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.

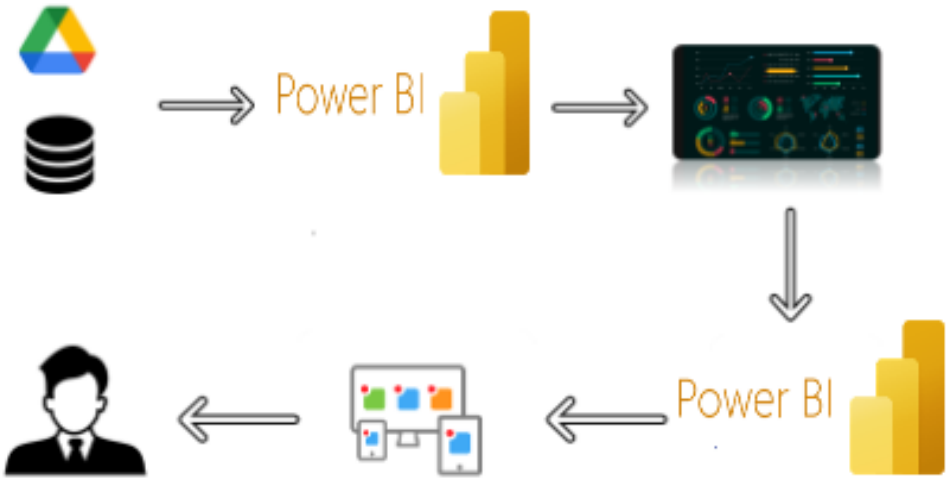
Scenario 1:
ABC Greenhouses has been facing challenges with inconsistent plant growth across its different greenhouse locations. By leveraging Power BI, the company plans to identify the best combination of soil type, sunlight hours, and watering frequency that leads to the highest growth milestones. 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 & 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

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: Download 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.

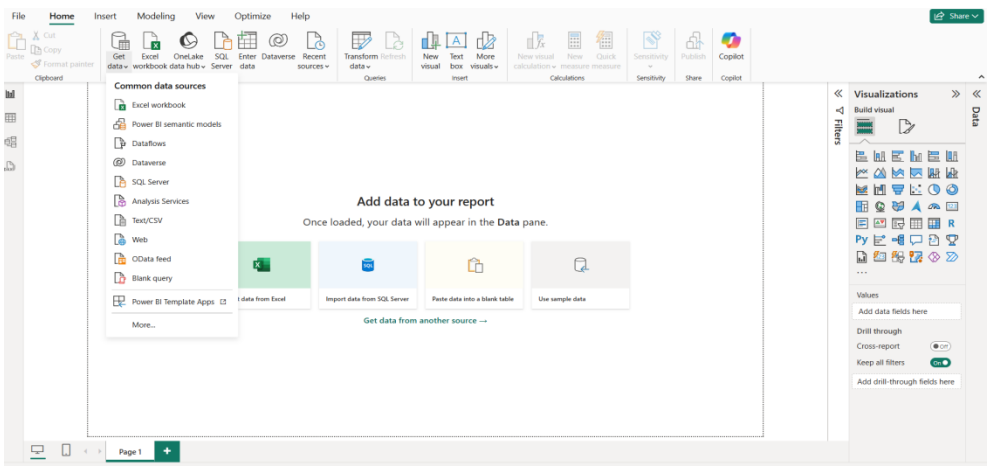
Activity 2: Connect Data with Power BI

With Power BI, users can seamlessly connect to a wide range of data sources, including databases, cloud services, spreadsheets, and streaming data. This capability allows organizations to consolidate disparate data sources into a single, unified platform, breaking down data silos and enabling holistic analysis.

Milestone 2: 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.

Data Loading:



Data Cleaning:

Data cleaning involves identifying and rectifying errors, inconsistencies, and missing values within datasets to ensure accuracy and reliability.

