Data Collection and Preprocessing Phase

Date	13-01-2026
Team ID	
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 dataset used in this project is the *Plant Growth Data Classification* dataset obtained from Kaggle. It contains structured data related to environmental and management factors affecting plant growth. Key attributes include soil type, sunlight hours, water frequency, fertilizer type, temperature, humidity, and plant growth milestones. The dataset is provided in CSV format and is suitable for data analysis and visualization using Power BI.
Data Cleaning	The dataset was inspected for data quality issues such as missing values, duplicate records, and inconsistent entries. It was observed that the dataset did not contain any missing or null values and had no duplicate records. Since the data was already clean and well-structured, no rows were removed. Basic validation was performed to ensure that numerical values such as temperature, humidity, and sunlight hours were within reasonable ranges.
Data Transformation	Power Query Editor in Power BI was used to prepare the data for analysis. The following transformations were performed: • Filtering data to validate categorical values such as soil type and fertilizer type • Sorting records to verify data distribution

	• Creating calculated columns (e.g., temperature range, humidity range, water frequency numeric values) These transformations enhanced the dataset's usability for visualization and analysis.
Data Type Conversion	Appropriate data types were assigned to all columns to ensure accurate analysis. Categorical fields such as soil type, water frequency, and fertilizer type were set as *Text* data types, while numerical fields such as temperature, humidity, and sunlight hours were set as *Decimal Number*. Growth milestone values were verified and standardized to ensure consistency.
Column Splitting and Merging	Column splitting and merging were not required for this dataset, as all attributes were already available in separate and well-defined columns. The existing column structure was sufficient for analysis and visualization purposes.
Data Modeling	Data modeling was performed within Power BI using a single-table model. DAX measures were created to calculate key metrics such as average temperature, average humidity, average sunlight hours, growth milestone count, and growth milestone percentage. Calculated columns were also created to define temperature ranges, humidity levels, and plant growth categories, enabling effective analysis and decomposition tree visualization.
Save Processed Data	After completing data exploration, cleaning, and transformation, the processed dataset was saved within the Power BI .pbix file. This ensured that all transformations, calculated columns, and measures were preserved for future analysis, dashboard updates, and reporting.