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

**INTRODUCTIN:**

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.

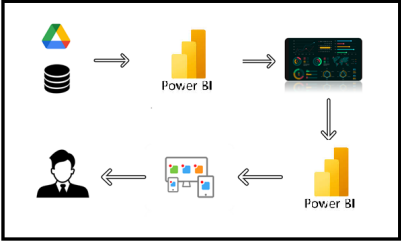
**Scenario 2:**

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    Amount of Data Rendered to DB

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

### ***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.

**Downloading the dataset**

Please use the link to download the dataset:

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

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.

### **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 :

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

* 3.2 Data Cleaning

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

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: predicting plant growth stage with envitonmental and management 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 with inflation rates 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 temperature\_rate is 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. Unpivot Year Columns :

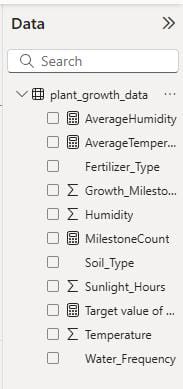
The data has inflation rates spread across multiple year columns, should unpivot these columns so that each year and inflation rate appear as separate rows. This will allow better flexibility in visualizations (like line charts, etc.).

In Power Query Editor, selected all year columns

Right-click and choose Unpivot Columns.

Renamed the generated column (usually Attribute) to Year and the other one (usually Value) to \_tempareture growth.4. Create Calculated Columns/Measures:

For more advanced analysis, such as comparing inflation rates across years or calculating Year over year changes, created calculated columns or measures using DAX. The DAX measures like Average inflation rate, minimum inflation rate, maximum inflation rate, and target inflation rate.

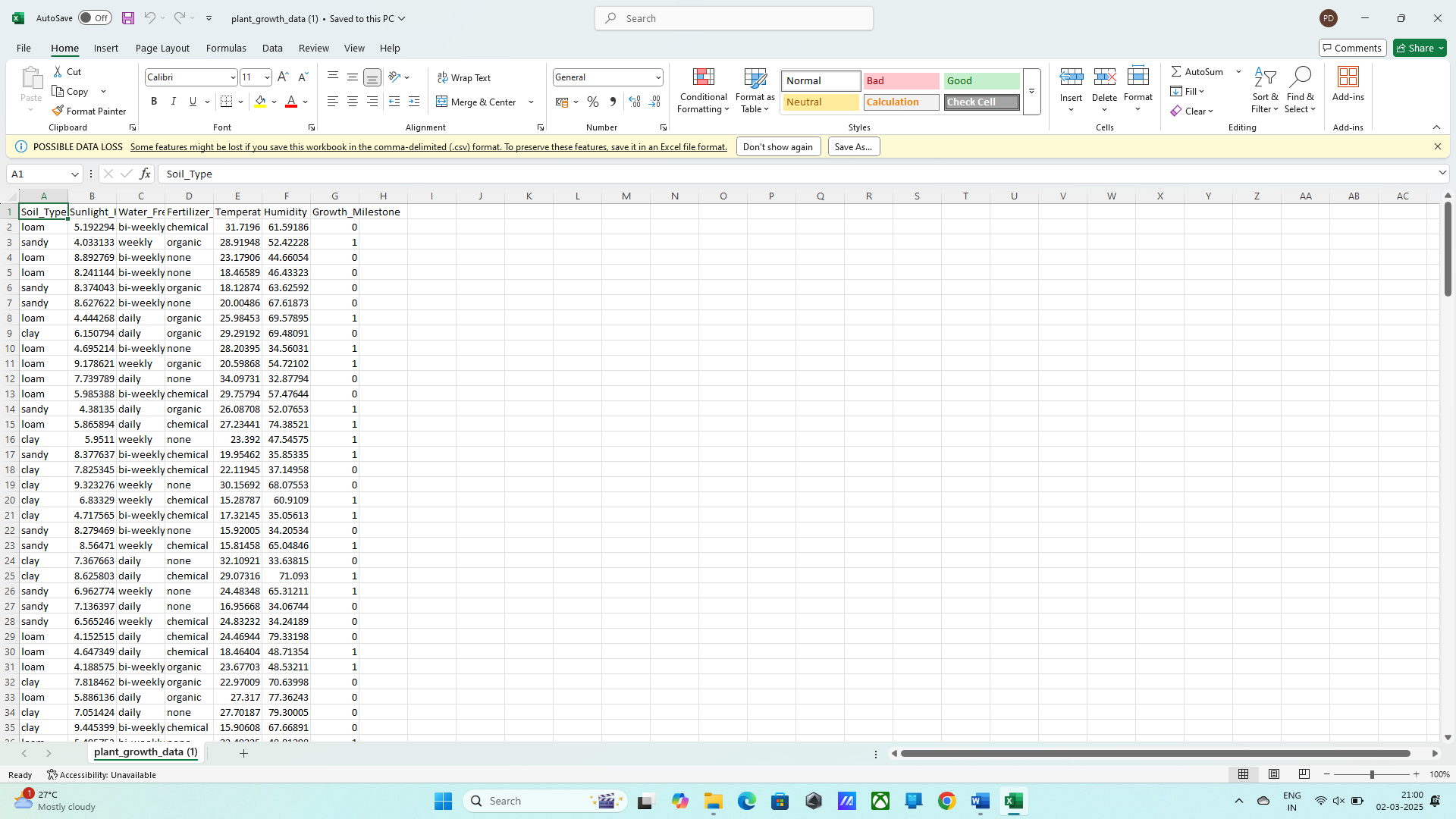


5. Save Data Preparation Steps:

Once cleaned, transformed, and unpivoted your data, click Close & Apply in the Power Query Editor to load changes into Power BI.

