**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 6 October 2025 |
| Team ID | xxxxxx |
| Project Title | Predicting Plant Growth Stages with Environmental and Management Data Using Power BI |
| Maximum Marks | 10 Marks |

**Data Exploration and Preprocessing Template**

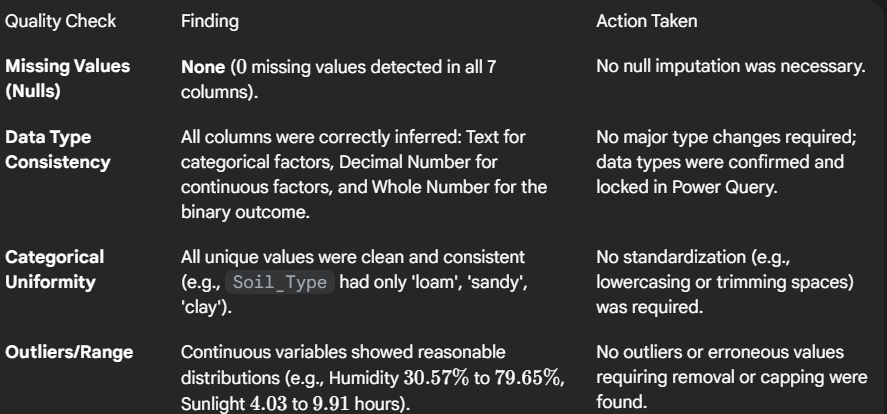
Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

1. Data Collection Plan and Raw Data Sources Identified

|  |  |
| --- | --- |
| **Element** | **Description** |
| Data Source(s) Identified | Single flat file: plant\_growth\_data.csv |
| Data Volume/Size | 193 records (rows) and 7 variables (columns). |
| Variable Types | 3 Categorical/Text (Soil\_Type, Water\_Frequency, Fertilizer\_Type). 3 Continuous/Numerical (Sunlight\_Hours, Temperature, Humidity). 1 Binary Outcome/Integer (Growth\_Milestone). |
| Acquisition Method | Direct file upload into Power BI Desktop via the "Get Data Text/CSV" connector. |

**2. Initial Data Quality Report (Power Query Analysis)**

The initial inspection was performed in the Power BI Power Query Editor before loading the data into the model.



**Conclusion on Data Quality:** The dataset is exceptionally clean and high-quality, allowing the project to proceed directly to data modeling and analysis without significant cleaning or imputation efforts.

**3. Data Exploration and Preprocessing (Modeling Steps)**

The preprocessing phase focused on transforming data for usability and effective visualization in Power BI.

