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

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.

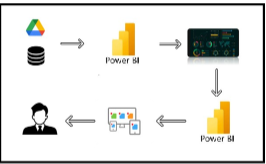
**Scenario 2:**

GreenEarth 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. GreenEarth Farms will use these insights to adjust their farming practices, ensuring their crops achieve the best possible growth under organic farming conditions.

**Scenario 3:**

FutureGrow 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. FutureGrow 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,  
1) Data Collection  
o Collect the dataset,  
o Connect Data with Power BI  
2) Data Preparation  
o Prepare the Data for Visualization

3) Data Visualizations  
o Visualizations  
4) Dashboard  
o Responsive and Design of Dashboard  
5) Report  
o Report Creation  
6) Performance Testing  
o Amount of Data Rendered to DB  
o Utilization of Data Filters  
o No. of Calculation fields  
o No. of Visualizations/Graphs  
7) Project Demonstration & Documentation  
o Record explanation Video for project end to end solution  
o Project Documentation-Step by step project development procedure

***Data Flow:***

***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, evaluate outcomes and generate insights from the data.

***Activity 1.1: understanding the data:***

Data contains all the meta information regarding the columns described in the CSV files.

***Dataset link:*** ***<https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification>***

***Column Description of the Dataset:***

* Soil\_Type: The type or composition of soil in which the plants are grown.
* Sunlight\_Hours: The duration or intensity of sunlight exposure received by the plants.
* Water\_Frequency: How often the plants are watered, indicating the watering schedule.
* Fertilizer\_Type: The type of fertilizer used for nourishing the plants.
* Temperature: The ambient temperature conditions under which the plants are grown.
* Humidity: The level of moisture or humidity in the environment surrounding the plants.
* Growth\_Milestone: Descriptions or markers indicating stages or significant events in the growth process of the plants.

***Milestone-2: Data Preparation***

Data preparation is a critical stage in the data analysis process, encompassing activities aimed at cleaning, transforming, and organizing raw data into a structured format suitable for analysis. This process involves identifying and addressing issues such as missing values, outliers, inconsistencies, and inaccuracies in the dataset, ensuring data quality and reliability.

***Activity-1: Preparation the data for visualizations:***

Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. This process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency. Since the data is already cleaned, we can move to visualization***.***

**Data Loading**: <https://drive.google.com/file/d/1VzDGGP-RxV_j0eI-uHzZvBwrl0ScNn6Z/view?usp=drive_link>

**Data Cleaning:** <https://drive.google.com/file/d/1-FHXFzIfoxyI8Eg-R_DbALnPQwuNkb5k/view?usp=drive_link>

To prepare data for visualizations in Power BI, specifically for inflation analysis across countries and years, followed these steps:

**1. Import Data into Power BI:**

* **File**: Imported Plant Growth data from an CSV or database file into Power BI.
  + Go to **Home** > **Get Data** > Choose file format (e.g. CSV).
  + Load the data into Power BI.

**2. Clean the Data (Remove Nulls, Duplicate Data):**

* Go to **Transform Data** (Power Query Editor) to clean the data.
* **Remove Nulls**: Select columns and filter out null values.
  + In Power Query, select the column > Click on the **Filter** icon > Uncheck **null** to remove them.
* **Remove Duplicates**: If there are duplicate rows, remove them by:
  + Go to **Transform** > **Remove Duplicates**.
* **Rename Columns**: Ensure column names are consistent and clear.
* **Change Data Types**:
  + Ensure that columns like Year are formatted as **Whole Numbers** and formatted as a **Decimal**.

Done this in Power Query or directly in Power BI by selecting the column in the Data View and changed the data type.