***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

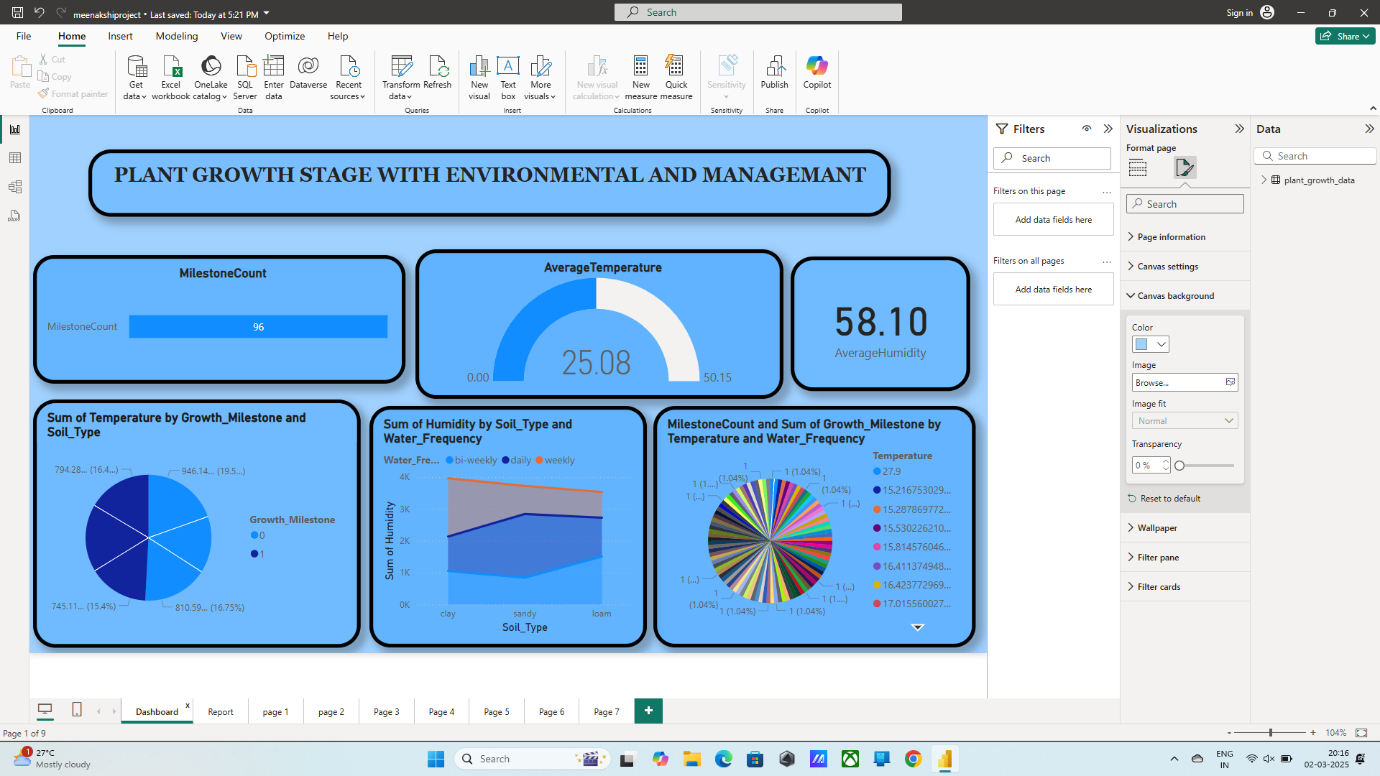


### Plant Growth Classification

Activity 1.1: Water Frequency According to Its Soil Type

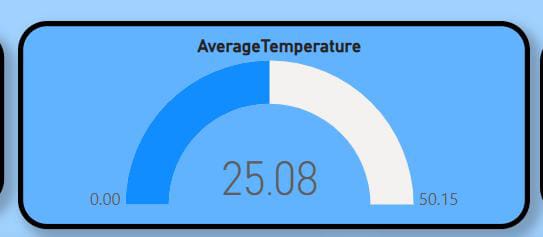
Insight: Loam soil with high water frequency results in the highest total water frequency, indicating that loam soil may require or benefit from more frequent watering compared to sandy and clay soils.

