Predicting Plant Growth Stages

with Environmental and Management Data Using Power BI

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.

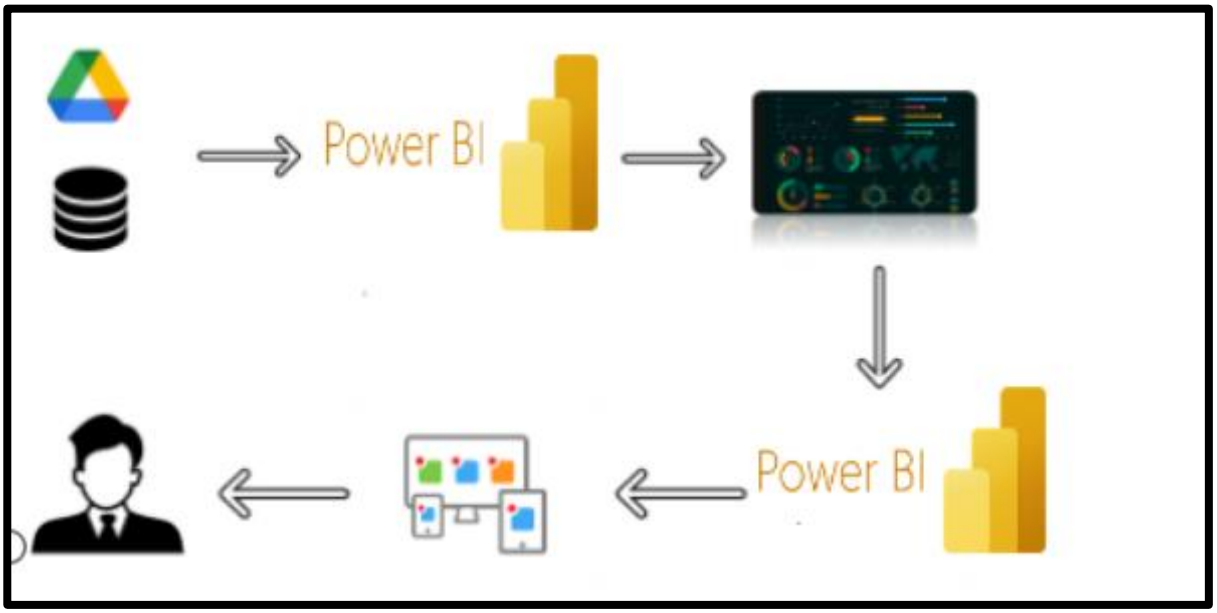
Scenario 2:

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,

- Data Collection
  - o Collect the dataset,
  - o Connect Data with Power BI
- Data Preparation
  - o Prepare the Data for Visualization
- Data Visualizations
  - o Visualizations
- Dashboard
  - o Responsive and Design of Dashboard
- Report
  - o Report Creation
- Performance Testing
  - o Utilization of Data Filters
  - o No. of Calculation fields
  - o No. of Visualizations/Graphs
- Project Demonstration & Documentation
  - o Record explanation Video for project end to end solution
  - o 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

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 phase in the data lifecycle, encompassing activities that transform raw data into a format suitable for analysis. This multifaceted process involves several steps including data cleaning, integration, transformation, and enrichment. Data cleaning involves identifying and rectifying errors, inconsistencies, and missing values within datasets to ensure accuracy and reliability.

Activity 1: Prepare the Data for Visualization

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.

Explanation video link 1:

https://drive.google.com/file/d/1VzDGGP-RxV_j0eI-uHzZvBwrl0ScNn6Z/view

Explanation video link 2:

https://drive.google.com/file/d/1-FHXFzIfoxyI8Eg-R_DbALnPQwuNkb5k/view

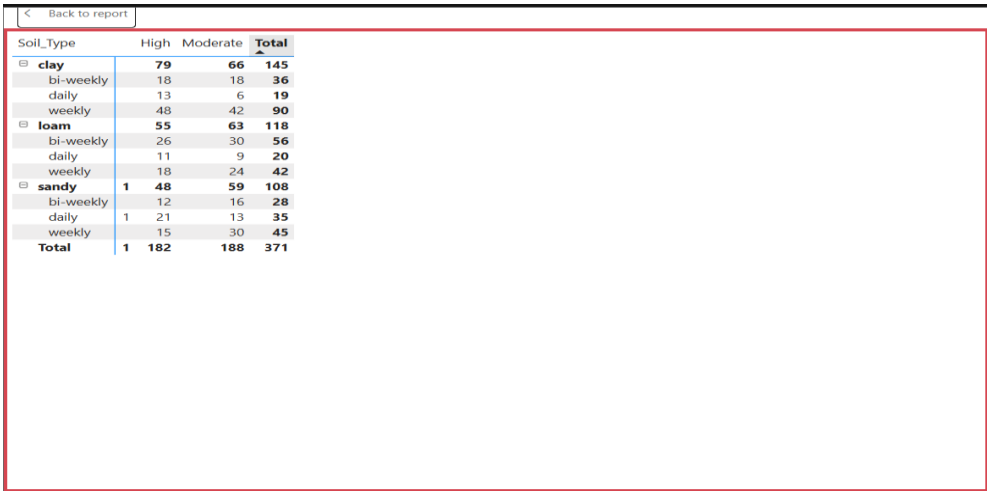
Milestone 3: Data Visualization

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: No of Unique Visualizations

The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the performance and efficiency of Social Pulse_ Illuminating the Digital Footprint - Unveiling Social Media Engagement Dynamics include bar charts, line charts, heat maps, scatter plots, pie charts,Maps etc. These visualizations can be used to compare performance, track changes over time, show distribution, and relationships between variables, breakdown of revenue and demographics, workload, resource allocation and location.

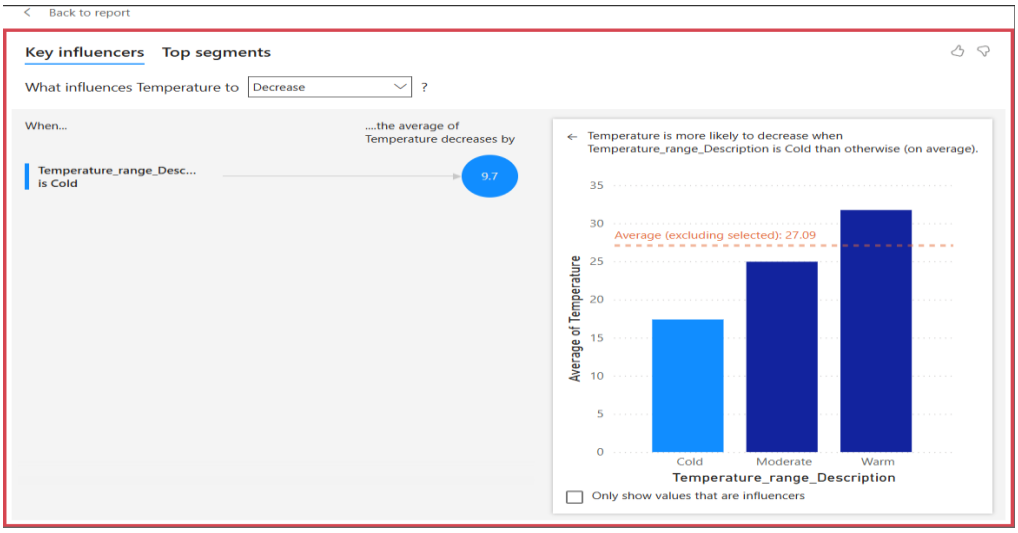
Activity 1.1: Water Frequency According to Its Soil Type



