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

Team Details:

Team ID: LTVIP2025TMID19940

Team Leader: Seedrala Noharika Nag Sri

(21481A54A4) Mail id: harikaseedrala@gmail.com

Team member : Sure Venkata Subramanyasai Yaswanth

(21481A54A7) Mail id: sureyaswanth2004@gmail.com

Team member: Sadhu Madhava Rao (21481A54A3)

Mail id: madhavaraosadhu82@gmail.com

Technologies used:

Excel/CSV – Used as the dataset source.

Power Query – For data cleaning and transformation.

DAX (Data Analysis Expressions) – For creating calculated measures and aggregations.

Power BI – For data visualization and dashboard creation.

Introduction:

XYZ Company, renowned for its innovative approach in agriculture, is embarking on a project to optimize plant growth through advanced data analytics and visualization techniques using Power BI. The project focuses 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.

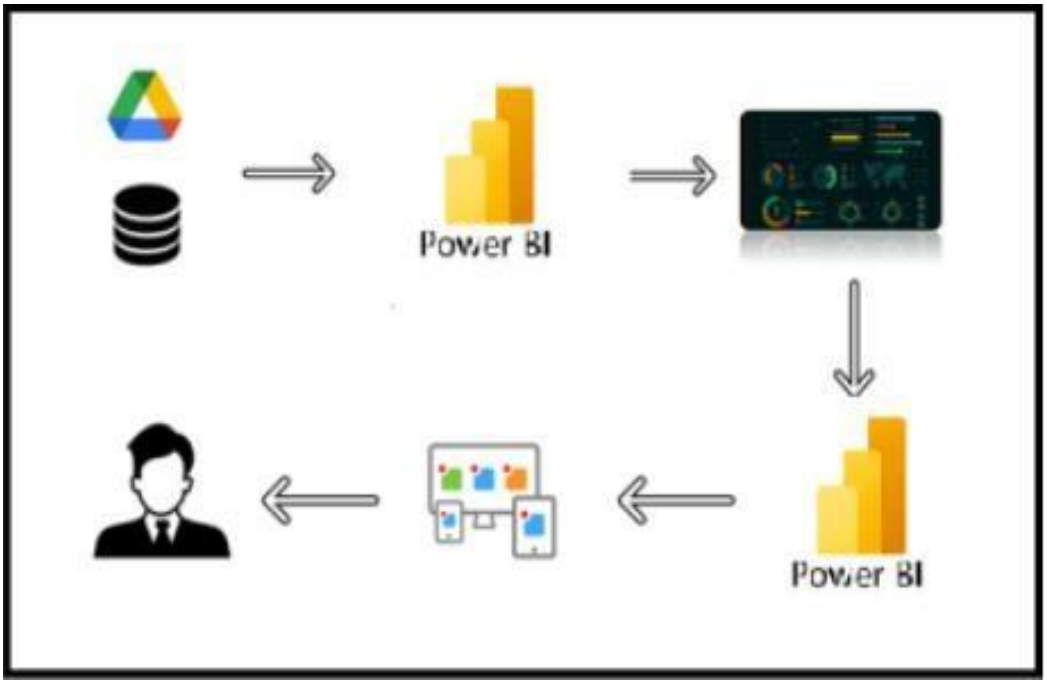
The analysis will be conducted using a decomposition tree to break down growth milestone counts by various factors, providing a clear view of the impact of each variable. Additionally, the project will include the development of several calculated columns and measures to enhance the dataset's analytical depth. Visualizations such as clustered bar charts, pie charts, scatter plots, and column charts will be utilized to present the findings effectively. By implementing this solution, XYZ Company aims to enhance crop yields, optimize resource allocation, and promote sustainable agricultural practices, ultimately solidifying its position as a leader in agricultural innovation.

Scenario 1 - ABC Greenhouses has been facing challenges with inconsistent plant growth across its different greenhouse locations. By leveraging Power BI, the company plans to identify the best combination of soil type, sunlight hours, and watering frequency that leads to the highest growth milestones. The decomposition tree will help break down growth milestone counts by these factors, revealing that loam soil combined with daily watering and 6-8 hours of sunlight yields the best results. This insight will enable ABC Greenhouses to standardize these conditions across all locations, improving overall plant health and productivity.