     Activity 1.2: plant growth stage with environtmental and management



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

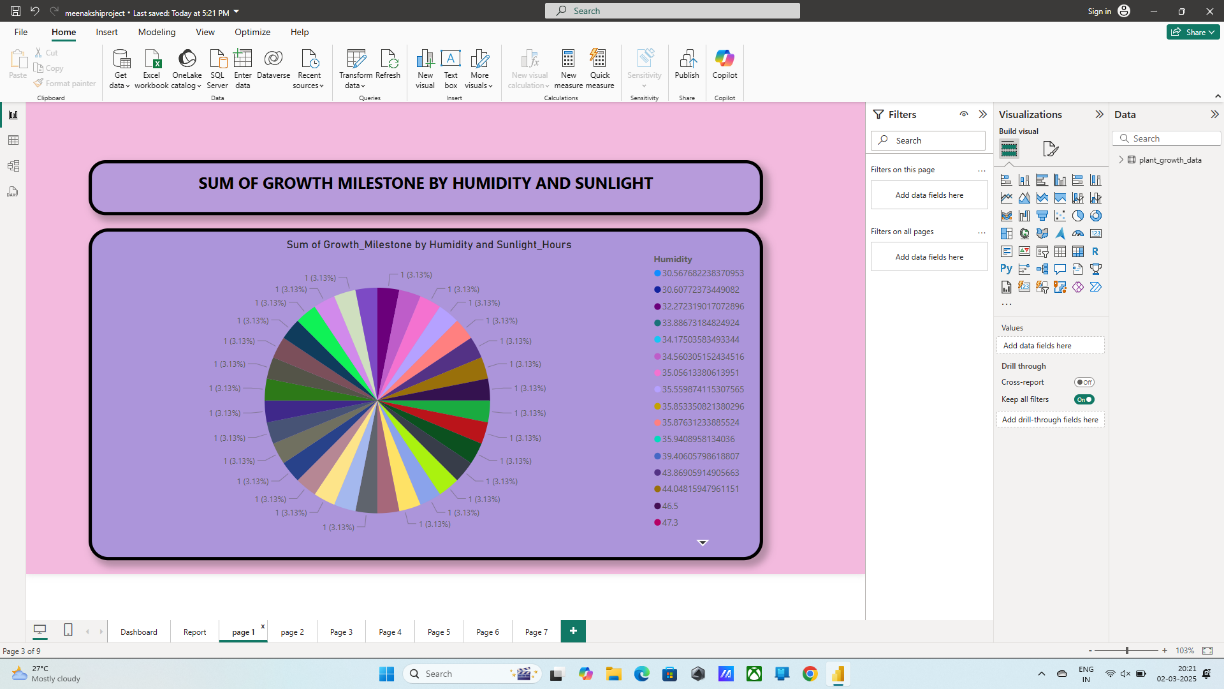
 Activity 1.3: sum of the plant growth



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

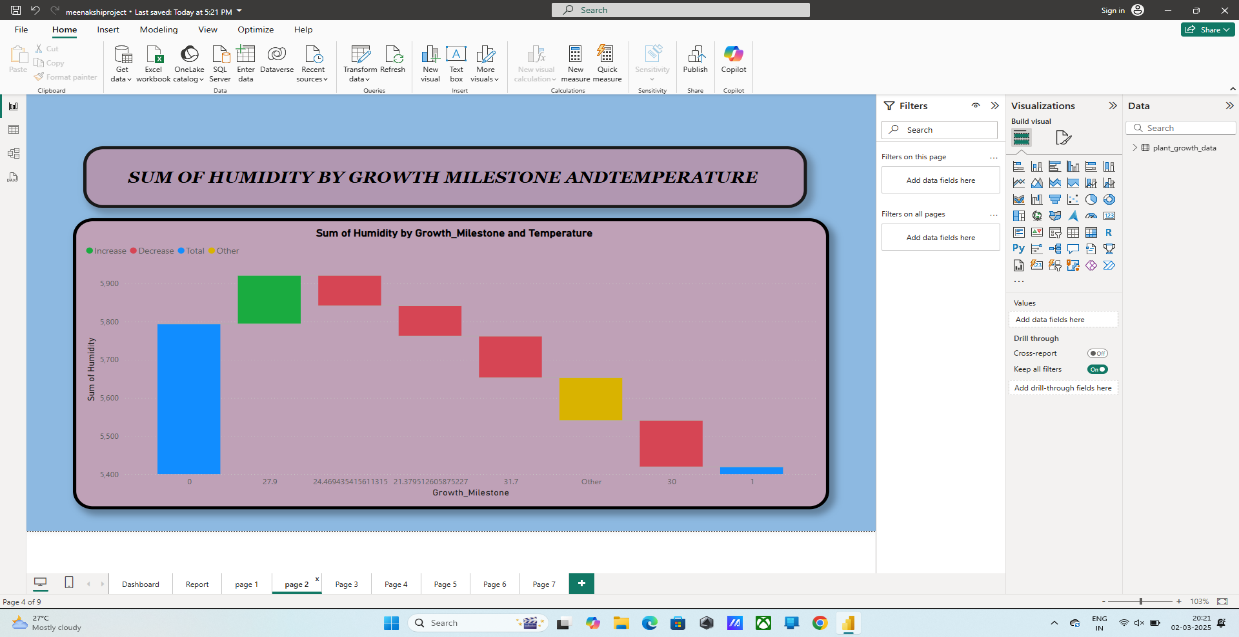
Insight: Warm temperature range increases plant growth, highlighting the positive impact of warm temperatures on plant development.