Growth Stage	Soil Type	Sunlight Hours	Water Frequency	Fertilizer Type	Temperature	Humidity	Growth Height	Date	Growth Score
Vegetative	Loam	5	Weekly	Chemical	34.17640505	77.697643...	165.6194588	01/01/2023	82
Vegetative	Silt	7	Bi-weekly	Chemical	25.21378234	31.609341...	64.04951732	02/01/20...	82
Mature	Sandy	9	Weekly	Organic	25.36183624	30.10072531	37.11319902	03/01/20...	80
Vegetative	Sandy	5	Weekly	Organic	30.39222547	76.4505903	170.0698616	04/01/20...	100
Flowering	Sandy	7	Monthly	Organic	30.85802149	49.91110382	61.68135535	05/01/20...	32
Mature	Loam	7	Weekly	None	32.88362835	65.758253...	189.3861214	06/01/20...	72
Mature	Sandy	7	Weekly	Organic	20.50199835	43.208467...	90.40318503	07/01/20...	65
Mature	Loam	10	Monthly	Chemical	23.47315813	41.7950195	156.0982958	08/01/20...	52
Mature	Silt	7	Monthly	Organic	16.2131641	58.415033...	106.8295199	09/01/20...	64
Flowering	Sandy	10	Bi-weekly	None	27.1541911	78.272323...	183.6863182	10/01/2023	96
Mature	Sandy	7	Monthly	Organic	19.60796703	62.89984...	71.39132253	11/01/2023	23
Germination	Loam	6	Daily	Chemical	24.86658544	32.626449...	86.29890972	12/01/2023	90
Flowering	Sandy	5	Daily	Chemical	29.52406362	33.878061...	99.4597348	13/01/2023	53
Flowering	Peat	10	Weekly	Chemical	26.64170161	35.667107...	117.7573152	14/01/2023	59
Flowering	Sandy	5	Monthly	Chemical	27.14467712	44.304153...	181.1178693	15/01/2023	46
Germination	Sandy	9	Weekly	Chemical	21.20401534	47.514278...	94.44378254	16/01/2023	56
Vegetative	Sandy	6	Monthly	Chemical	19.31124327	77.340924...	52.92570655	17/01/2023	1
Mature	Peat	6	Daily	None	18.55588897	32.189695...	51.72702605	18/01/2023	14
Flowering	Loam	10	Weekly	Organic	17.04809086	78.753417...	169.5972526	19/01/2023	100
Vegetative	Sandy	10	Monthly	Chemical	20.03500775	47.156076...	187.485153	20/01/20...	39
Mature	Clay	6	Weekly	Chemical	15.17973578	72.11667838	99.66515841	21/01/2023	1
Germination	Silt	7	Daily	Organic	18.02025103	45.416952	156.9370949	22/01/20	55

Milestone 3: Data Visualization

Plant Growth Classification

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: Water Frequency According to Its Soil Type

Back to report | WATER FREQUENCY ACCORDING TO ITS SOIL TYPE

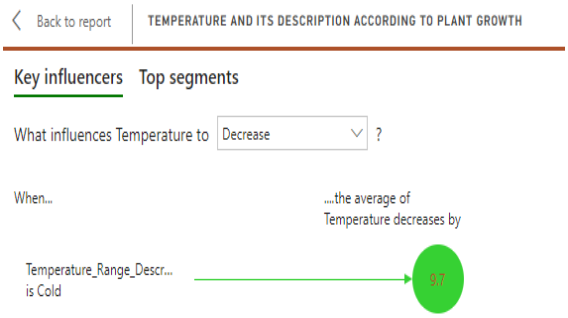
Soil_Type	High	Moderate	Total
loam	**191.61**	**205.65**	**397.25**
weekly	45.06	51.26	**96.31**
daily	64.02	54.73	**118.75**
bi-weekly	82.53	99.66	**182.19**
sandy	**215.43**	**217.46**	**432.88**
bi-weekly	33.29	60.49	**93.78**
weekly	35.27	70.65	**105.92**
daily	146.87	86.32	**233.19**
clay	**280.24**	**207.13**	**487.37**
bi-weekly	59.80	68.15	**127.95**
daily	95.42	40.66	**136.09**
weekly	125.02	98.32	**223.34**
Total	**687.28**	**630.24**	**1,317.51**

Insight: Loam soil with high water frequency results in the highest total water frequency, indicating that loam soil may require or benefit from more frequent watering compared to sandy and clay soils.