Scenario 2 Green Earth Farms has noticed varying growth rates in their organic crops and wants to ensure consistency in their yield. By analyzing the dataset, the company discovers that organic fertilizer combined with loam soil and bi-weekly watering leads to the most significant growth milestones. The decomposition tree further reveals that maintaining temperatures between 20-30°C and humidity levels between 50-70% optimizes plant growth. Green Earth Farms will use these insights to adjust their farming practices, ensuring their crops achieve the best possible growth under organic farming conditions.

Scenario 3 - Future Grow Tech has been developing smart farming solutions but needs to validate their technology's effectiveness under different conditions. By using Power BI to analyze the dataset, the company identifies that their smart sensors for monitoring soil moisture and adjusting water frequency in real-time significantly improve growth milestones. The decomposition tree analysis reveals that these sensors work best with sandy soil and weekly organic fertilizer application, under moderate temperature and humidity conditions. Future Grow Tech will integrate these findings into their product development, enhancing their technology to offer precise and effective agricultural solutions.

Technical Architecture:



Project Flow

To accomplish this, we have to complete all the activities listed below,

- Data Collection & Extraction from Database
    - Collect the dataset,
    - Storing Data in DB
    - Perform SQL Operations
    - Connect DB with Power Bi
- Data Preparation
    - Prepare the Data for Visualization
- Data Visualizations
    - No of Unique Visualizations
- Dashboard
    - Responsive and Design of Dashboard
- Report
    - Responsive and Design of Dashboard
- Performance Testing
    - No of Visualizations/ Graphs
- Project Demonstration & Documentation
    - Record explanation Video for project end to end solution
    - Project Documentation-Step by step project development procedure

Milestone 1: Data Collection & Extraction from Database

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

Activity 1: Collect the dataset

Please use the link to download the dataset: Link

Activity 1.1: Understand the data

Data contains all the meta information regarding the columns described in the CSV files. we have provided a CSV files:

1. Plant_Growth_Data_Classification

The dataset consists of environmental and management factors influencing plant growth, including:

- **Soil Type** (loam, sandy, clay)
- **Sunlight Hours**
- **Water Frequency** (daily, weekly, bi-weekly)
- **Fertilizer Type** (chemical, organic, none)
- **Temperature**
- **Humidity**
- **Growth Milestone** (indicating growth success)

Activity 2: Connect Data with Power BI

1. Open **Power BI Desktop**.
2. Click on **Home > Get Data**.
3. Choose **Excel/CSV/Database** (depending on the dataset format).
4. Browse and select the dataset file.
5. Click **Load/Transform** to import the data into Power BI for further processing.

Milestone 2: Data Preparation

1. **Data Cleaning:**
   - Checked for missing or inconsistent values.
   - Removed duplicates and handled null values where applicable.
2. **Created New Columns:**
   - `Water_Frequency_Numeric`: Converted categorical values (daily, weekly, etc.) into numerical format.
   - `Temperature Range`: Grouped temperature values into predefined categories.
   - `Humidity Range`: Categorized humidity levels.
   - `Humidity_Range_Description`: Provided descriptive labels for humidity ranges.

- `Temperature_Range_Description`: Descriptive labels for temperature ranges.
- `Growth_Milestone_Description`: Classified growth stages based on conditions.
- `Plant_Growth_Category`: Categorized plants based on growth success.

3. **Created Measures:**

- `Average_tempareture`: Calculates the mean temperature across the data
- `Average Humidity`: Computes the mean humidity level.
- `Average Sunlight`: Measures the average sunlight hours.
- `Growth_Milestone_Measure`: Aggregates growth milestones for analysis.
- `Growth_Milestone_Percentage`: Calculates the percentage of plants achieving growth milestones.
- `Growth_Milestone_Count`: Counts the total number of growth milestone occurrences.

4. **Data Transformation:**

- Applied Power Query transformations for better data structuring.

Ensured all columns were in the correct data type.