**3. Save Data Preparation Steps:**

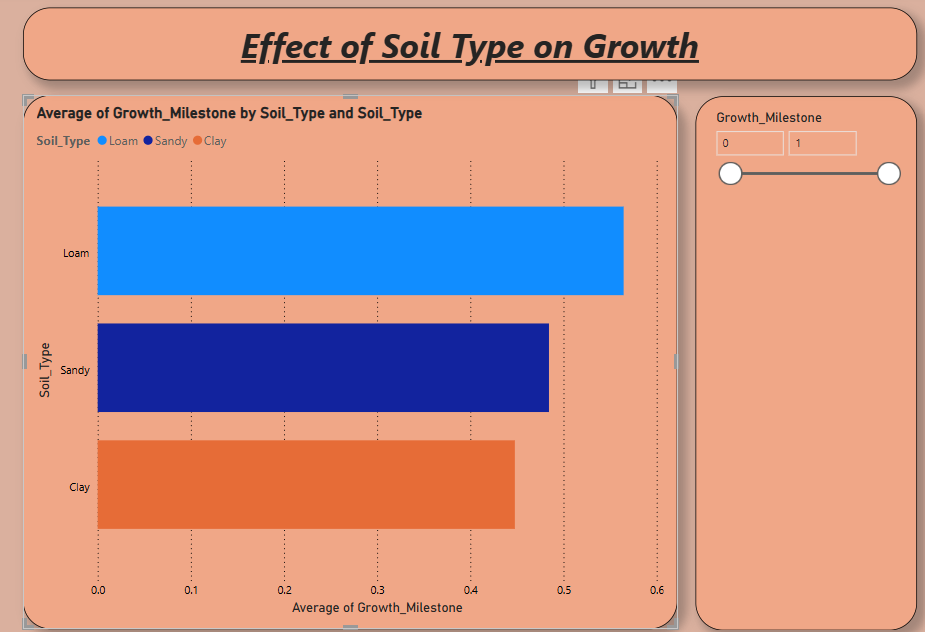
Once cleaned, transformed, click **Close & Apply** in the Power Query Editor to load changes into Power BI.

***Milestone-3: Data Visualization:***

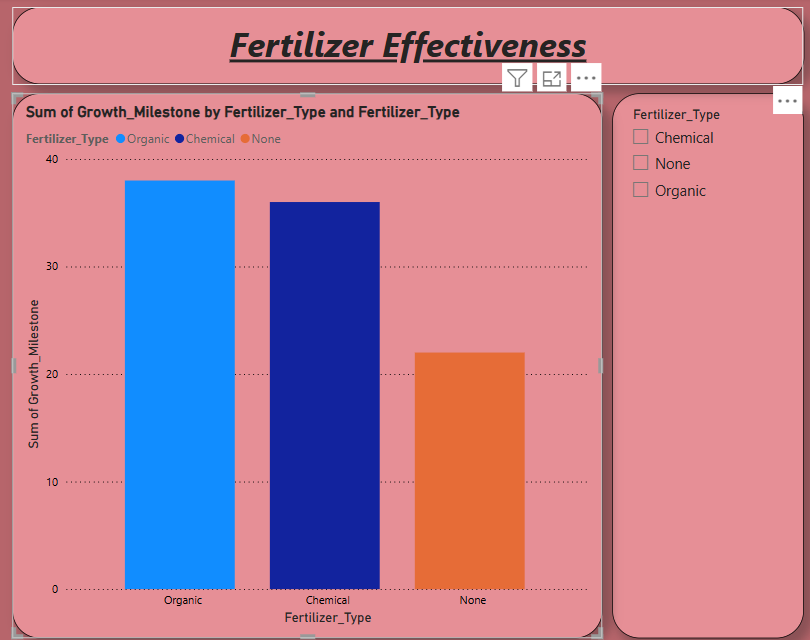
Data visualization is the process of creating graphical representations of data to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and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

**Activity-1.1. Bar Chart: Effect of Soil Type on Growth**

* **Purpose**: Compare different **Soil\_Type** values to see how each one affects the overall plant growth and milestone achievements.
* **X-Axis**: Soil\_Type
* **Y-Axis**: Average Growth Stage (using the AverageGrowthBySoilType DAX calculation).