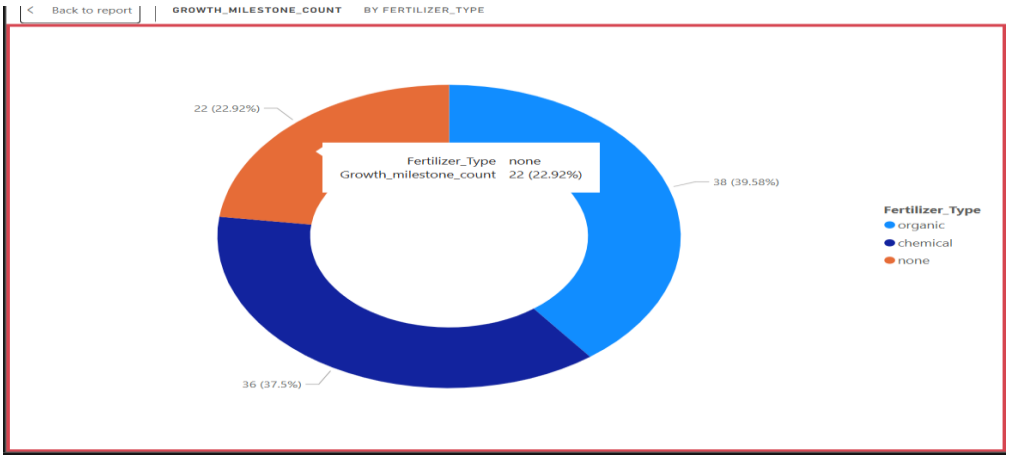
Soil_Type	High	Moderate	Total
clay	79	66	145
bi-weekly	18	18	36
daily	13	6	19
weekly	48	42	90
loam	55	63	118
bi-weekly	26	30	56
daily	11	9	20
weekly	18	24	42
sandy	1 48	59	108
bi-weekly	12	16	28
daily	1 21	13	35
weekly	15	30	45
Total	1 182	188	371

Activity 1.2: Average Temperature by Temperature Range Description

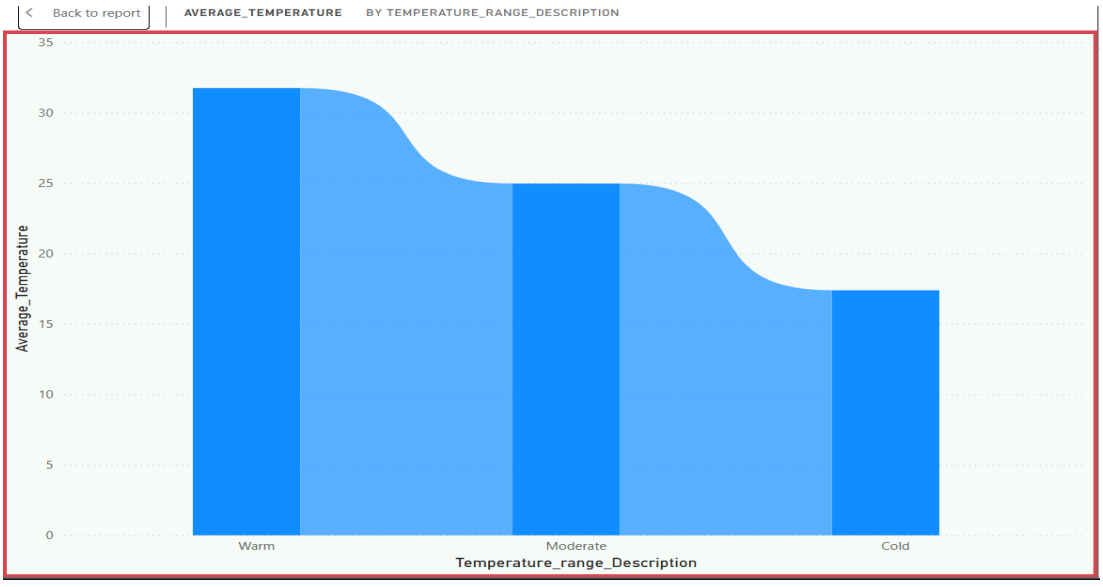
Activity 1.3: Temperature and Its Description According to Plant Growth