Milestone 4: Data Visualization

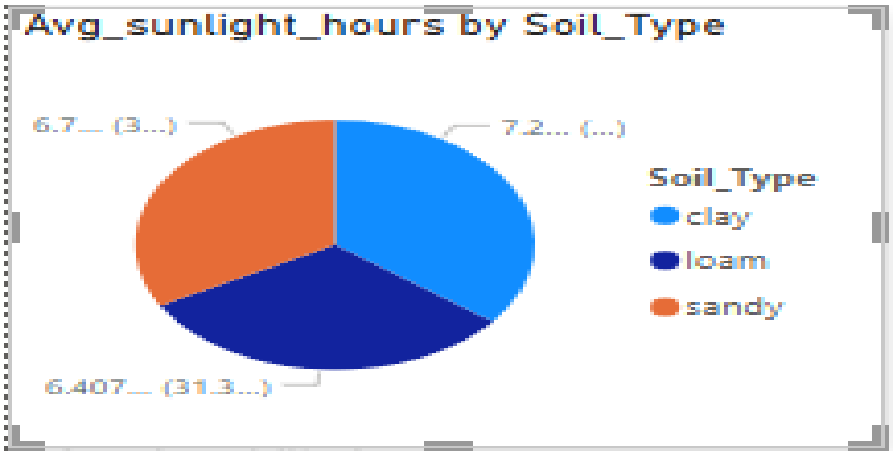
Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

Activity 1: No of Unique Visualizations

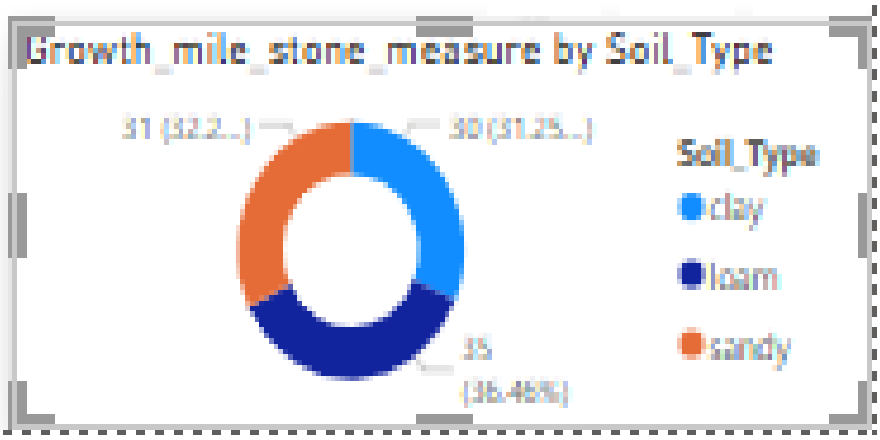
The number of unique visualizations that can be created with this dataset is extensive. Various types of visualizations help analyze plant growth performance and environmental impacts. These visualizations can be used to compare plant growth under different conditions, track growth trends over time, analyze distributions, and identify relationships between variables. Common visualizations include:

Visualizations Used:

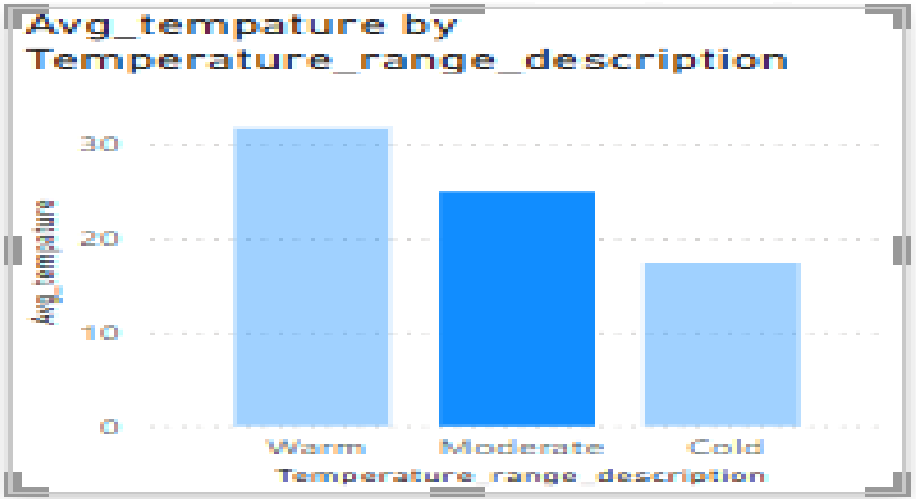
Pie Chart: To show the proportion of Average sunlight hours by soil type.