 Activity 1.4: sun Growth Milestone by humidity and sunlight



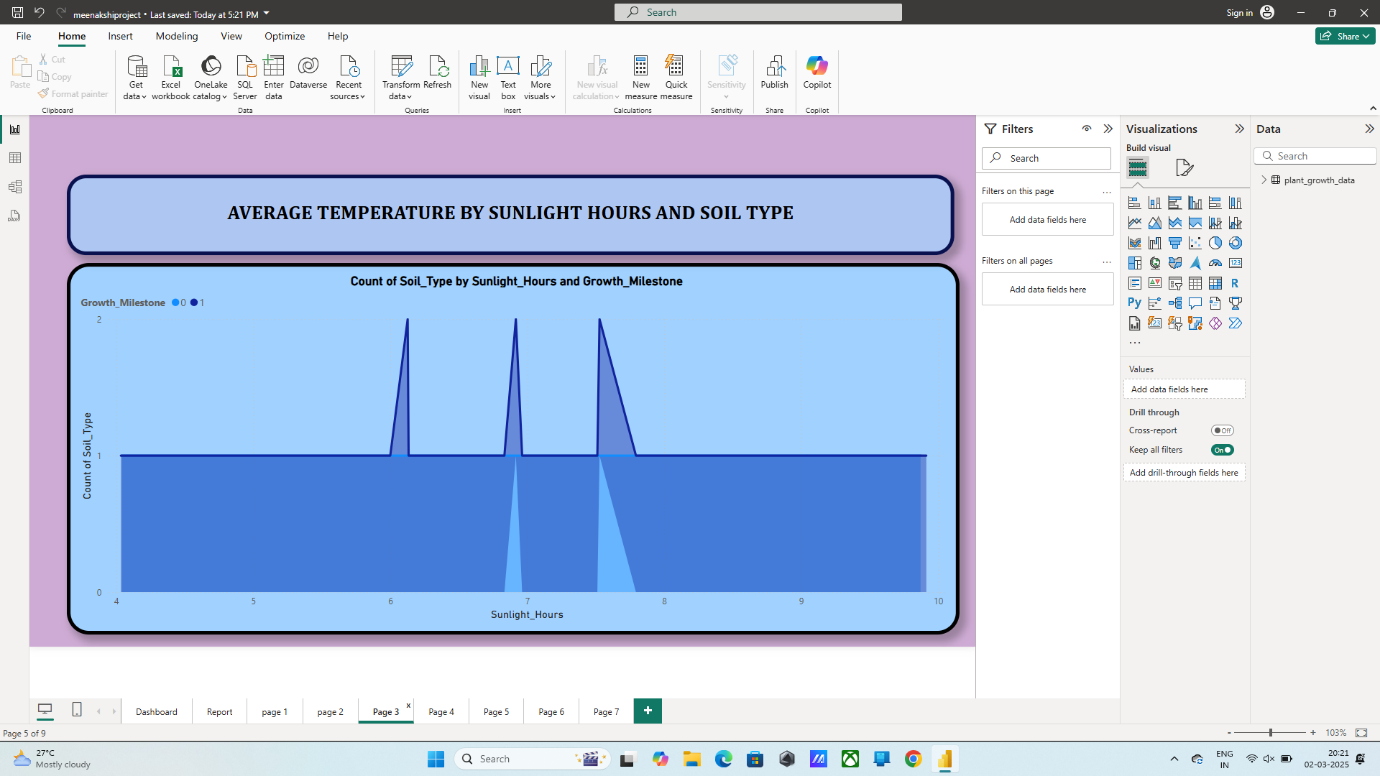
Insight : Chemical fertilizers account for the highest growth milestone count, indicating their effectiveness in promoting plant growth milestones compared to organic fertilizers or no fertilizer.

Activity 1.5:  sum of humidity by growth milestone and temperature



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

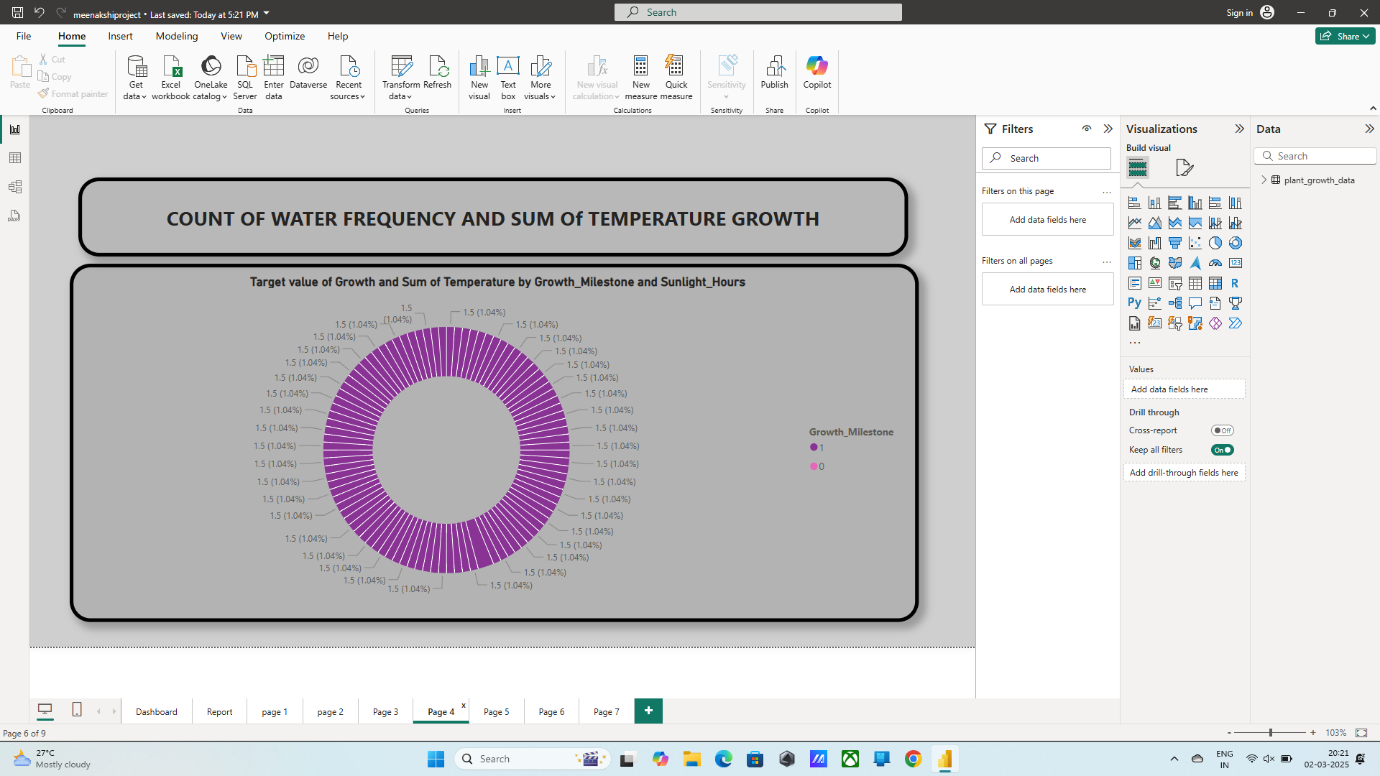
Activity 1.6: avarege temperature by sunlight hours and soil type



