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

Introduction

XYZ Company, known for its innovative agricultural solutions, is launching a project aimed at optimizing plant growth using advanced data analytics and visualization techniques through Power BI. This project involves analyzing a detailed dataset that includes critical environmental and management factors such as **soil type, sunlight hours, water frequency, fertilizer type, temperature,** and **humidity**.

By utilizing this data, the company aims to **predict plant growth milestones**, which are key to understanding the conditions necessary for optimal growth. The project will include the development of **interactive dashboards** and **predictive models** to uncover trends and insights that can support better agricultural practices and improve greenhouse management.

The analysis will be performed using **decomposition trees**, allowing for a breakdown of growth milestone counts by different variables to understand the effect of each one. In addition, several **calculated columns** and **measures** will be created to enhance the analytical depth of the dataset. Visualizations such as **clustered bar charts**, **pie charts**, **scatter plots**, and **column charts** will be used to effectively communicate the findings.

Through this approach, XYZ Company aims to **increase crop yield**, **optimize the use of resources**, and encourage **sustainable farming practices**, thereby reinforcing its leadership in agricultural innovation.

Scenario 1: ABC Greenhouses

ABC Greenhouses has been experiencing inconsistent plant growth across its various greenhouse locations. To solve this, the company plans to use Power BI to identify the most effective combination of **soil type**, **sunlight duration**, and **watering frequency** that leads to the highest growth milestones.

Using a **decomposition tree**, the team analyzes growth milestone counts and discovers that **loam soil**, when combined with **daily watering** and **6–8 hours of sunlight**, yields the best results.

Scenario 2: GreenEarth Farms

GreenEarth Farms has observed variability in the growth rates of its organic crops and wants to ensure more consistent yields. By analyzing the dataset, the company finds that a combination of **organic fertilizer**, **loam soil**, and **bi-weekly watering** results in the most significant growth milestones.

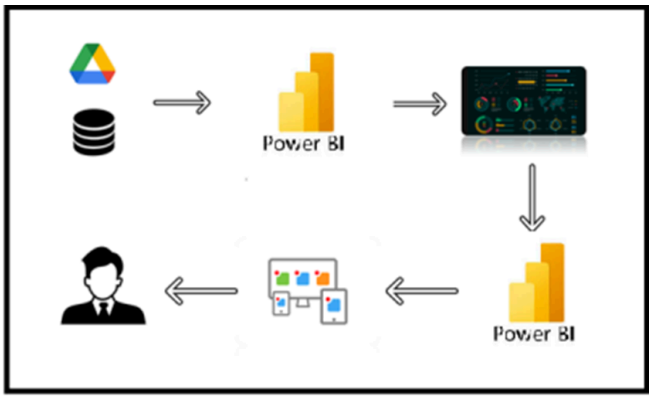
Further analysis using the decomposition tree reveals that maintaining **temperatures between 20–30°C** and **humidity levels between 50–70%** further enhances plant growth. With these findings, GreenEarth Farms will adjust their farming methods to support the best possible growth under organic conditions.

Scenario 3: FutureGrow Tech

FutureGrow Tech, a company focused on smart farming innovations, is working to validate the performance of its technology under varying environmental conditions. Using Power BI to study the dataset, they find that their **smart sensors**, which monitor soil moisture and automatically adjust watering schedules, significantly improve plant growth milestones.

The decomposition tree shows that these sensors are most effective when used with **sandy soil**, a **weekly organic fertilizer schedule**, and under **moderate temperature and humidity** conditions. Based on these results, FutureGrow Tech will incorporate the insights into their product development to enhance the precision and effectiveness of their smart farming solutions.

Technical Architecture:



Project Flow

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

1. Data Collection
    a. Collect the dataset,
    b. Connect Data with Power BI
2. Data Preparation
    - Prepare the Data for Visualization
- Data Visualizations
    - Visualizations
- Dashboard
    - Responsive and Design of Dashboard
- Report

- Report Creation
- Performance Testing
  - Utilization of Data Filters
  - No. of Calculation fields
  - 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, evaluate outcomes and generate insights from the data.

Activity 1: Downloading 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.

Activity 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.