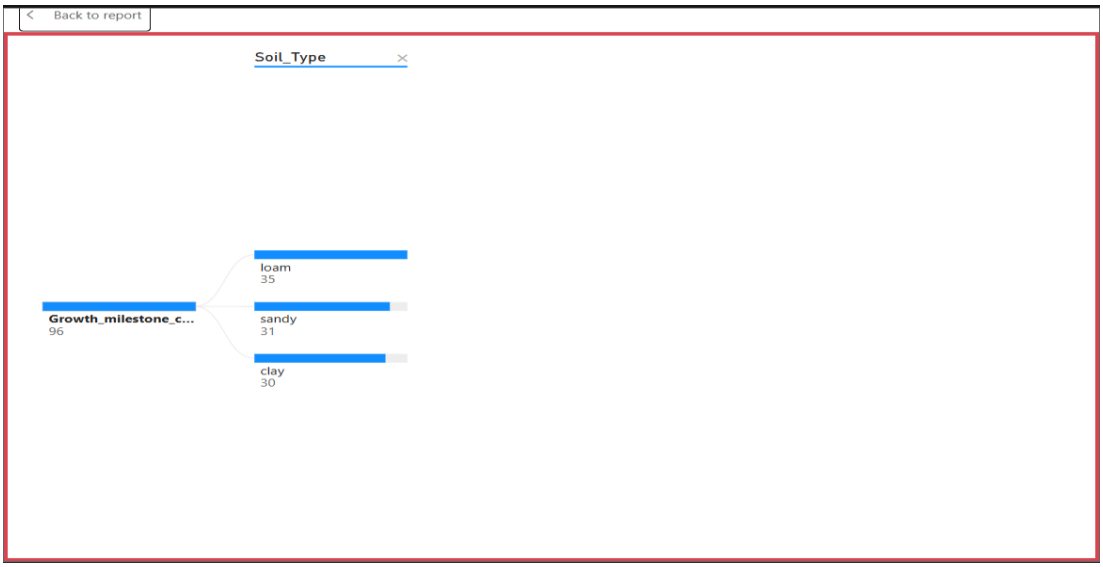
Activity 1.4: Growth Milestone Count by Fertilizer Type



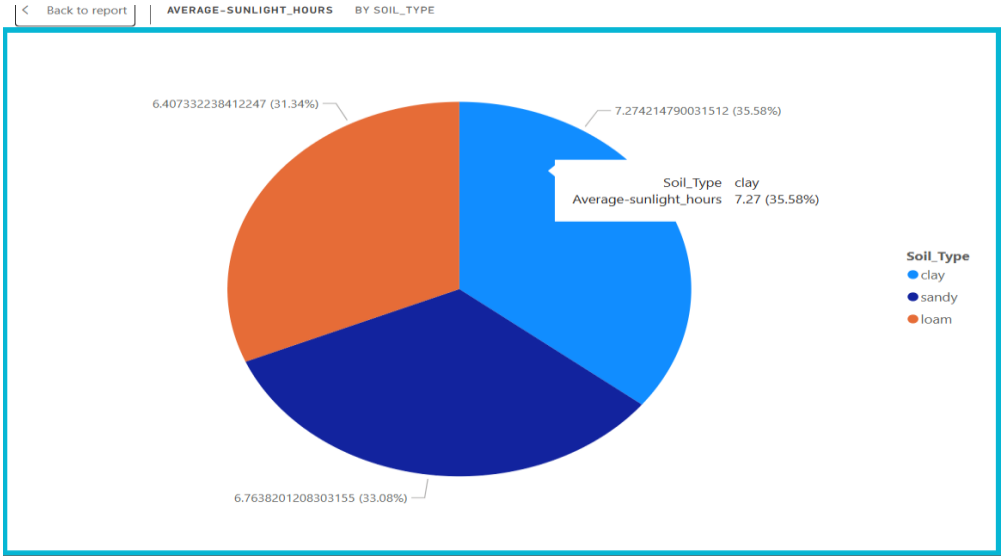
Activity 1.5: Average Humidity by Humidity Level Description



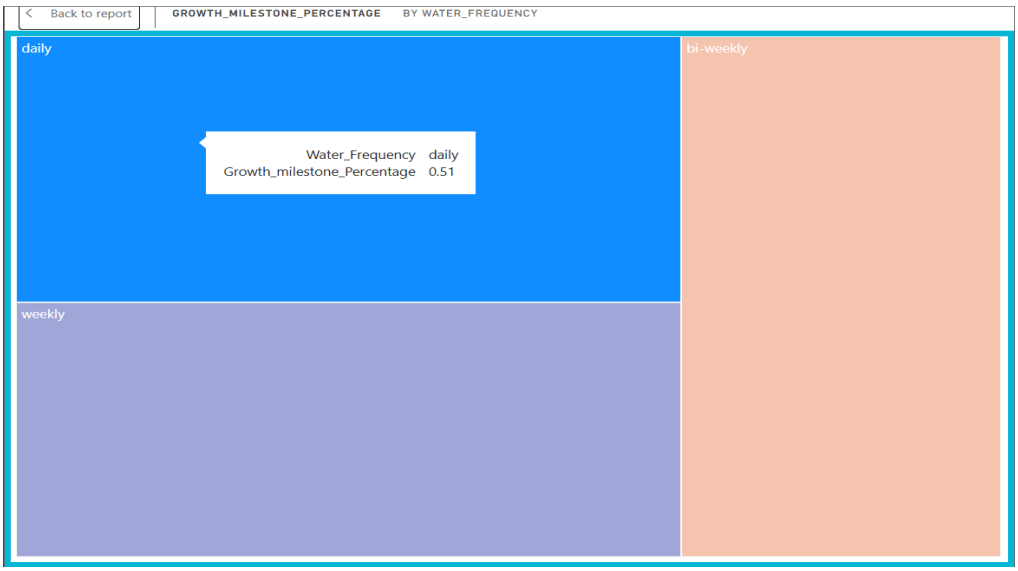
Activity 1.6: Growth Milestone Count According to Its Soil Type