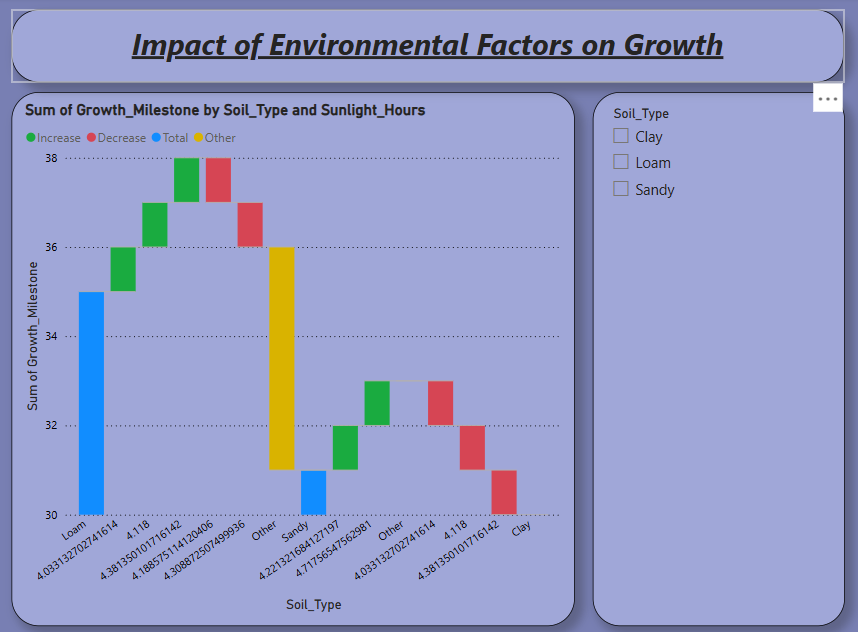
 **Activity-1.2. Clustered Column Chart: Fertilizer Effectiveness**

* **Purpose**: Compare the effectiveness of different fertilizers in achieving growth milestones.
* **X-Axis**: Fertilizer\_Type
* **Y-Axis**: Average Growth Milestone (using the FertilizerEffectiveness measure)



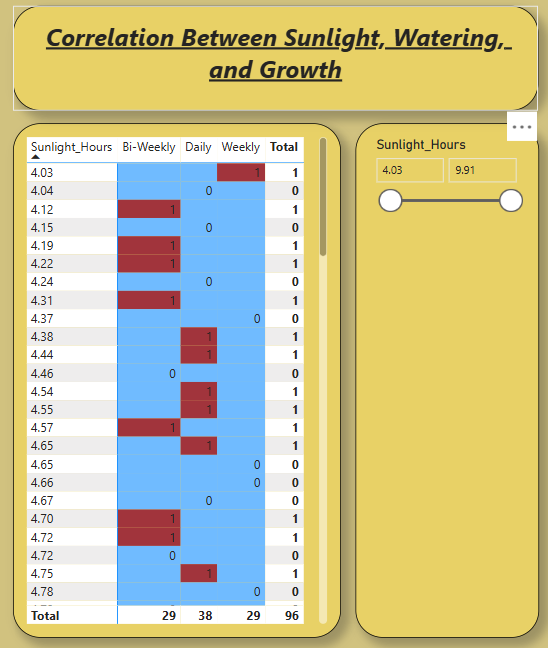
**Activity-1.3. Waterfall Chart: Impact of Environmental Factors on Growth**

* **Purpose**: Show the cumulative impact of various environmental factors (such as Temperature, Humidity, and Sunlight\_Hours) on plant growth progression.
* **X-Axis**: Sunlight Hours
* **Y-Axis**: Growth Milestone.