3.1: Data Loading
3.2 Data Cleaning

Milestone 2: 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: Plant Growth Classification

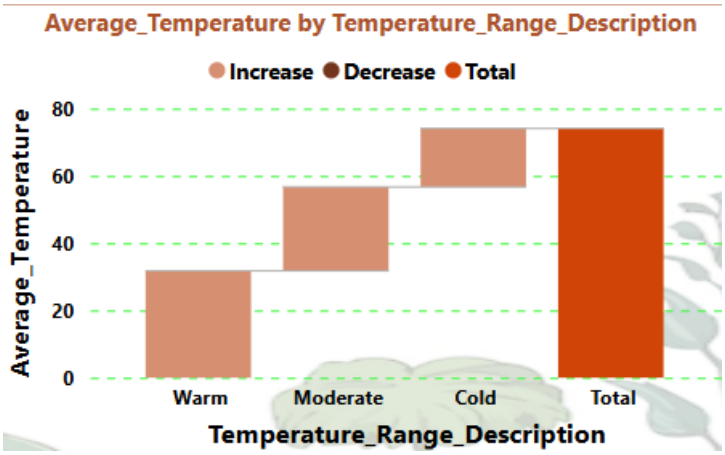
Activity 1.1: Water Frequency According to Its Soil Type

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**
clay	**280.24**	**207.13**	**487.37**
Total	**687.28**	**630.24**	**1,317.51**

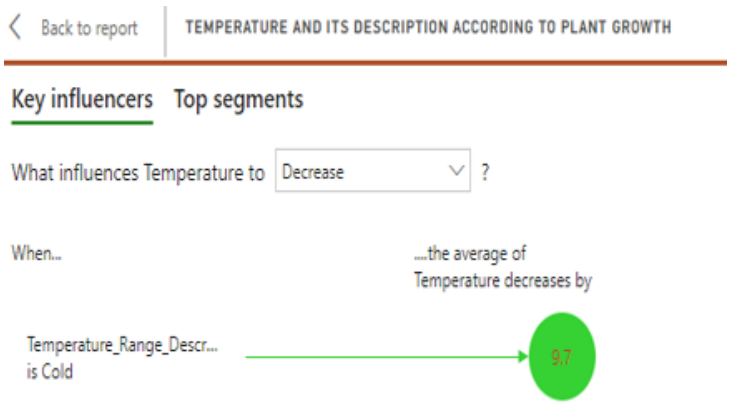
Insight: Loam soil with high water frequency results in the highest total water frequency, indicating that loam soil may require or benefit from more frequent watering compared to sandy and clay soils.

Activity 1.2: Average Temperature by Temperature Range Description

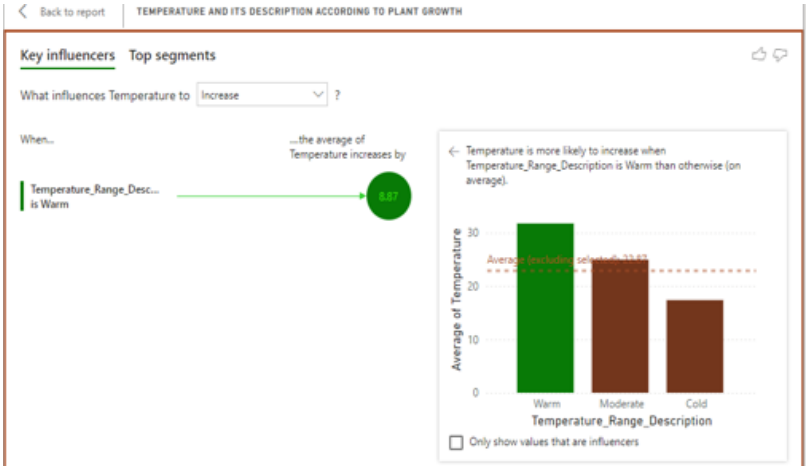


Insight: Moderate temperature ranges have the highest average temperature increase, suggesting that plants in this range experience the most growth.

Activity 1.3: Temperature and Its Description According to Plant Growth

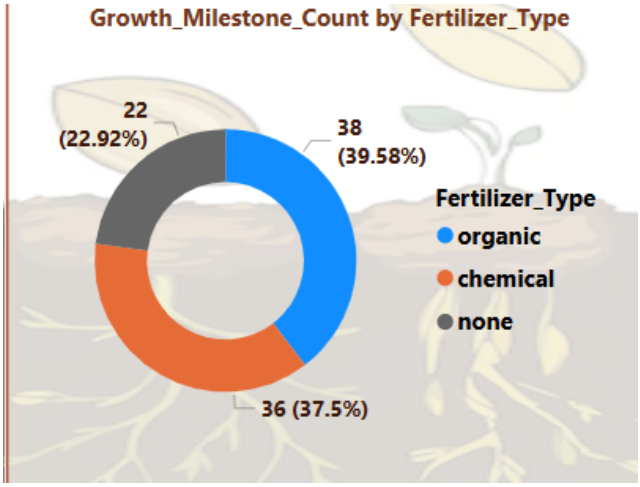


Insight: Cold temperature range decreases plant growth, highlighting the negative impact of cold temperatures on plant development.