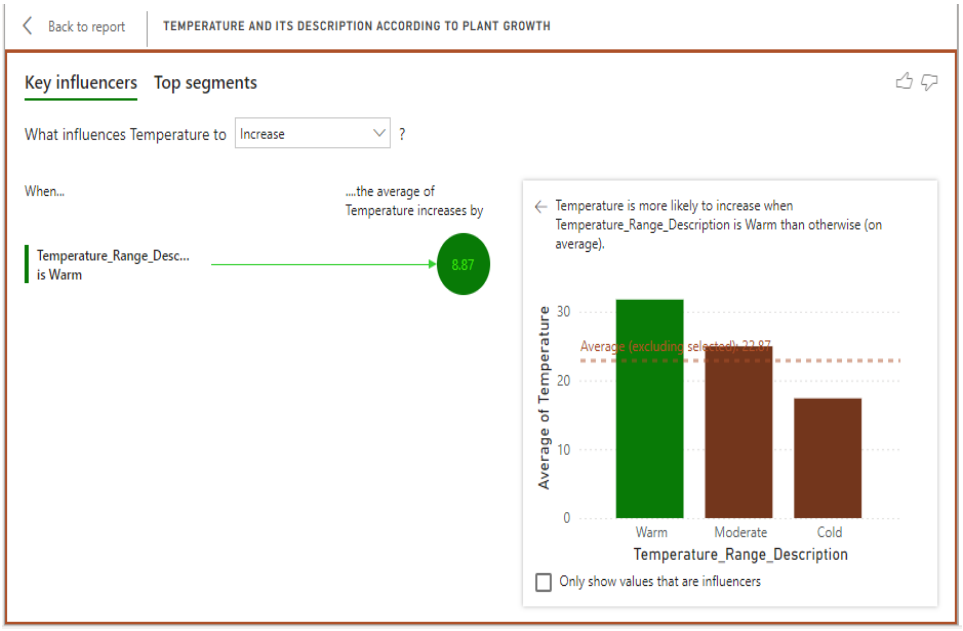
Activity 1.2: Average Temperature by Temperature Range Description

Insight: Moderate temperature ranges have the highest average temperature increase, suggesting that plants in this range experience the most growth.

Activity 1.3: Temperature and Its Description According to Plant Growth

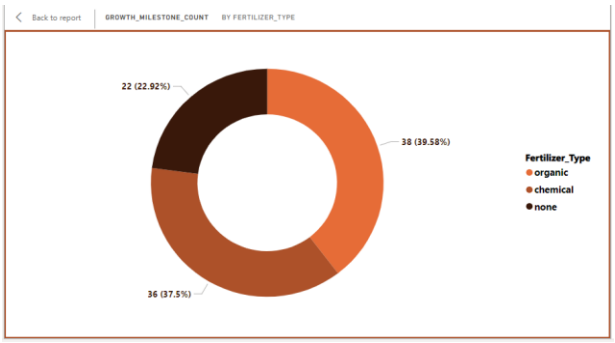


Insight: Cold temperature range decreases plant growth, highlighting the negative impact of cold temperatures on plant development.



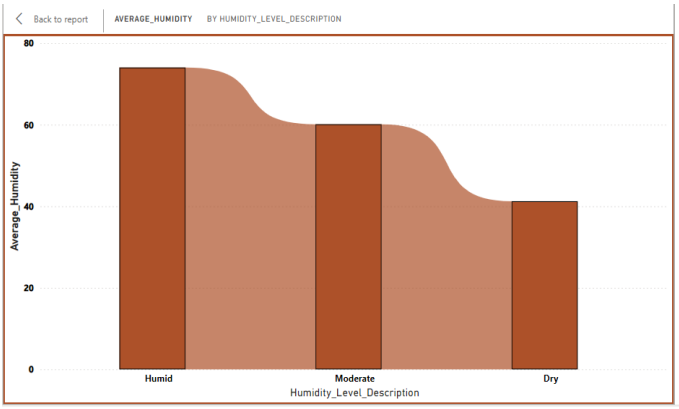
Insight: Warm temperature range increases plant growth, highlighting the positive impact of warm temperatures on plant development.

Activity 1.4: Growth Milestone Count by Fertilizer Type



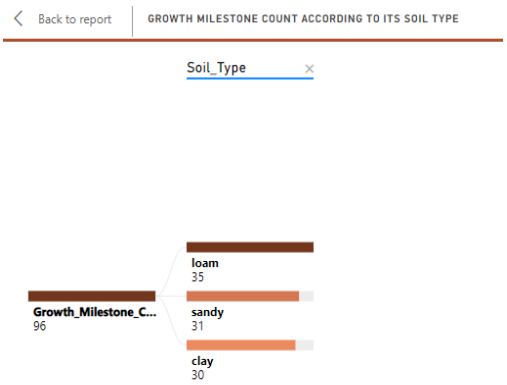
Insight : Chemical fertilizers account for the highest growth milestone count, indicating their effectiveness in promoting plant growth milestones compared to organic fertilizers or no fertilizer.

