionally creating dashboards for a summarized view. Throughout the process, it's essential to consider the audience's needs and ensure the report effectively communicates insights from the data. Finally, iterate based on feedback to continually improve the report's design and usefulness.

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

1. Trends Analysis:

- Power BI time series and trend visualizations show changes over time (e.g., growth milestone counts across months or seasons), helping to anticipate future patterns or cycles.

2. Performance Benchmarking

- Reports visualize how different segments perform (e.g., loam vs. sandy soil), enabling you to benchmark locations, teams, or products against each other.

3. Anamaly Detection:

- Outliers or sudden changes in performance (such as unexpected drops in growth stages) are quickly identified, allowing fast investigation and response.

4. Continuous Improvement:

- Regular observations provide a feedback loop, so you can track the impact of interventions and refine strategies over time.

Example:

  - **Soil Type Dominance**:

- Loam soil leads all categories with 96 growth milestones, substantially outperforming sandy(31) and clay(30) soils.

  - **Water Frequency Effectiveness**:

- Daily watering achieves a slightly higher growth milestone percentage(0.51) compared to weekly(0.49) and biweekly(0.48) schedules.

  - **Sunlight Exposure Patterns:**

- Sandy soil receives the highest average sunlight hours(7.27), but loamy soil achieves more growth milestone even with slightly less sunlight.

  - **Overall Growth Achievement**:

- The total number of achieved growth milestone in the dataset is 96, indicating strong effectiveness of the optimal conditions identified.

7. Performance Testing

7.1 Utilization of Data Filters

Here are some key elements for the effective utilization of data filters, based on the provided dashboard visuals:

- **Segmented Analysis:** Data filters allow users to break down and analyze plant growth milestones by specific factors such as soil type, humidity level, water frequency, and fertilizer type. This enables targeted insights rather than generalized conclusions

- **Comparison Across Categories:** Data filters allow users to break down and analyze plant growth milestones by specific factors such as soil type, humidity level, water frequency, and fertilizer type. This enables targeted insights rather than generalized conclusions.

- **Enhanced Visualization:** Applying filters dynamically updates the dashboard elements (like bar charts, pie charts, and KPIs), helping users visualize changes in real time based on selected criteria.

- **Insight Generation:** Filters help highlight key influencers and outliers (such as the soil or humidity levels that produce the highest growth), as presented in the summarized reports on the dashboards.

- **User-Driven Exploration:** With filters in place, users can explore 'what-if' scenarios, helping them understand the impact of environmental or management changes (e.g., switching from weekly to daily watering) on plant growth outcomes

- **Quick Identification of Trends:** By narrowing down the dataset, filters make it easier to identify patterns and trends (e.g., which combination of factors consistently yields higher growth milestones).

Effectively used, data filters provide clarity, customization, and actionable insights for data-driven decision-making in complex, multi-factor environments like plant growth analysis.

7.2 Number of Calculation Fields

There are 5 calculation fields(Measure)

Average_Humidity: Calculate a measure for average of humidity.
Average_Sunlight_Hours: Calculate a measure for average of Sunlight_Hours.
Average_Temperature: Calculate a measure for average of Temperature.
Growth_Milestone_Count: Calculate the rows that have Growth_Milestone=1.
Growth_Milestone_Percentage: Calaculate the percentage of Growth_Milestone_Count from all rows.

7.3 Number of Visualizations

From the images provided:

- **Dashboard Page :**
    - **Matrix:** Water frequency According to Its Soil Type
    - **Waterfall Chart:** Average_Temperature by Temperature_Level_Description
    - **Key Influencers:** Temperature and its description according to plant growth
    - **Donut Chart:** Growth_Milestone_Count by Fertilizer_Type.
    - **Ribbon Chart:** Average_Humidity by Humidity_Level_Description.
    - **Decomposition Tree:** Growth_Milestone_Count according to its Soil_Type.
  - **Total = 6 visual elements**

- **Report Page :**
    - **Pie Chart:** Average_Sunlight_Hours by Soil_Type
    - **TreeMap:** Growth_Milestone_Percentage by Water_Frequemcy,
    - **LineChart:** Growth_Milestone_Count by Humidity_Level_Description.
    - **Gauge:** Growth_Milestone_Count.
    - **Card:** Average_Sunlight_Hours, Average_Humidity, Average_Temperature
  - **Total = 5 visual elements**

Overall Total Visuals: 11

The visual load is optimized to ensure that performance remains fast and interactive for end users.

8. Conclusion / Observation

The Power BI project analyzing plant growth according to different soil type ,fertilizer type, temperature, humidity, sunlight hours and water frequency effectively demonstrates how data can be transformed into meaningful insights through interactive dashboards and reports.

Key Observations:

- The dataset was efficiently cleaned and structured, enabling smooth visual representation.
- A variety of visualization types (Map, Clustered Column chart, Pie chart, Cards, Area charts) provided diverse perspectives on the plant growth.
- Slicers and filters enhanced user interactivity, allowing users to focus on specific soil types and fertilizers.
- The report page emphasizes , while the dashboard focuses on **summary insights and quick metrics**.
- Performance was maintained with a limited number of calculated fields and optimized visuals.

Conclusion:

The project clearly demonstrates that moderates humidity, daily watering, sandy soil, and optimal temperature are the main contributors to maximizing plant growth milestones. Over-watering, excessive humidity, or extreme temperatures negatively affect growth. Organic fertilizers show a positive trend, supporting sustainable practices. These insights provide valuable guidelines for agricultural planning and effective plant management.

9. Future Scope

The current Power BI project provides a strong foundation for analyzing plant growth data; however, there is significant potential to expand and enhance the analysis in the future. Some possible future developments include:

- **Advanced Predictive Modeling:** Implement machine learning models to predict plant growth milestones under different environmental scenarios. This will make the system more adaptive, helping farmers and gardeners preemptively address sub-optimal conditions.

- **Decision Support Tools:** Develop user-friendly dashboards or mobile applications for farmers and gardeners, offering personalized care schedules and alerts based on plant and soil parameters.

- **Sustainability and Environmental Impact Analysis:** Analyze the impact of different agricultural practices on soil health and environmental sustainability, including carbon footprint and water usage analysis, to support eco-friendly agricultural practices.

- **Integration with Weather Forecasting:** Combine growth insights with weather predictions to adjust agricultural routines in advance (e.g., optimizing watering before a rain event).

- **Expansion to Diverse Plant Species:** Extend analysis beyond current plant types to include a broader range of crops, flowers, or trees, allowing for more comprehensive agricultural recommendations across various plant categories.

- **Regional or Climate-Specific Analysis:** Customize the model for different geographical locations and climate zones. This would optimize recommendations by accounting for varying environmental conditions, ensuring relevance for diverse regions.

10. Appendix

10.1 Source Code

No custom source code was used in this Power BI project. The entire analysis, data transformation, and visualizations were created using built-in Power BI features such as Power Query, DAX, and the drag-and-drop visual interface.

10.2 GitHub & Project Demo Link

- **GitHub Repository:**https://github.com/AbhinavTomar1208/Predicting-Plant-Growth-Stages-With-Environmental-And-Management-Data-Using-Power-BI
- **Project Demo Video/Presentation:** https://drive.google.com/file/d/1SwJFDhzkm7Vsw7AFzjgOCiLyvAdy0kPe/view?usp=sharing