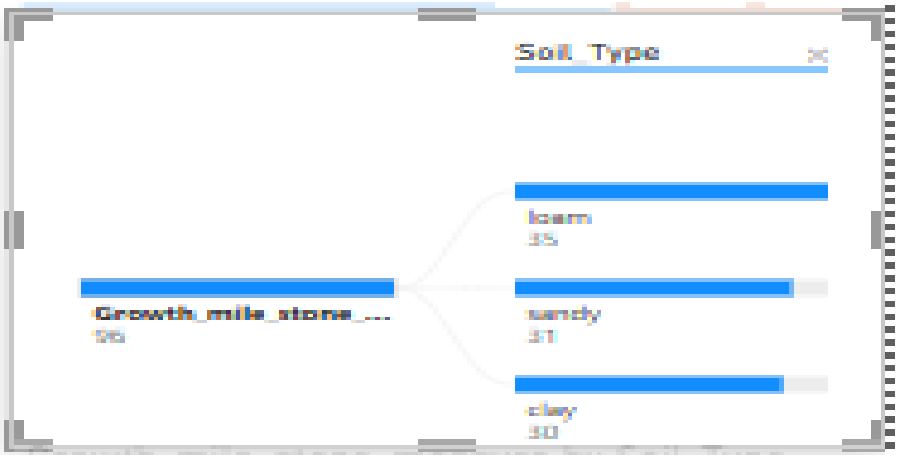
Donut Chart: To display Growth milestone measure by soil type.



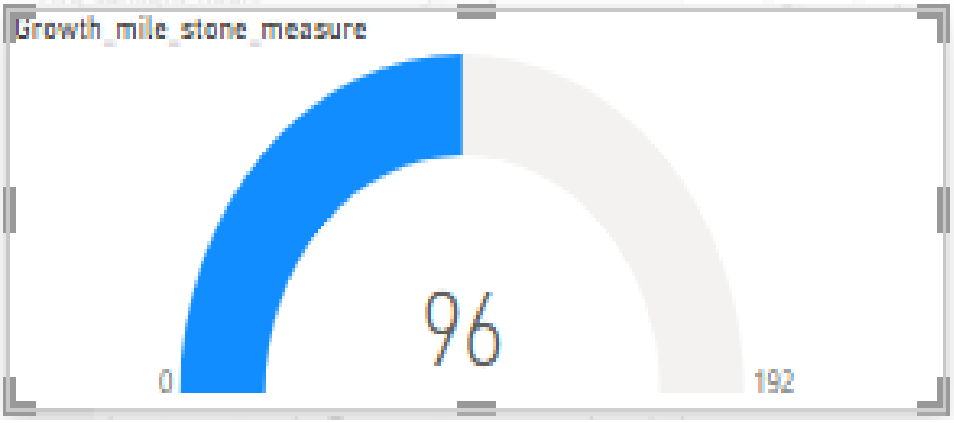
Stacked Column Chart: To compare Average temperature by Temperature range description.



Decomposition Tree: To analyze the growth milestone by soil type.



Gauge Chart: To monitor key performance of growth milestone measure(count).



Tree map: To show the distribution of plant growth milestone description by water frequency.