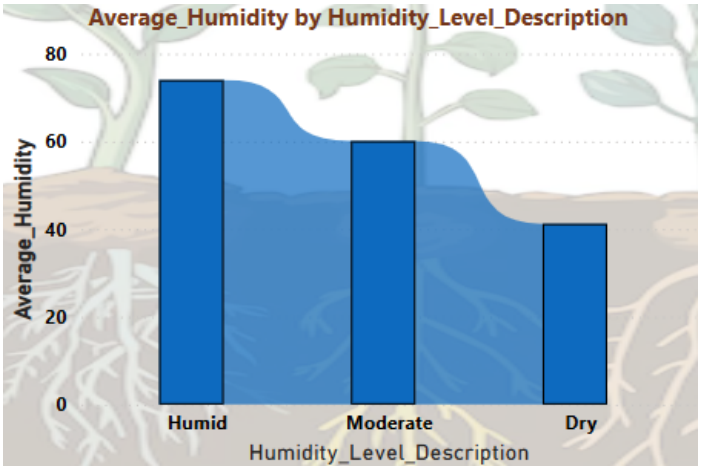
Insight: Warm temperature range increases plant growth, highlighting the positive impact of warm temperatures on plant development.

Activity 1.4: Growth Milestone Count by Fertilizer Type



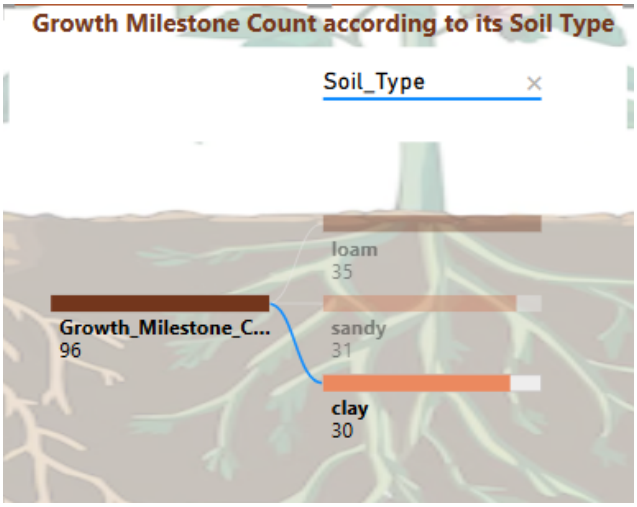
Insight : Chemical fertilizers account for the highest growth milestone count, indicating their effectiveness in promoting plant growth milestones compared to organic fertilizers or no fertilizer.

Activity 1.5: Average Humidity by Humidity Level Description



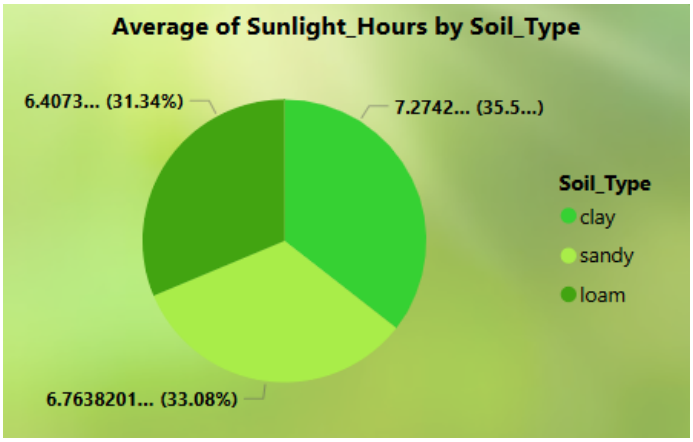
Insights : Humid conditions lead to the highest average humidity, emphasizing the importance of maintaining high humidity levels for optimal plant growth.

Activity 1.6: Growth Milestone Count According to Its Soil Type



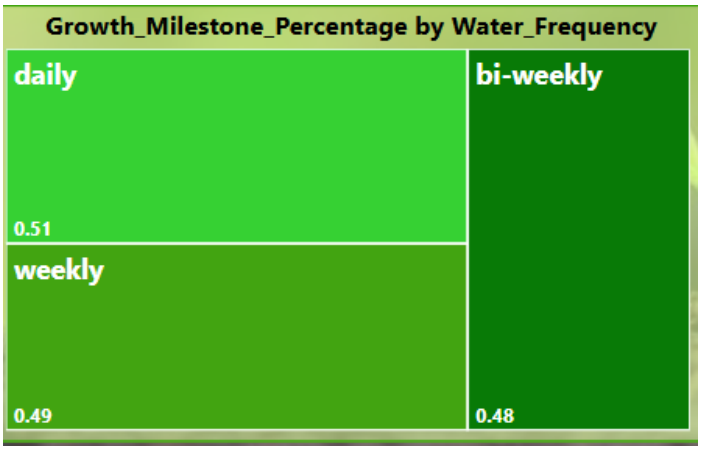
Insights : Loam soil has the highest growth milestone count, suggesting it is the most conducive soil type for achieving plant growth milestones.