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

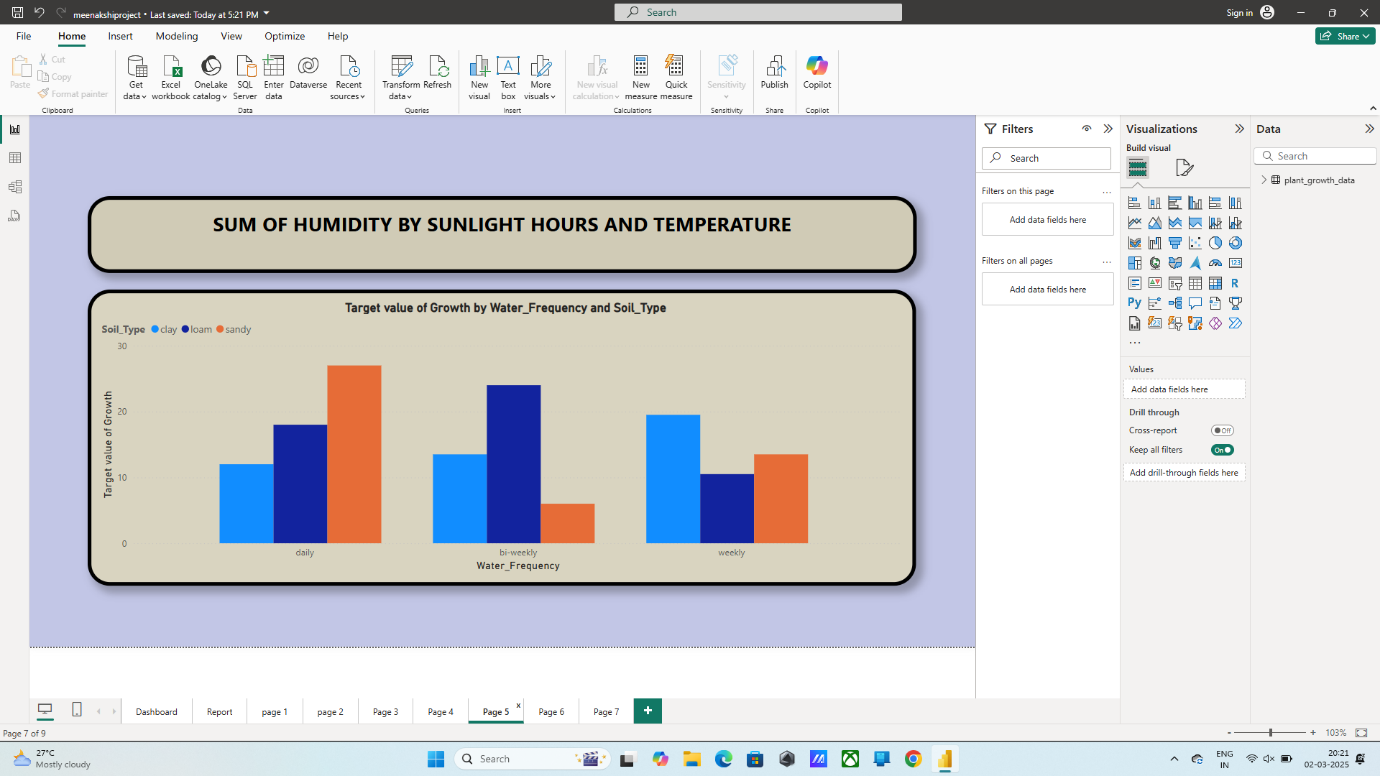
 Activity 1.7:

  count of water frequency and sum of temperature growth



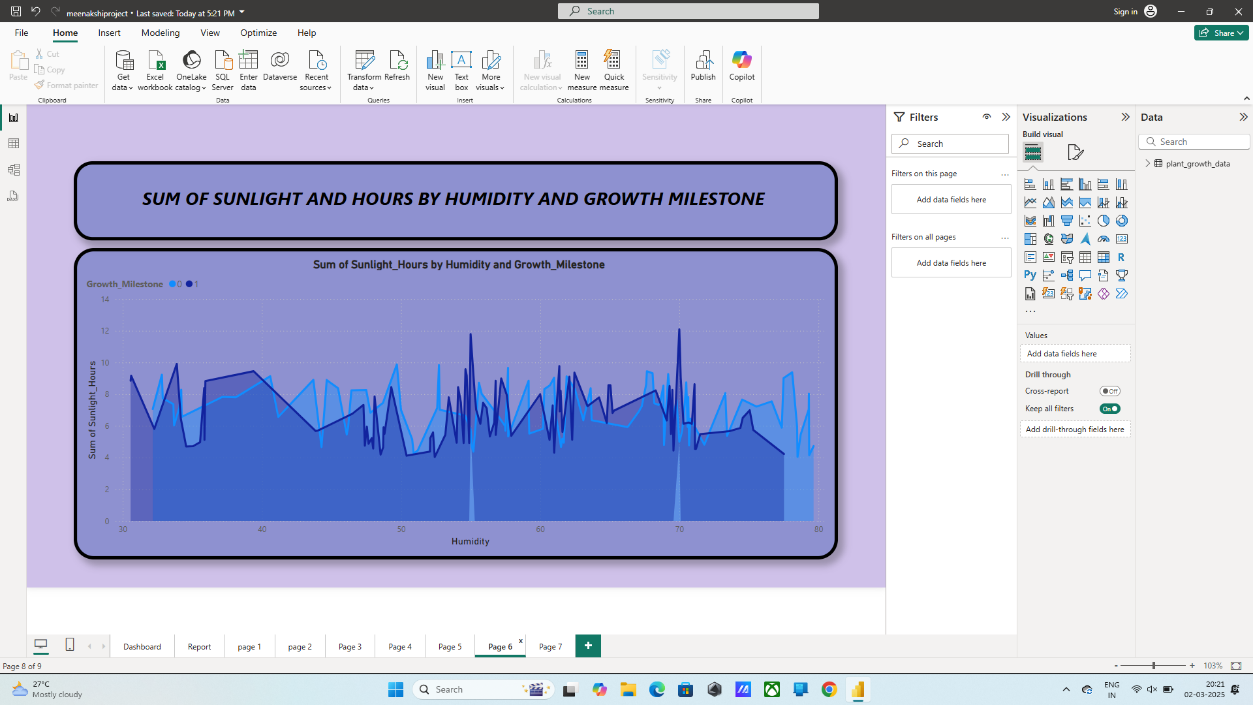
Insights : Sandy soil receives the highest average sunlight hours, which may contribute to its growth performance under sufficient light conditions.

Activity 1.8: Sum of humidity by sunlight hours and temperature



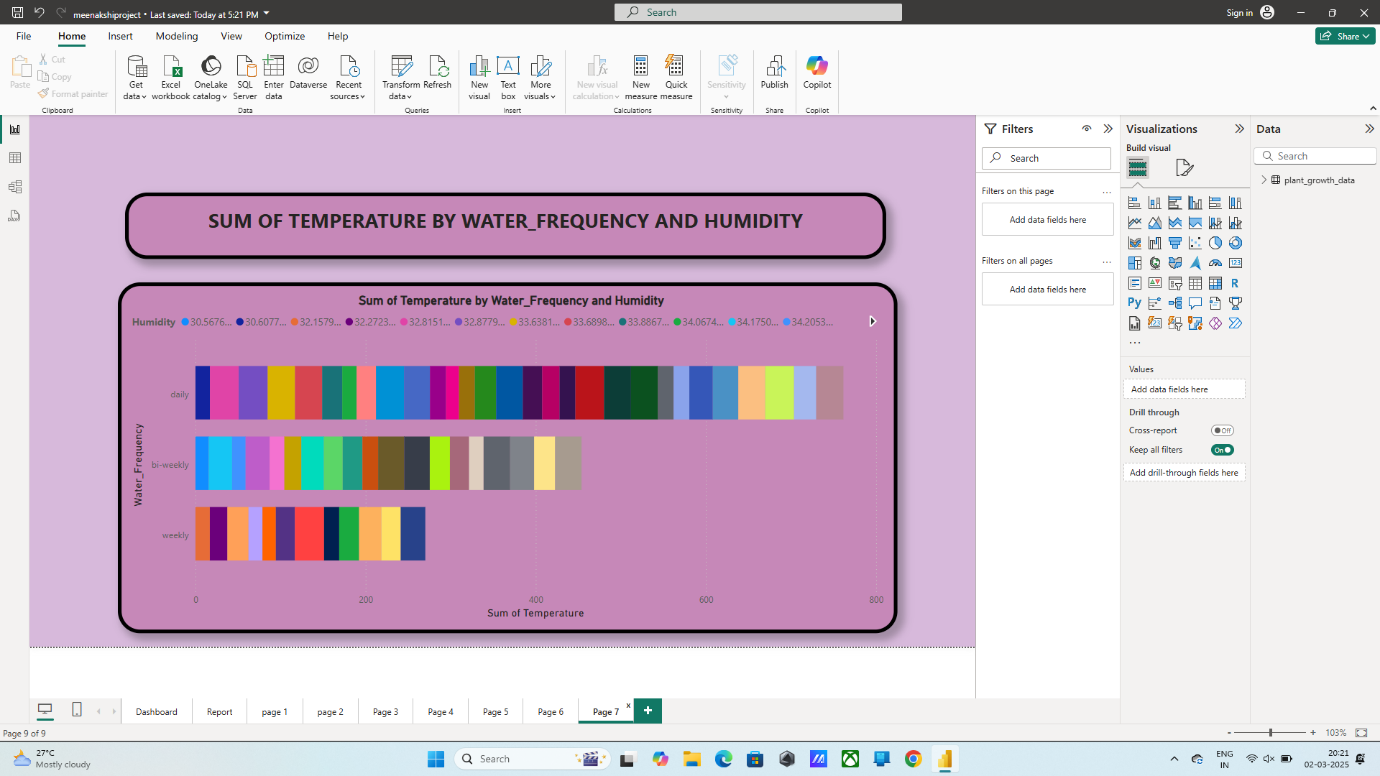
Insights : Daily watering results in the highest percentage of growth milestones, showing that frequent watering is crucial for plant growth.

