

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	3 Marks

Data Quality Report

Data Source	Data Quality Issue	Severity	Resolution Plan
kaggle	Missing values observed in temperature and humidity columns for some plant records.	Moderate	Records with missing temperature and humidity values were identified using Power Query filters and removed to maintain analytical accuracy.
kaggle	Duplicate entries found for certain plant growth records.	High	Duplicate rows were detected using Power Query's duplicate removal feature and eliminated to prevent biased results.
kaggle	Inconsistent naming in categorical columns such as soil type and fertilizer type.	Low	Standardization was performed by correcting spelling variations and unifying category labels using Power Query text transformation functions.
kaggle	Incorrect data types assigned to numerical fields such as sunlight hours and water frequency.	Moderate	Data types were rectified in Power Query by explicitly setting numerical columns to whole number or decimal data types.
kaggle	Outlier values present in temperature and humidity measurements.	Low	Logical filtering was applied to remove extreme values that were outside realistic environmental ranges.