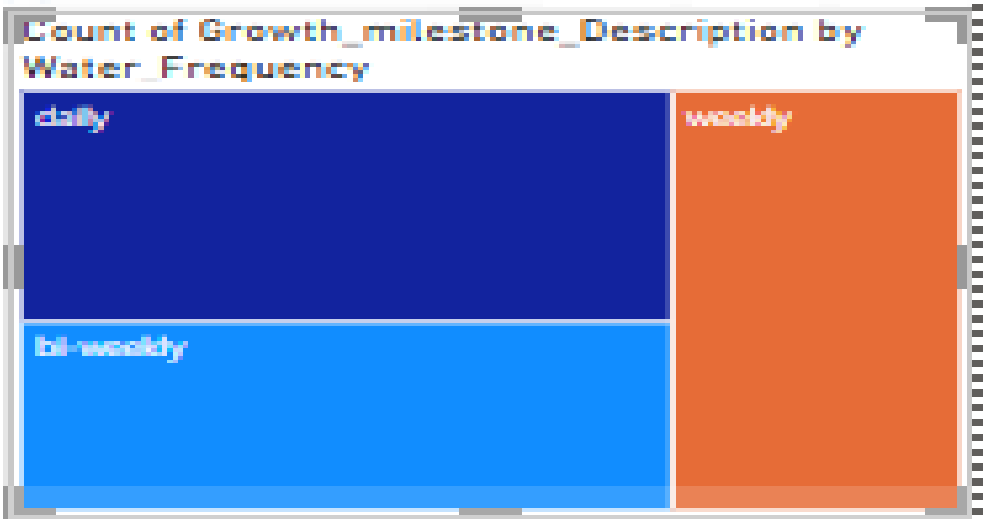
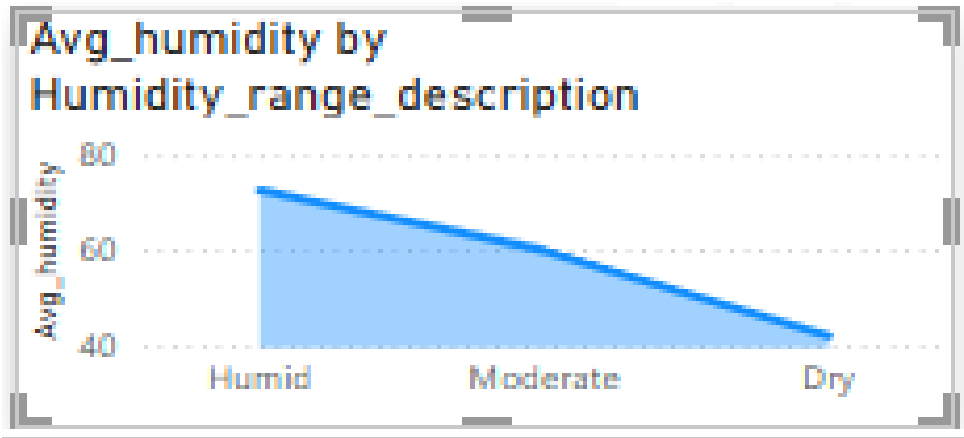


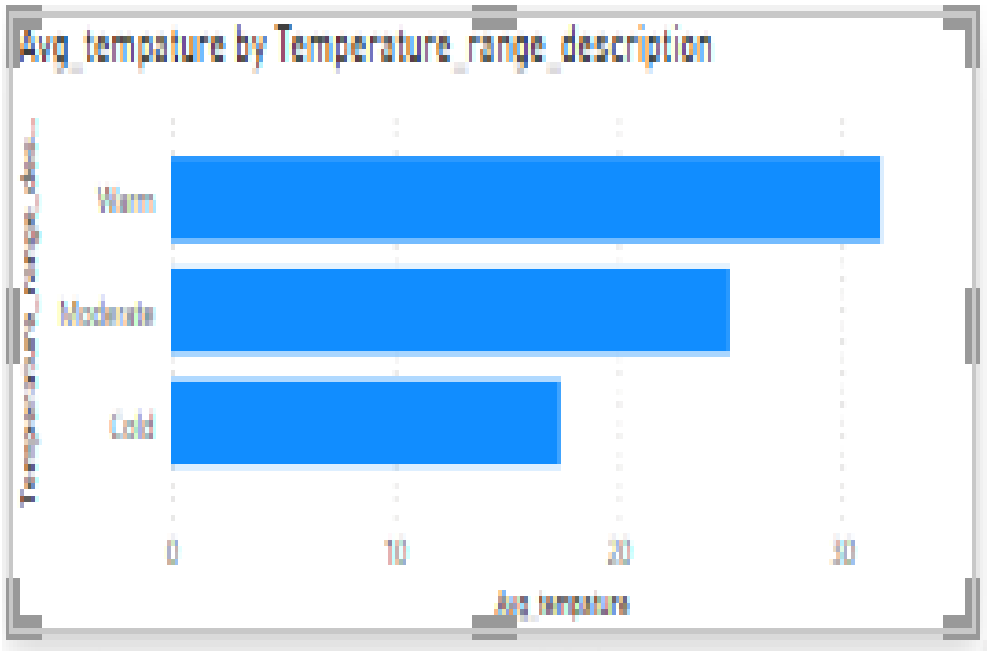
Table: To provide a detailed view of dataset values.

