Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	Sukriti
Project Title	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI
Maximum Marks	10 Marks

Data Exploration and Preprocessing Template

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	Name of Dataset -Plant Growth Data Classification Size of Dataset - 11.49 kB 7 Columns – 3 String, 3 Decimal, 1 Integer 193 rows
Data Cleaning	Grouping (Fields like Humidity, Temperature, and Water_Frequency were transformed into **categorical bins** (e.g., Humidity_Range, Temperature_Range, Water_Frequency_Total)), Duplicate Handling (duplicate removal have been applied), Derived Columns (Columns such as Growth_Milestone_Percentage, Water_Frequency_Total, and Humidity_Range suggest transformations or calculated columns from original numerical values.), Outlier Detection (Temperature_Range)
Data Transformation	Filtering (values like "High," "Low," "Moderate"), sorting (Power BI often auto-sorts previewed data.) and creating calculated columns. (Fields like Water_Frequency_Total, Growth_Milestone_Percentage, and Humidity_Range are clearly derived from base columns.)

Data Type Conversion	Categorical → Numeric Water_Frequency (text or categorical) Water_Frequency_Numeric (numeric version of the same)
Column Splitting and Merging	Derived (Humidity_Range, Water_Frequency_Total, Temperature_Range)
Data Modeling	Measures are created (e.g. Average_Humidity, Average_Sunlight_hours).
Save Processed Data	plant_growth_data