

Data Quality and Preprocessing Report – Predicting Plant Growth Stages with Environmental and Management Data

> Purpose

▪ This document details the steps taken to ensure that the dataset used in the Power BI dashboard is clean, consistent, and ready for analysis. It outlines how missing values, formatting issues, and categorical variables related to environmental and management features were handled.

> Unit Consistency & Conversion

- Converted temperature values to Celsius where necessary.
- Standardized sunlight hours and humidity as numerical fields.
- Rounded environmental values to 2 decimal places for display in Power BI.

> Data Cleaning Steps

Step	Action Taken
Column Renaming	Used readable names like Temperature_Range , Soil_Type , Growth_Milestone
Date Formatting	Ensured all date/year fields are in a consistent format
Duplicates	No duplicate records found after initial data inspection
Irrelevant Columns	Dropped fields such as plant ID, geographic codes, etc., not required for analysis
Missing Values	Handled using average imputation or removal based on field relevance
Category Mapping	Grouped soil types, water frequency, and fertilizer types into descriptive categories