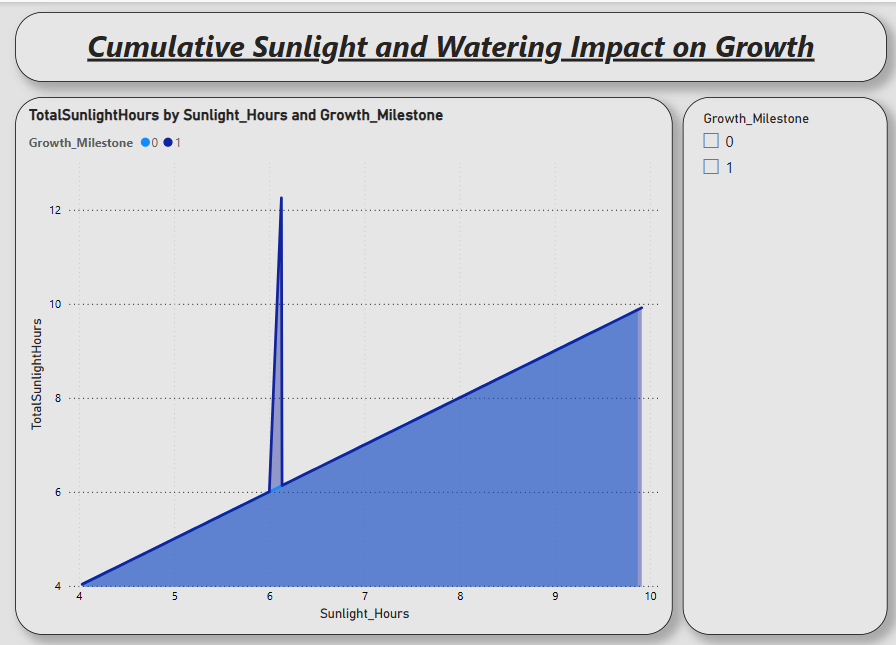
**Activity-1.4. Heatmap: Correlation Between Sunlight, Watering, and Growth**

* **Purpose**: Display correlations between multiple variables like sunlight, watering frequency, and growth stages.
* **X-Axis**: Sunlight\_Hours
* **Y-Axis**: Water\_Frequency
* **Color**: Growth\_Milestone



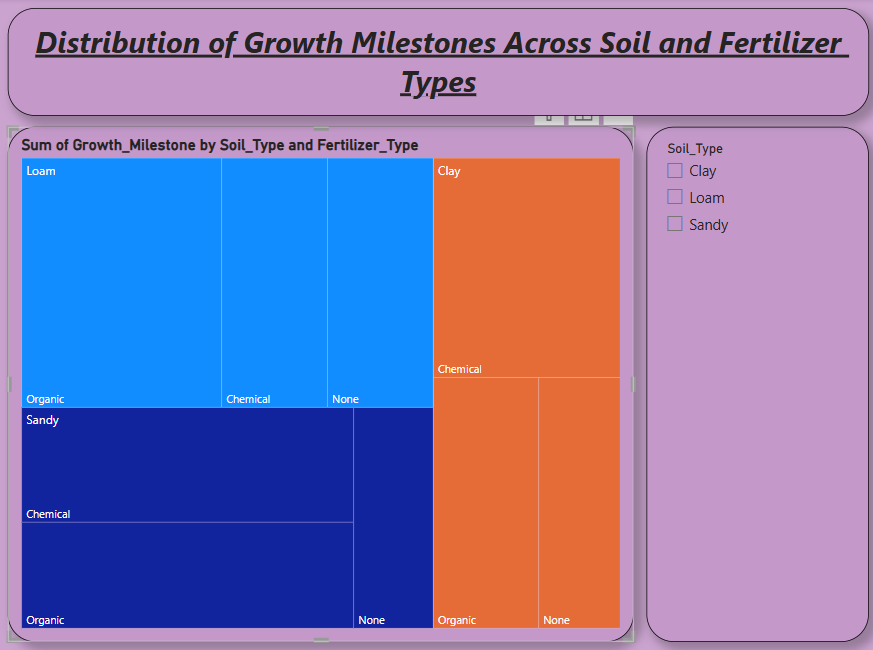
**Activity-1.5. Stacked Area Chart: Cumulative Sunlight and Watering Impact on Growth**

* **Purpose**: Show the cumulative effect of sunlight and watering frequency on the growth stages over time.
* **X-Axis**: Date
* **Y-Axis**: Cumulative Sunlight Hours
* **Legend**: Growth\_Milestone