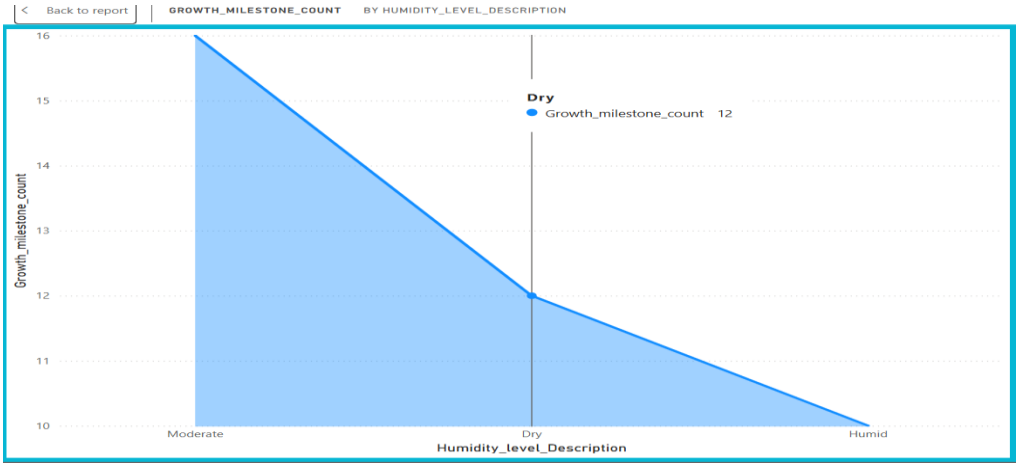
Activity 1.7: Average Sunlight Hours by Soil Type



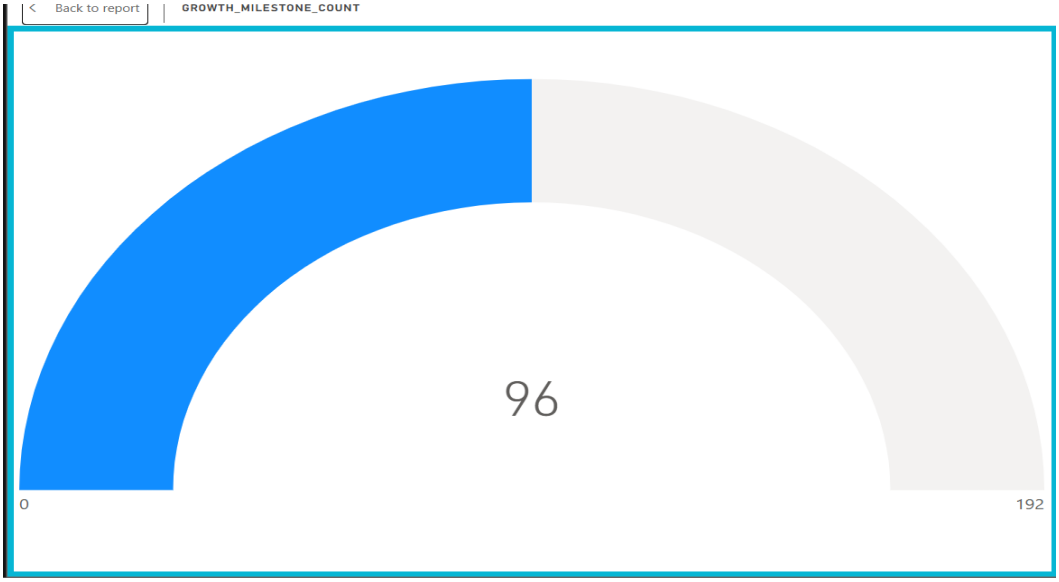
Activity 1.8: Growth Milestone Percentage by Water Frequency



