Data Collection and Preprocessing Phase

Date	25 December 2025
Project Title	Predicting Plant Growth Stages with Environmental and Management Data Using Power BI.
Maximum Marks	3 Marks

Data Quality Report:

S. No.	Data Source	Column Name	Data Quality Issue	Severity	Resolution Plan
1.	plant_growth_data	All Columns	Missing Values Check	High	No missing values were found.
2.	plant_growth_data	All Rows	Duplicate Records	Medium	No duplicate rows were detected.
3.	plant_growth_data	Temperature	Outlier Validation	Low	These temperatures are normal for plant growth. No data removal required.
4.	plant_growth_data	Growth_Milestone	Class Imbalance	High	The distribution is balanced(49.7%Success vs.50.3%Failure). No re-sampling techniques required.
5.	plant_growth_data	Soil_Type	Consistency Check	Low	Standardized values found: loam, sandy, clay. No inconsistencies found.

SUMMARY:

The raw data source plant_growth_data.csv was subjected to a rigorous quality assessment. The dataset consists of 193 records and 7 attributes with zero missing values and zero duplicate records.

Key Qualities:

- **Completeness:** The data is 100% complete, requiring no imputation strategies.
- **Validity**: Numerical ranges for Temperature (15-35°C) and Humidity (30-80%) are biologically valid and require no outlier removal.
- **Reliability:** The target variable (Growth_Milestone) is perfectly balanced (~50/50 split), which is ideal for the proposed Power BI predictive analysis.
- **Conclusion:** The data is deemed **High Quality** and is approved for immediate use in the Data Modeling phase without further remediation.