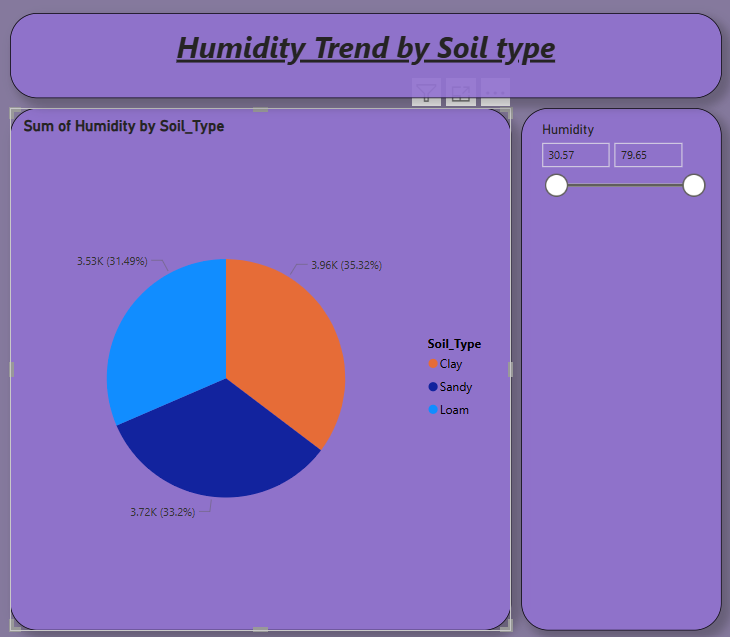
**Activity-1.6. Treemap: Distribution of Growth Milestones Across Soil and Fertilizer Types**

* **Purpose**: Use a treemap to represent the distribution of plants at various growth stages across different soil types and fertilizers.
* **Category**: Soil\_Type
* **Subcategory**: Fertilizer\_Type
* **values**: sum of growth milestone



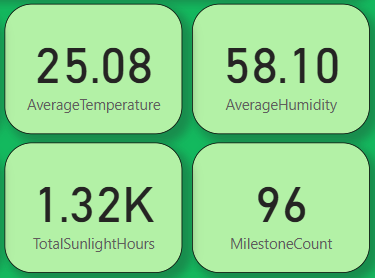
**Activity-1.7. Pie Chart: Humidity Trend by Soil type**

* **Purpose**: Using a Pie chart to represent the distribution of humidity and soil type.
* **Legend:** Soil\_Type
* **Values:** Sum of Humidity

****

**Activity-1.8.Cards:**

A **Card** visualization is great for displaying a single value, such as the Averages like Average temperature, Average Humidity, Totalsunlight.



***Milestone-4: 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:***

Created a responsive and well-designed dashboard in Power BI is key to effectively communicating of data insights***.***

