**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 24 June 2025 |
| Team ID | xxxxxx |
| Project Title | Predicting plant growth stages with environmental and management data using power bi |
| Maximum Marks | 3 Marks |

**Data Quality Report Template**

The Data Quality Report Template will summarize data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Dataset | Mention the issues faced in the selected dataset. | Low/ Moderate/ High | Give the solution for that issue technically. |
| Plant Growth Data Classification, Real-Dataset (Kaggle) | Inconsistent text formatting in categorical columns **(Soil\_Type, Water\_Frequency, Fertilizer\_Type)** | |  | | --- | | Moderate |  |  | | --- | |  | | |  | | --- | | Standardize text values using  .str.lower().str.strip() in pandas to remove  case and whitespace issues. |  |  | | --- | |  | |
| Plant Growth Data Classification, Real-Dataset (Kaggle) | |  | | --- | | No date or timestamp  Column available | | Moderate | |  | | --- | | Add a Date column during data collection  to allow time-based analysis and trend  visualization. |  |  | | --- | |  | |
| Plant Growth Data Classification, Real-Dataset (Kaggle) | |  | | --- | | **Growth\_Milestone**  Column contains  binary values only  (e.g., 0 and 1) |  |  | | --- | |  | | Low | |  | | --- | | Refine milestone tracking method to  allow a range of values for better growth  trend analysis. |  |  | | --- | |  | |
| Plant Growth Data Classification, Real-Dataset (Kaggle) | |  | | --- | | Missing engineered  features such as  environmental  impact scores |  |  | | --- | |  | | Low | |  | | --- | | Create new derived columns like  **Env\_Index = (Humidity \* Temp) / 100**  for richer analysis. |  |  | | --- | |  | |
| Plant Growth Data Classification, Real-Dataset (Kaggle) | |  | | --- | | Categorical columns not  ready for machine  learning model  input |  |  | | --- | |  | | Low | Apply Label Encoding or One-Hot Encoding to categorical fields during preprocessing. |