Activity 1.5: Average Humidity by Humidity Level Description



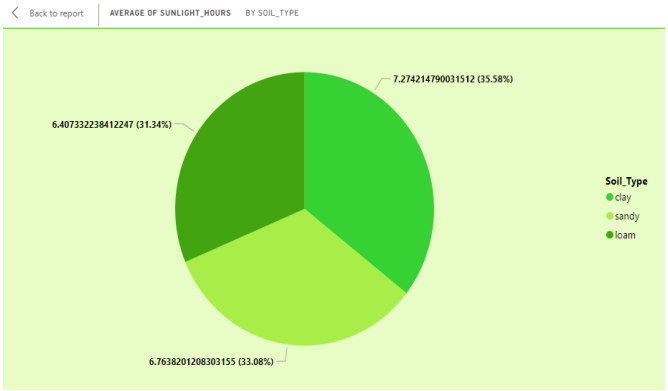
Insights : Humid conditions lead to the highest average humidity, emphasizing the importance of maintaining high humidity levels for optimal plant growth.

Activity 1.6: Growth Milestone Count According to Its Soil Type



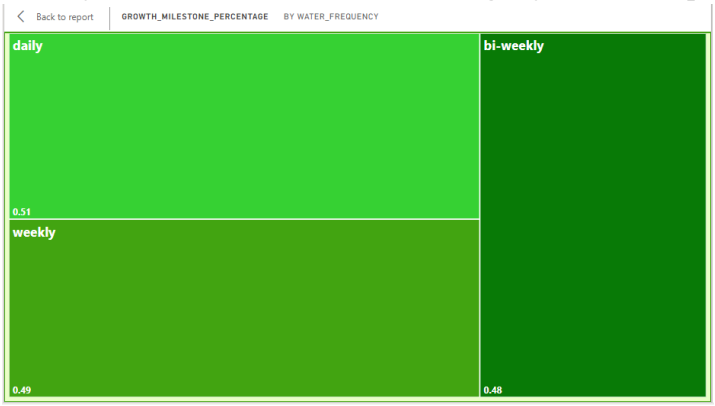
Insights : Loam soil has the highest growth milestone count, suggesting it is the most conducive soil type for achieving plant growth milestones.

Activity 1.7: Average Sunlight Hours by Soil Type



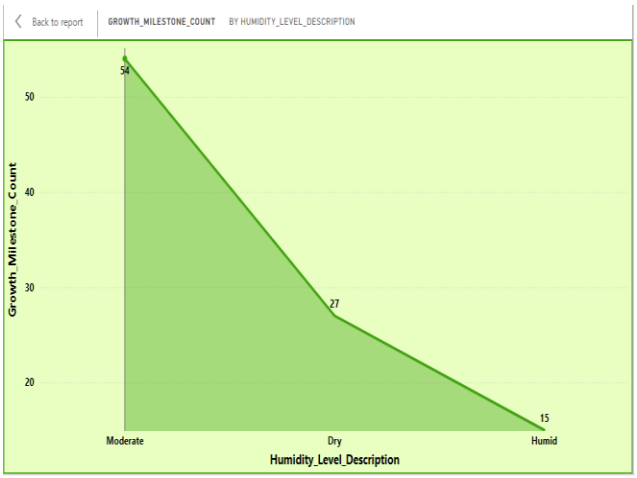
Insights : Sandy soil receives the highest average sunlight hours, which may contribute to its growth performance under sufficient light conditions.

Activity 1.8: Growth Milestone Percentage by Water Frequency



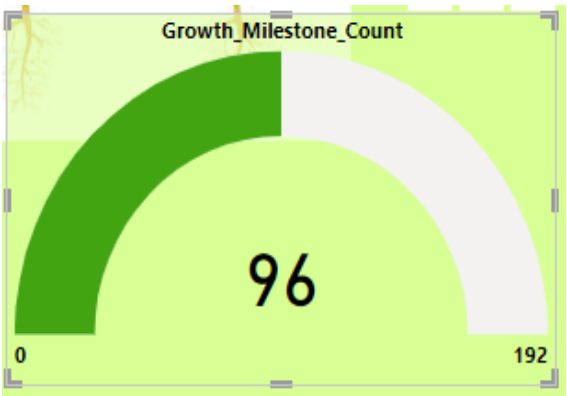
Insights : Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.