**1. Key Visuals Placement:**

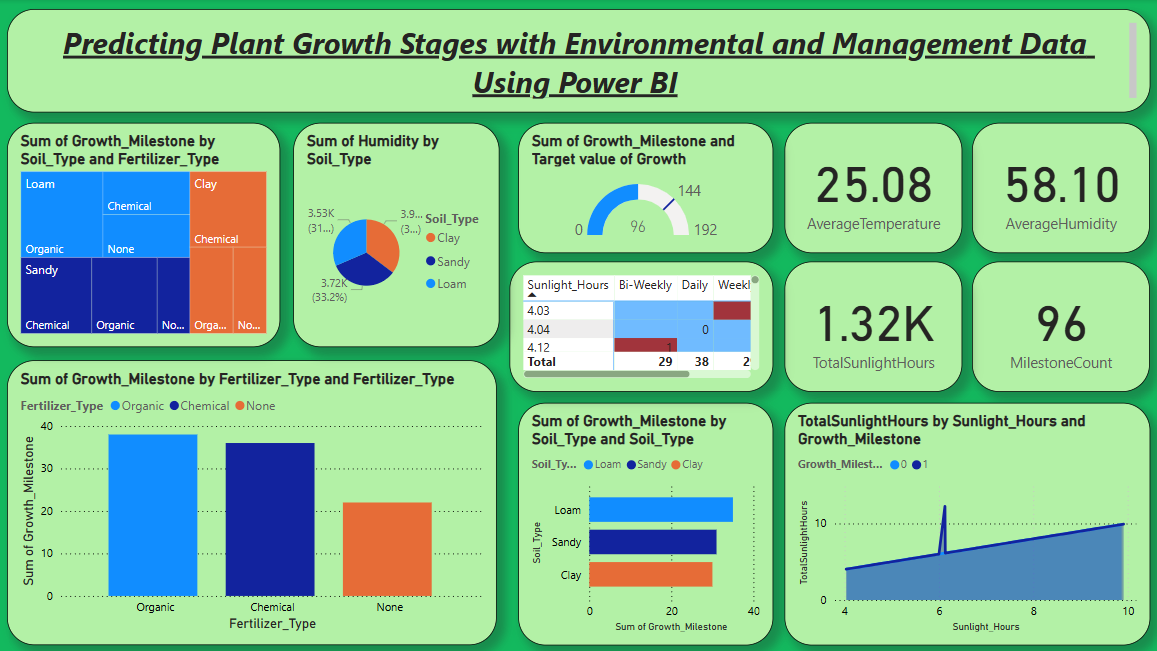
* **Top-Level KPIs**: Place key metrics such as **Current Growth Stage**, **Average Temperature**, or **Water Frequency** at the top of the dashboard. Use cards or gauges for these, making them large and immediately noticeable.
* **Charts and Trends**: Below the KPIs, place time-series charts like **line charts** or **scatter plots** to show trends such as **Growth Milestone over Time** or **Sunlight vs Growth Stage**.
* **Interactive Filters**: Add filters like **Soil Type**, **Fertilizer Type**, or **Date** to the side panel or at the top in a compact, accessible manner. Ensure users can easily select these to drill into specific segments.
* **Custom Visuals**: Consider using custom visuals, such as **timeline slicers** or **heatmaps**, for complex data such as cumulative impacts of sunlight or humidity on growth.

**2. Filters and Slicers:**

* **Dropdown or List Slicers**: Place slicers for **Soil\_Type**, **Fertilizer\_Type**, **Water\_Frequency**, or **Growth\_Milestone** in a dedicated area to allow users to focus on specific segments. The slicers should be responsive, meaning they adjust based on available space and change layout when viewed on mobile.
* **Date Range Filters**: Use a timeline slicer or date range filter to help users zoom into specific time periods when analyzing growth stages over time.

**3. Final Touches:**

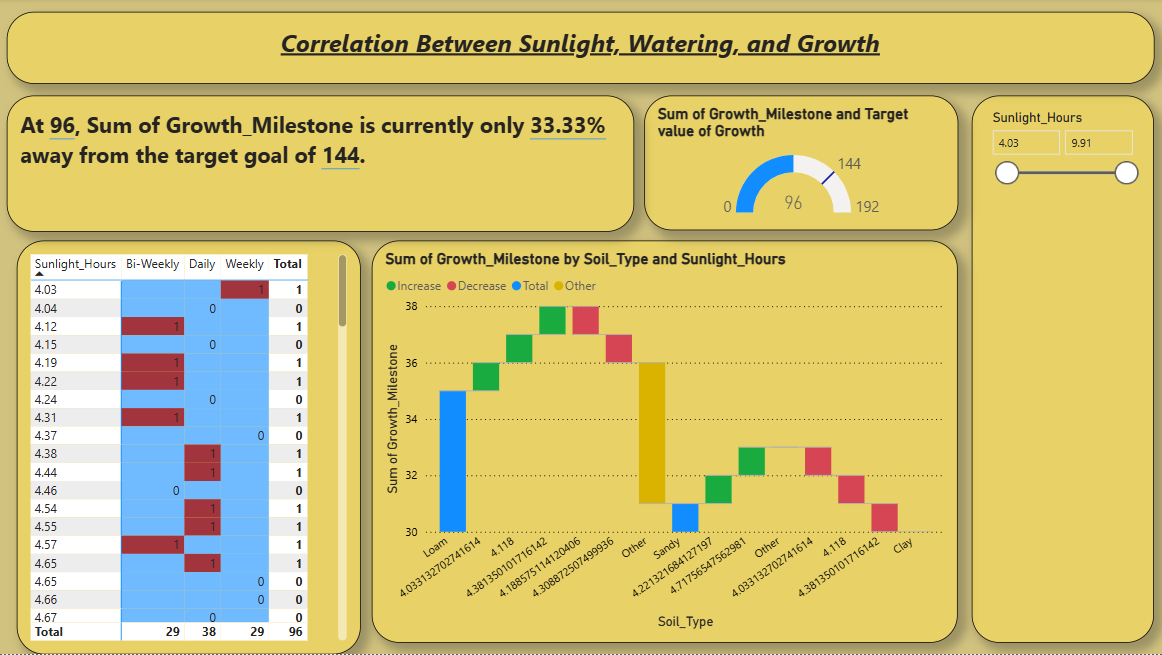
* **Interactive Elements**: Use buttons, bookmarks, and custom visuals to enhance the user experience.
* ***Theme: Apply a consistent theme (color palette,*** fonts, etc.) for professionalism and clarity.
* **Annotations**: Consider adding small text boxes or annotations near key visuals to guide users through insights or to highlight important .