**Activity 1.9**: Growth Milestone Count by Humidity Level Description

n

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

  Activity 1.10 : sum of temperature by water frequency and humidity



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

 Activity 1.11

Dashboard

: 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.

NOTE: Video Explanations for the above Visualizations are in Dashboard and Report sections.

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.

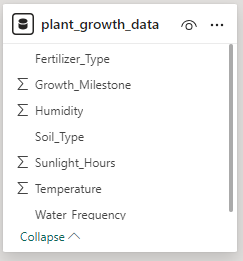
Responsive and Design of Dashboard

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 usefull

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.

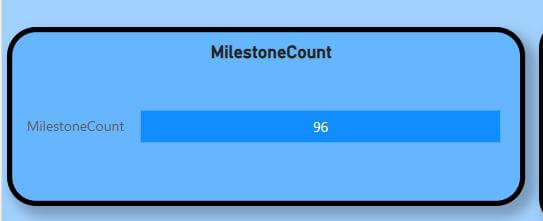
D

Save Data preparation steps :

Once cleaned, transformed, and unprivoted your data, click close & Apply in the power query editor to load changes into power BI

***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.



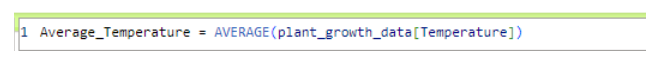
Activity 2.1: Average Humidity as “Measure”



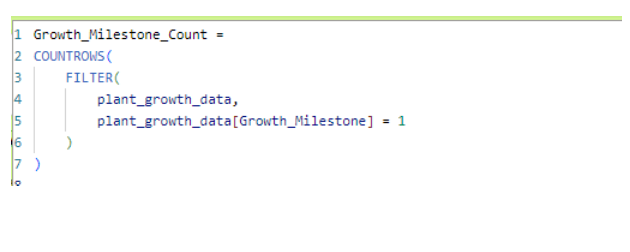
Activity 2.2: Average Sunlight Hours as “Measure”



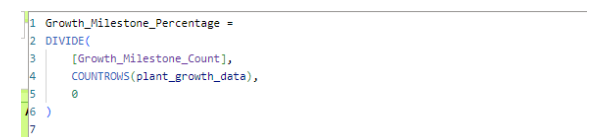
Activity 2.3: Average Temperature as “Measure”



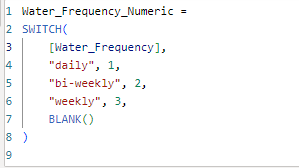
           Activity 2.4: Growth Milestone Count as “Measure”



        Activity 2.5: Growth Milestone Percentage as “Measure”



        Activity 2.6: Water Frequency Numeric as “New Column”



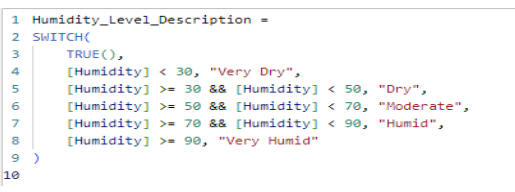
        Activity 2.7: Temperature Range as “New Column”



        Activity 2.8: Humidity Range as “New Column”



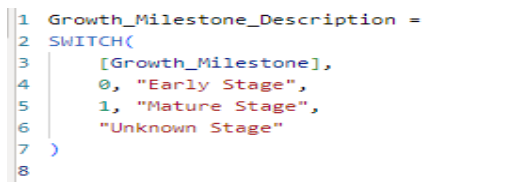
  Activity 2.9: Humidity Level Description as “New Column”



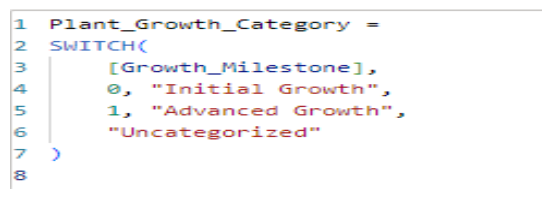
         Activity 2.10: Temperature Range Description as “New Column”