Activity 1.7: Average Sunlight Hours by Soil Type



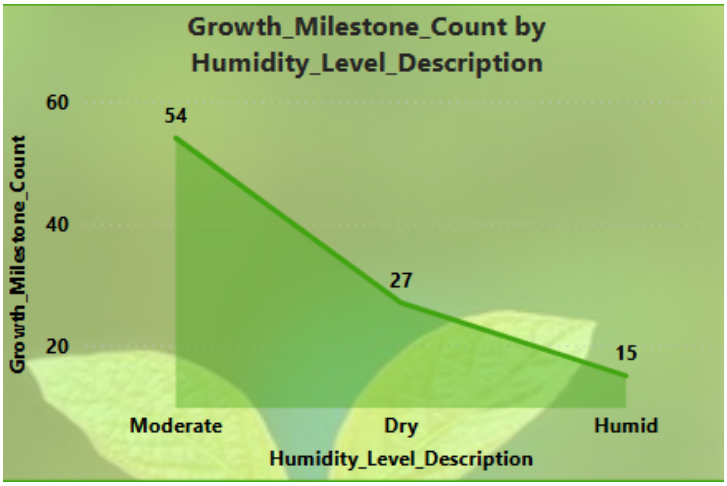
Insights : Sandy soil receives the highest average sunlight hours, which may contribute to its growth performance under sufficient light conditions.

Activity 1.8: Growth Milestone Percentage by Water Frequency



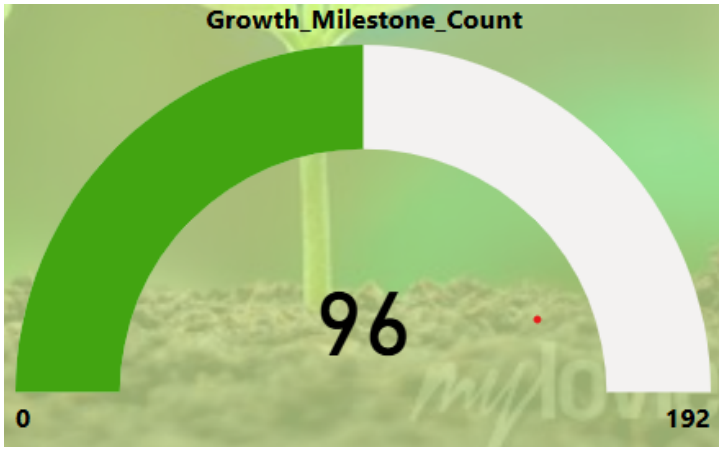
Insights : Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.

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



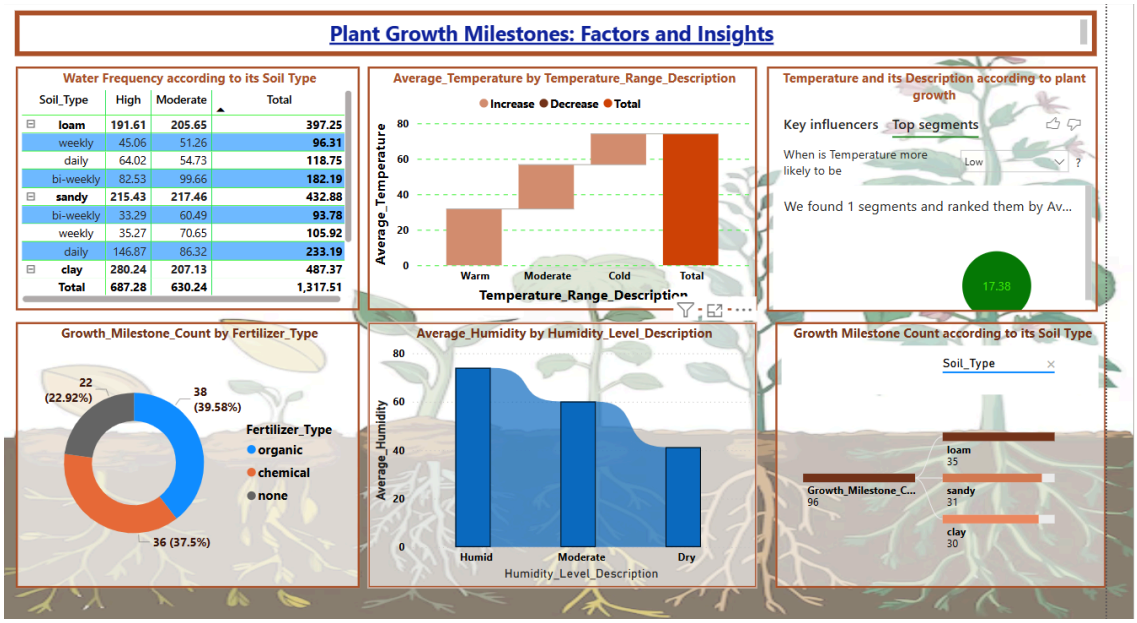
Insights: The total growth milestone count is 96, with the highest contributions coming from environments with optimal humidity and temperature conditions.