Humidity_range	Temperature_range	Avg_sunlight_hours	Water_Frequency
High	High	6.24	bi weekly
High	Moderate	7.30	bi weekly
Low	High	6.66	bi weekly
Low	Moderate	7.61	bi weekly
Moderate	High	6.21	bi weekly
Moderate	Moderate	6.81	bi weekly
Total		**6.83**	

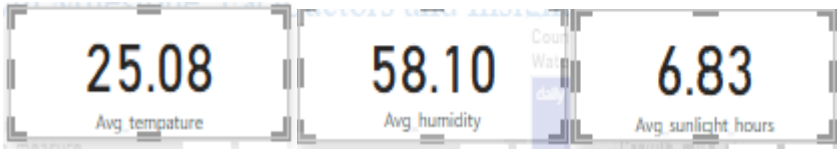
Area Chart: To track Average humidity by humidity range description.



Stacked Bar Chart: To compare Average temperature by temperature range description.



Cards: To display summary metrics like **Average Temperature, Humidity, and Sunlight**.



Milestone 5: Dashboard

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data, and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

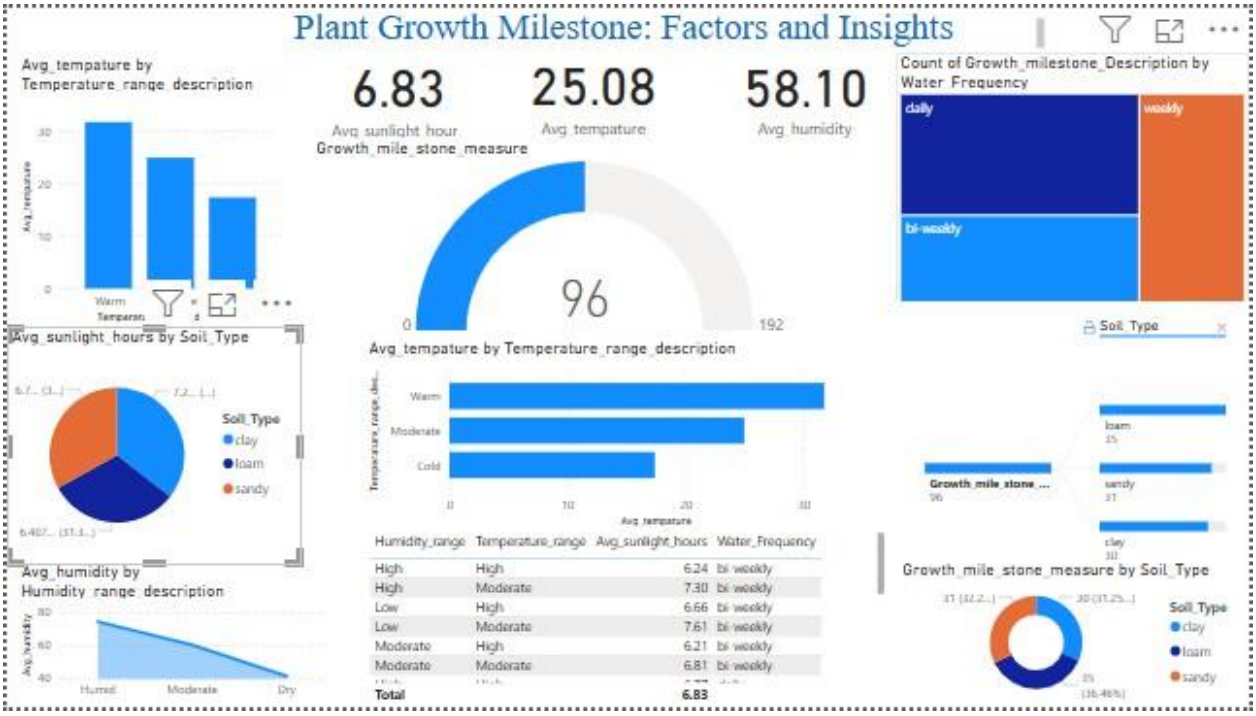
Activity :1- Responsive and Design of Dashboard