Activity 1.9: Growth Milestone Count by Humidity Level Description



Insights: Humid conditions lead to the highest growth milestone count, reinforcing the importance of maintaining adequate humidity levels for optimal growth

Activity 1.10 : Growth Milestone Count



Insights: The total growth milestone count is 96, with the highest contributions coming from environments with optimal humidity and temperature conditions.

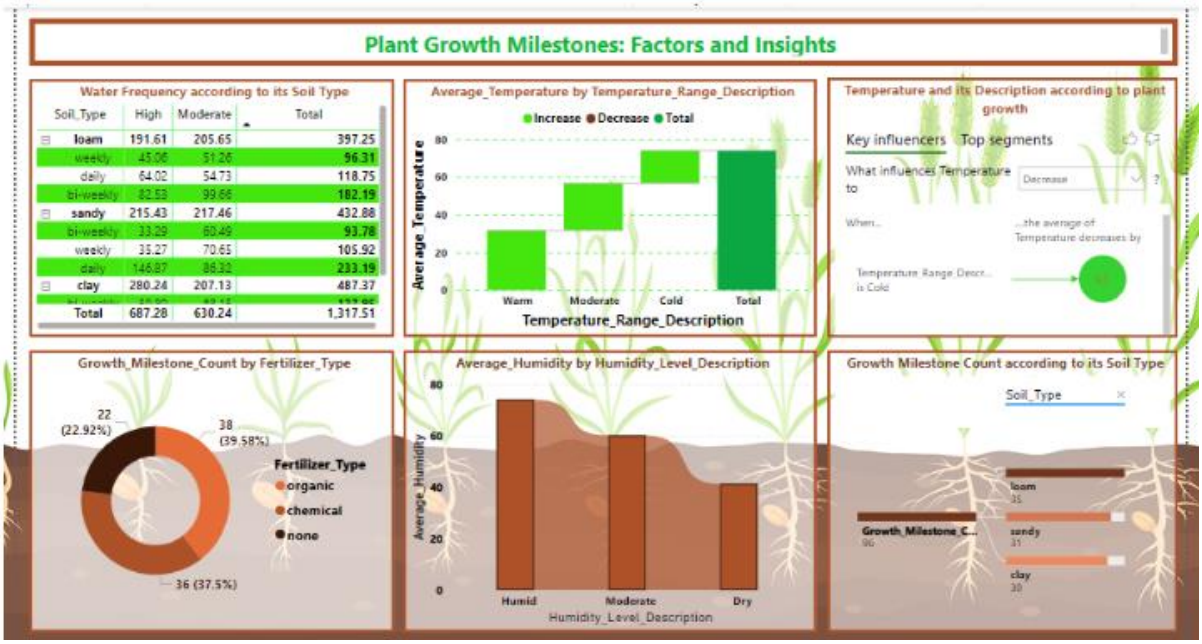
Activity 1.11: Average Humidity, Average Temperature, Average Sunlight Hours

Insights:-The data indicates a temperate and comfortable climate with an average temperature of 25.08°C, moderate humidity of 58.10%, and a reasonable amount of sunlight averaging 6.83 hours per day.

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



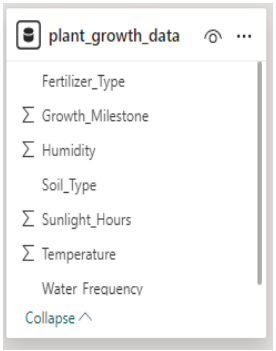
Milestone 6: Report

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.

Report:

Milestone 7: 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.



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_Milestone_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"
)

```

No of Visualizations/ Graphs

- Water Frequency According to Its Soil Type
- Average Temperature by Temperature Range Description
- Temperature and Its Description According to Plant Growth
- Growth Milestone Count by Fertilizer Type
- Average Humidity by Humidity Level Description
- Growth Milestone Count According to Its Soil Type
- Average Sunlight Hours by Soil Type

- Growth Milestone Percentage by Water Frequency
- Growth Milestone Count by Humidity Level Description
- Growth Milestone Count
- Average Humidity, Average Temperature, Average Sunlight Hours