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

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

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.

Activity 1: Design of Report

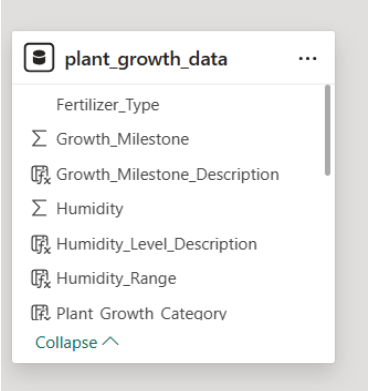
Designing a report in Power BI involves connecting to data sources, creating visualizations like charts and graphs, customizing their appearance and interactivity, organizing them logically on the canvas, formatting elements for consistency and clarity, and optionally creating dashboards for a summarized 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 6: Performance Testing

Activity 1: Amount of Data Loaded

"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 2: 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 2.11: Average Humidity as “Measure”

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

Activity 2.12: Average Sunlight Hours as “Measure”

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

Activity 2.13: Average Temperature as “Measure”

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

Activity 2.14: Growth Milestone Count as “Measure”

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

```

Activity 2.15: Growth Milestone Percentage as “Measure”

```
Growth_Milestone_Percentage =
DIVIDE(
    [Growth_Milestone_Count],
    COUNTROWS(plant_growth_data),
    0
)

```

Activity 2.16: Water Frequency Numeric as “New Column”

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

```

Activity 2.17: Temperature Range as “New Column”

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

```

Activity 2.18: Humidity Range as “New Column”

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

```

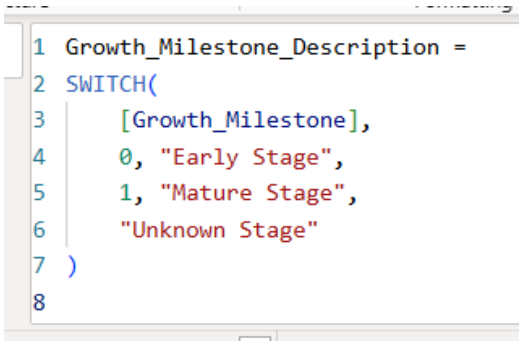
Activity 2.19: Humidity Level Description as “New Column”

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

```

Activity 2.20: Temperature Range Description as “New Column”

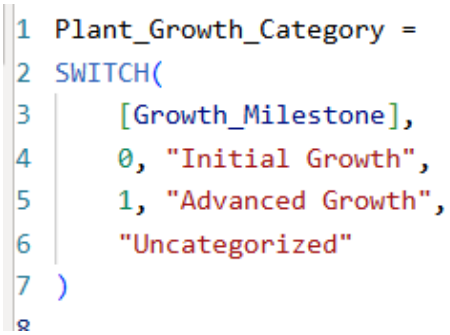
```
Temperature_Range_Description =
SWITCH(
    TRUE(),
    [Temperature] < 10, "Very Cold",
    [Temperature] >= 10 && [Temperature] < 20, "Cold",
    [Temperature] >= 20 && [Temperature] < 30, "Moderate",
    [Temperature] >= 30 && [Temperature] < 40, "Warm",
    [Temperature] >= 40, "Hot"
)

```

Activity 2.21: Growth Milestone Description as “New Column”



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

Activity 2.22: Plant Growth Category as “New Column”



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

Activity 2.2: 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

Milestone 7: Project Demonstration & Documentation

Below mentioned deliverables are submitted along with other deliverables

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

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