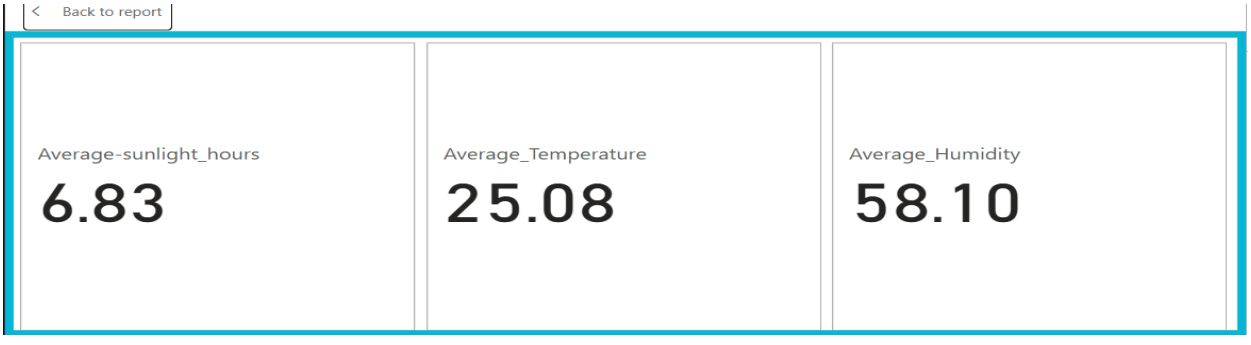
Activity 1.9: Growth Milestone Count by Humidity Level Description



Activity 1.10 : Growth Milestone Count



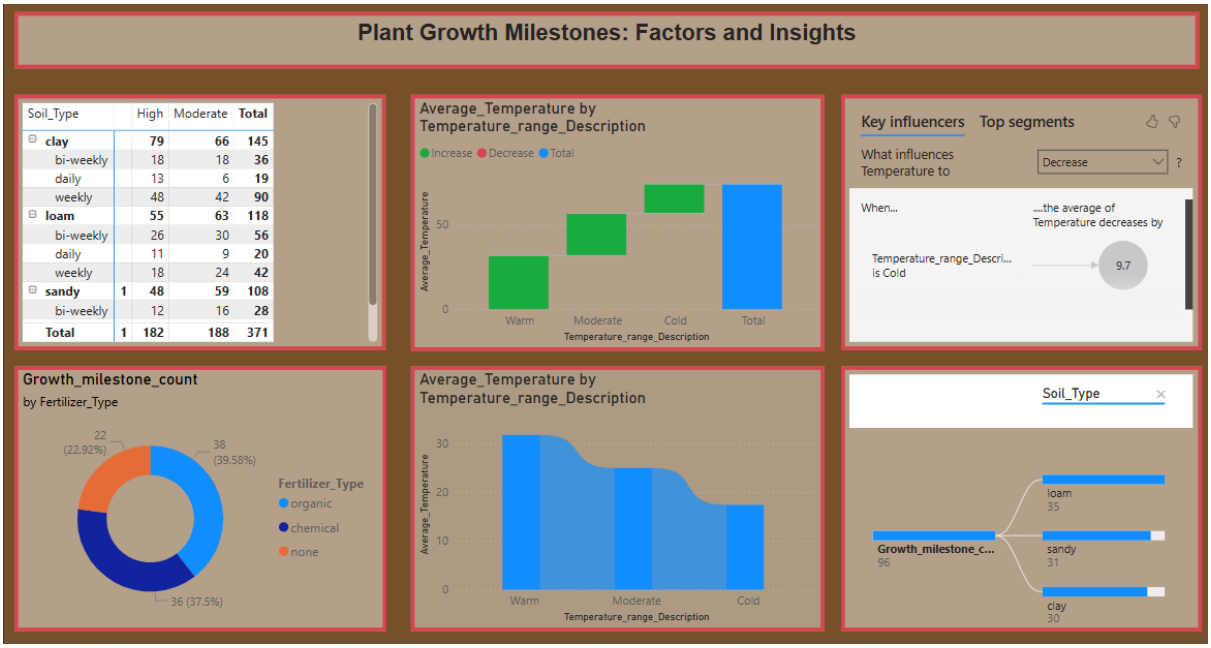
Activity 1.11: Average Humidity, Average Temperature, Average Sunlight Hours