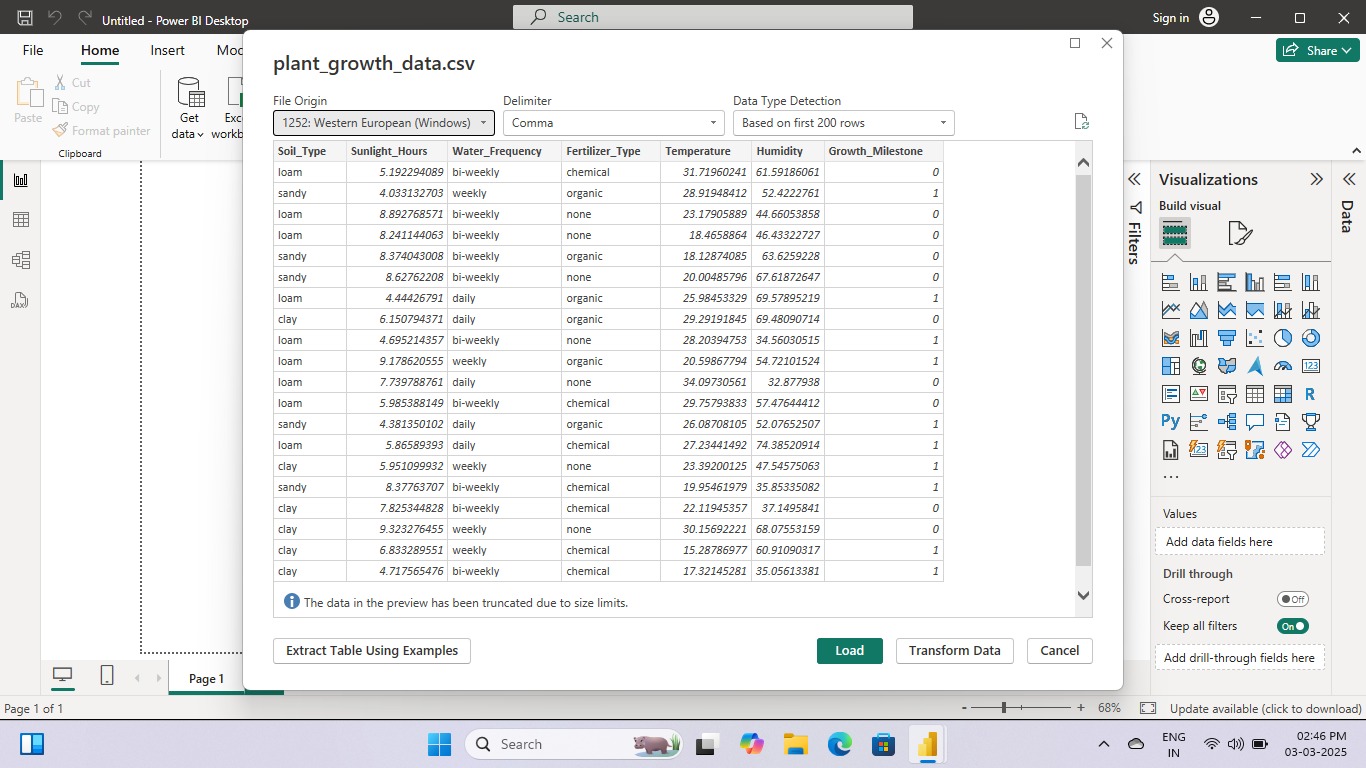
        Activity 2.11: Growth Milestone Description as “New Column”



       Activity 2.12: Plant Growth Category as “New Column”



* 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



### ## Data Collection

### 1. Environmental data:

### - Temperature

### - Humidity

### - Light intensity

### - Soil moisture

### 2. Management data:

### - Irrigation schedules

### - Fertilization records

### - Pruning practices

### - Pest/disease management

### 3. Plant growth stage data:

### - Observations of plant growth stages (e.g., seedling, vegetative, reproductive)

### ## Data Preparation

### 1. Clean and preprocess data

### 2. Handle missing values

### 3. Normalize/scale data (if necessary)

### 4. Split data into training and testing sets

### ## Machine Learning Model

### 1. Choose a suitable algorithm (e.g., decision trees, random forests, neural networks)

### 2. Train the model using the training data

### 3. Evaluate the model's performance using the testing data

### ## Power BI Implementation

### 1. Connect to data sources (e.g., databases, spreadsheets)

### 2. Create data models and measures

### 3. Build visualizations (e.g., charts, tables, maps) to display predicted plant growth stages

### 4. Use Power BI's machine learning capabilities (e.g., Power BI ML, Azure Machine Learning) to integrate the trained model

### ## Documentation

### 1. Data dictionary: describe data sources, variables, and data types

### 2. Model documentation: explain the machine learning algorithm, training data, and hyperparameters

### 3. Power BI report documentation: describe the visualizations, measures, and data models used

### 4. User guide: provide instructions for using the Power BI report to predict plant growth stages

### Example Power BI Report:

### ### Page 1: Predicted Plant Growth Stages

### 1. Map visualization: display predicted plant growth stages by location

### 2. Bar chart: show predicted plant growth stages by crop type

### 3. Table: display detailed information about predicted plant growth stages (e.g., date, temperature, humidity)

### ### Page 2: Environmental and Management Data

### 1. Line chart: display temperature and humidity trends over time

### 2. Bar chart: show irrigation and fertilization schedules

### 3. Table: display detailed information about environmental and management dat

### 4. Model Performance

### 1. Confusion matrix: evaluate the model's performance using the testing data

### 2. ROC curve: display the model's receiver operating characteristic curve

### 3. Table: display detailed information about the model's performance metrics

### This is just a general outline, and the specifics will depend on your data and requirements.

Here are some additional details:

Data Visualization

Page 1: Predicted Plant Growth Stages

1. Map Visualization: Use a shape map or a filled map to display the predicted growth stages.

- Color code the map to represent different growth stages (e.g., green for vegetative, yellow for reproductive).

- Use a legend to explain the colors used.

2. Bar Chart: Use a stacked bar chart or a clustered bar chart to display the predicted growth stages.

- Color code the bars to represent different growth stages.

- Use a legend to explain the colors used.

3. Table: Use a table visualization to display detailed information about predicted plant growth stages.

- Include columns for plant ID, growth stage, date, temperature, humidity, and other relevant information.

Page 2: Environmental and Management Data

1. Line Chart: Use a dual-axis line chart to display temperature and humidity trends over time.

- Color code the lines to represent temperature and humidity.

- Use a legend to explain the colors used.

2. Bar Chart: Use a stacked bar chart or a clustered bar chart to display irrigation and fertilization schedules.

- Color code the bars to represent different schedules.

- Use a legend to explain the colors used.

3. Table: Use a table visualization to display detailed information about environmental and management data.

- Include columns for date, temperature, humidity, irrigation, fertilization, and other relevant information

**Project Demonstration & Documentation**

Below mentioned deliverables to be submitted along with other deliverables

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

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

Create document as per the template provided