Data Collection and Preprocessing Phase

Date	23 July 2025
Team ID	Xxxxxx
Project Title	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI
Maximum Marks	10 Marks

Data Exploration and Preprocessing

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

Section	Description
Data Overview	The "Plant Growth Data Classification" dataset is designed to support the prediction and analysis of plant growth stages based on a combination of environmental and management factors. Each row in the dataset represents the condition of a particular plant at a point in time, along with the observed growth milestone it has reached. The data helps in understanding how factors like soil, sunlight, watering, fertilizer, temperature, and humidity interact to influence plant growth.
Data Cleaning	Handle missing values, duplicates, and correct errors.

Data Transformation	We used **Power Query** in Power BI for all preprocessing steps. Here's what we did:
	
	• **New Columns Created:**
	
	Water_Frequency_Numeric: A calculated column added to classify water frequency into numerics:
	
	```1 Water_Frequency_Numeric = SWITCH([Water_Frequency],"daily",1,"bi-weekly",2,"weekly",3,BLANK())```
	
	Temperature Range: A calculated column added to classify the temperature into different ranges:
	
	```1 Temperature_Range = SWITCH(TRUE(),[Temperature] <15,"low",[Temperature]>=15 && [Temperature]<25,"Moderate",[Temperature]>=25,"High")```
	
	Humidity Range: A calculated column added to classify the humidity into different ranges:
	
	```1 Humidity_Range = SWITCH(TRUE(),[Humidity]<15,"Low",[Humidity]>=15 && [Humidity]<25,"Moderate",[Humidity]>=25,"High")```
	
	Humidity_Level_Description: A calculated column added to classify the humidity into different humid levels:
	
	```1 Humidity_Level_Decsription = 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")```
	
	Temperature_Range_Description: A calculated column added to classify temperature into different temperature levels:
	
	```1 Temperatur_Range_Decsription = SWITCH(TRUE(),[Temperature]<10,"Very Cold",[Temperature]>=10 && [Temperature]<20,"Cold",[Temperature]>=20 && [Temperature]<30,"Moderate",[Temperature]>=30 && [Temperature]<40,"Warm",[Temperature]>=40,"Hot")```

Growth_Milestone_Description: A calculated column added to classify Growth_Milestone into different levels:

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

Plant_Growth_Category: A calculated column added to classify Growth_Milestone into plant growth category:

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

New Measures Created:
Average_Humidity: Calculate a measure for average of humidity.

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

Average_Sunlight_Hours: Calculate a measure for average of Sunlight_Hours.

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

Average_Temperature: Calculate a measure for average of Temperature.

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

Growth_Milestone_Count: Calculate the rows that have Growth_Milestone=1.

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

Growth_Milestone_Percentage: Calaculate the percentage of Growth_Milestone_Count from all rows.

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

Data Type Conversion	There is no need of data type conversion as all are already assigned correct datatype.
Column Splitting and Merging	There is no split or merge used in dataset.
Save Processed Data	Save the cleaned and processed data for future use.