Milestone 4: Dashboard

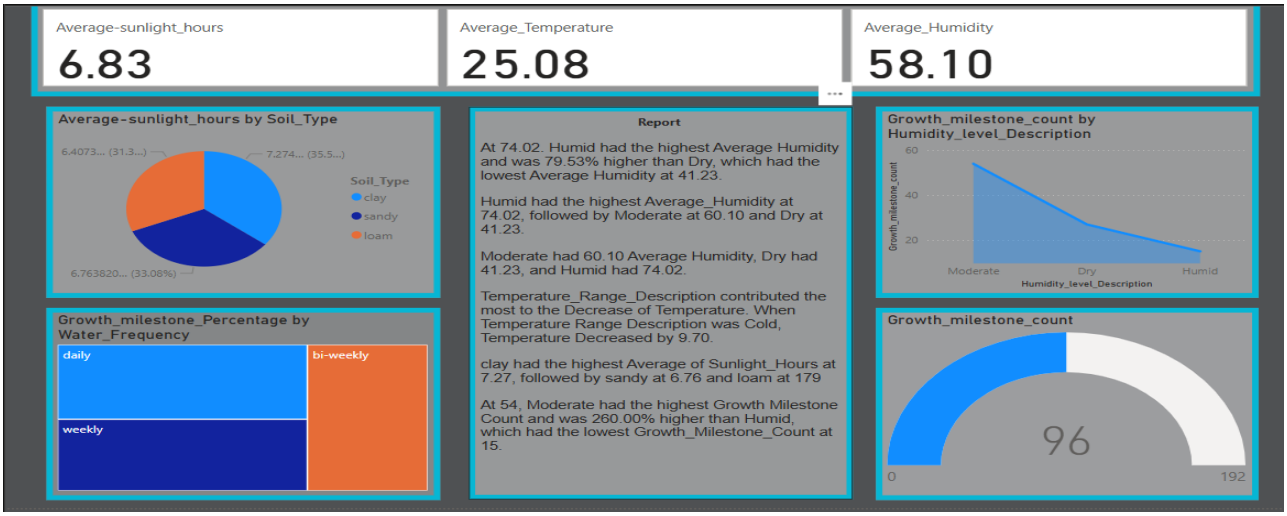
A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of 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



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 2.1: Average Humidity as “Measure”

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

Activity 2.2: Average Sunlight Hours as “Measure”

```
Average-sunlight_hours = AVERAGE(plant_growth_data[Sunlight_Hours])
```

Activity 2.3: Average Temperature as “Measure”

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

Activity 2.4: Growth Milestone Count as “Measure”

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

Activity 2.5: Growth Milestone Percentage as “Measure”

```
Growth_milestone_Percentage =
DIVIDE(
    [Growth_milestone_count],
    COUNTROWS(plant_growth_data),0
    )
```

Activity 2.6: Water Frequency Numeric as “New Column”

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

Activity 2.7: Temperature Range as “New Column”

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

Activity 2.8: Humidity Range as “New Column”

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

Activity 2.9: Humidity Level Description as “New Column”

```
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"
    )
```

Activity 2.10: Temperature Range Description as “New Column”

```
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"
)
```

Activity 2.11: Growth Milestone Description as “New Column”

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

Activity 2.12: Plant Growth Category as “New Column”

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

Activity 2 : No of Visualizations/ Graphs

1. Water Frequency According to Its Soil Type
2. Average Temperature by Temperature Range Description
3. Temperature and Its Description According to Plant Growth
4. Growth Milestone Count by Fertilizer Type
5. Average Humidity by Humidity Level Description
6. Growth Milestone Count According to Its Soil Type
7. Average Sunlight Hours by Soil Type
8. Growth Milestone Percentage by Water Frequency
9. Growth Milestone Count by Humidity Level Description
10. Growth Milestone Count
11. Average Humidity, Average Temperature, Average Sunlight Hours

Milestone 7: Project Demonstration & Documentation

https://drive.google.com/drive/folders/1IrFCMKArPn_w4BW9v9zNk6OIs-UhRGS8?usp=sharing