Data Collection and Preprocessing Phase

Date	15 dec 2025
Team ID	SWUID20250236463
Project Title	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI
Maximum Marks	10 Marks

Data Exploration and Preprocessing

Section	Description
Data Overview	**The dataset used in this project contains structured plant growth information along with environmental and management factors such as soil type, sunlight hours, water frequency, fertilizer type, temperature, humidity, and growth milestone. The data was collected in Excel format and imported into Power BI for analysis.**
Data Cleaning	**Missing values in temperature and humidity columns were identified and removed. Duplicate plant records were detected and eliminated. Inconsistent categorical values in soil type and fertilizer type were standardized to ensure data accuracy.**
Data Transformation	**Power Query Editor was used to filter irrelevant records, sort data for analysis, and create calculated and conditional columns such as temperature range, humidity level description, and sunlight category to support meaningful visualization.**
Data Type Conversion	**Data types were corrected using Power Query to ensure numerical columns such as temperature, humidity, and sunlight hours were set as numeric values, while categorical fields were set as text to avoid calculation errors.**
Column Splitting and Merging	**Where required, columns were split to separate combined information, and descriptive columns were merged to create readable labels for temperature and humidity categories used in visual analysis.**

Section	Description
Data Modeling	**A single-table data model was used for this project. DAX measures such as Growth Milestone Count, Growth Milestone Percentage, Average Temperature, Average Humidity, and Average Sunlight Hours were created to support analysis and dashboard visuals.**
Save Processed Data	**After completing preprocessing and validation, the cleaned dataset was saved within the Power BI model and the final Power BI file (.pbix) was stored for future reference and reuse.**