Created a professional, responsive, and well-designed Power BI dashboard that clearly communicates the plant growth insights.

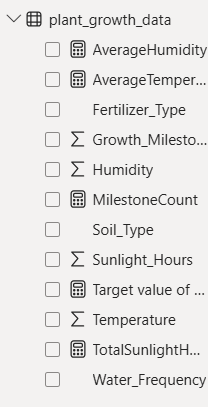
***Milestone-5: Report***

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.



***Milestone-6: Performance Testing***

"Amount of Data Loaded" refers to the quantity or volume of data that has been imported, retrieved, or loaded into a system, software application, database, or any other data storage or processing environment. It's a measure of how much data has been successfully processed and made available for analysis, manipulation, or use within the system.



***Activity-1:*  Utilization of DAX Expressions**

DAX (Data Analysis Expressions) in Power BI is a powerful formula language used to create custom calculations in calculated columns, measures, and tables. DAX expressions can be employed to manipulate data and perform complex calculations that are not possible with basic aggregations. They are similar to Excel formulas but are designed for relational data and can include functions for aggregation, time intelligence, and table manipulation. Understanding DAX is essential for unlocking the full potential of Power BI, as it allows users to create dynamic, interactive reports and dashboards that provide deep insights into data.

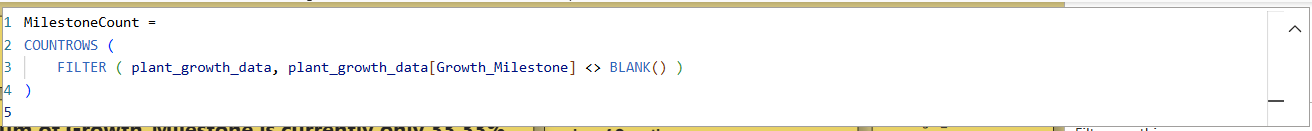
***Activity-1.1: Averagehumidity***



***Activity-1.2: AverageTemperature***



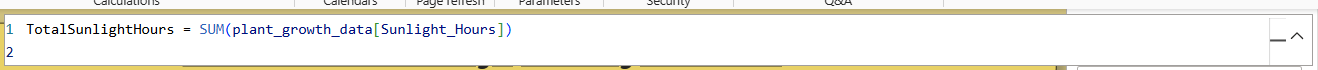
***Activity-1.3: MilestoneCount***



***Activity-1.4: Target value of Growth***



***Activity-1.5: TotalsunlightHours***



***Activity-2:* No of Visualizations/ Graphs:**

1. Bar Chart: Effect of Soil Type on Growth.

2.Clustered Column Chart: Fertilizer Effectiveness.

3.Waterfall Chart: Impact of Environmental .

4.Heatmap: Correlation Between Sunlight, Watering, and Growth.

5.Stacked Area Chart: Cumulative Sunlight and Watering Impact on Growth.

6.Treemap: Distribution of Growth Milestones Across Soil and Fertilizer Types.

7.Pie Chart: Humidity Trend by Soil type.

8.Cards.

***Milestone-7:* Project Demonstration & Documentation**

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

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