The responsiveness and design of the **Plant Growth Analysis Dashboard** are crucial to ensuring that the insights derived from the dataset are easily interpretable and actionable. Key considerations for designing a responsive and effective dashboard include:

- **User-Centered Design:** Ensuring that the dashboard is intuitive and easy to navigate.
- **Clear and Concise Information:** Presenting key metrics and visualizations in an easily understandable format.
- **Interactivity:** Implementing slicers and drill-throughs to allow users to filter data dynamically based on soil type, water frequency, temperature, and humidity.
- **Data-Driven Approach:** Utilizing real-time data updates to provide accurate and relevant insights on plant growth.
- **Accessibility:** Designing the dashboard to be user-friendly for all stakeholders, including farmers, researchers, and agricultural analysts.
- **Customization:** Allowing users to adjust visualization settings based on their analytical needs.
- **Security:** Ensuring that only authorized users have access to sensitive environmental data and plant growth analysis.

The ultimate goal is to create a **dashboard that is user-friendly, interactive, and data-driven**, providing actionable insights to improve decision-making in agricultural and environmental management.

Once you have created views on different sheets in Power Bi you can pull them into a dashboard.



Milestone 6: Report

A data report is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data Report can be told using a variety of mediums, presentations, interactive visualizations, and videos.

- Created a **comprehensive Power BI report** summarizing key insights.
- The report includes:
  - Growth patterns based on soil type and environmental conditions.
  - Correlation analysis between temperature, humidity, and plant growth.
  - Data-driven insights for better plant management decisions.

Milestone 7: Performance Testing

Performance testing is a crucial aspect of software development aimed at evaluating the speed, responsiveness, stability, and scalability of an application under various workload conditions. It involves simulating real-world usage scenarios to assess how the system behaves and performs under stress, peak loads, or normal conditions.

Activity 1: Utilization of Data Filters

The utilization of data filters plays a pivotal role in streamlining information processing and analysis across various domains. By selectively extracting or excluding specific data points based on predefined criteria, filters enable efficient data management and enhance decision-making processes.

- **Amount of Data Rendered to DB:**
  - Optimized dataset loading for better performance.

- **Number of Calculation Fields:**
    - Limited complex DAX calculations to maintain efficiency.

Activity 2: No of Visualizations/ Graphs

- **Growth Milestone by Soil Type** – To analyze how different soil types impact plant growth.
- **Water Frequency by Growth Milestone** – To examine how watering schedules affect plant growth.
- **Temperature Range by Plant Growth Category** – To compare plant growth stages across different temperature ranges.
- **Humidity Range by Growth Milestone** – To assess how varying humidity levels impact plant growth.
- **Fertilizer Type by Growth Milestone** – To analyze the effectiveness of different fertilizer types on plant growth.
- **Sunlight Hours by Soil Type** – To visualize how sunlight exposure varies across different soil types.
- **Temperature by Water Frequency** – To track temperature variations based on watering schedules.
- **Humidity by Soil Type** – To examine humidity variations across different soil compositions.
- **Growth Milestone by Temperature Range** – To determine plant growth success across different temperature conditions.
- **Water Frequency by Soil Type** – To analyze how watering patterns differ for various soil types.
- **Fertilizer Type by Soil Type** – To examine the correlation between soil type and fertilizer usage.
- **Growth Milestone by Humidity Range** – To assess the relationship between humidity and plant growth.

Milestone 8: Project Demonstration & Documentation

Below mentioned deliverables to be submitted along with other deliverables

Activity 1: - Record explanation Video for project end to end solution

Creating a record explanation video for a project's end-to-end solution is crucial for ensuring clarity and transparency in its implementation. This video serves as a comprehensive guide, detailing every aspect of the project from inception to completion.

Explanation video: Link

Activity 2: - Project Documentation-Step by step project development procedure

Create document as per the template provided

1. **Imported Dataset into Power BI**.
2. **Performed Data Cleaning & Transformation** using Power Query.
3. **Created New Columns & Measures** for deeper analysis.
4. **Designed Data Visualizations & Dashboard**.
5. **Developed & Published Power BI Report**.
6. **Conducted Performance Testing** to optimize efficiency.
7. **Finalized Documentation** for better project presentation.