**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 7 August 2025 |
| Team ID | xxxxxx |
| Project Title | Predicting Plant Growth Stages Using Environmental & Management Data in Power BI |
| Maximum Marks | 2 Marks |

**Data Collection Plan & Raw Data Sources Identification Template**

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

**Data Collection Plan Template**

|  |  |
| --- | --- |
| **Section** | **Description** |
| Project Overview | This project aims to develop an interactive dashboard to monitor and analyze plant growth under varying environmental and management conditions. Data on soil type, sunlight hours, water frequency, fertilizer type, temperature, humidity, growth milestones, and growth rate is collected from agricultural records and research datasets. The dashboard visualizes trends, identifies patterns, and supports decision-making for optimized plant care. |
| Data Collection Plan | * Sources: Agricultural field experiments, greenhouse monitoring systems, IoT sensors, and research datasets. * Parameters: Soil type, sunlight hours, water frequency, fertilizer type, temperature, humidity, growth milestones, growth status, watering level. * Frequency: Daily logging for environmental factors; weekly logging for growth milestones. * Format: CSV/Excel files exported from sensors and manual records; standardized templates for uniformity. * Validation: Real-time checks for missing or abnormal values before saving data. |
| Raw Data Sources Identified | * Government Agriculture Portals – e.g., Indian Council of Agricultural Research (ICAR) datasets. * Open Data Platforms – Kaggle, FAO (Food and Agriculture Organization) crop datasets. * Research Publications – Agricultural universities’ experimental data on crop growth stages. * IoT Sensor Feeds – Temperature, humidity, and soil moisture readings from greenhouse monitoring devices. * Manual Farm Logs – Watering schedules, fertilizer application records maintained by farmers. |

**Raw Data Sources Template**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source Name** | **Description** | **Location/URL** | **Format** | **Size** | **Access Permissions** |
| Dataset 1 | 1. Plant ID – Unique identifier for each plant.  2. Soil Type – Type of soil used (e.g., sandy, loamy, clay).  3. Sunlight Hours – Average daily sunlight exposure in hours.  4. Water Frequency – Watering schedule (daily, weekly, bi-weekly).  5. Fertilizer Type – Type of fertilizer applied.  6. Temperature – Average environmental temperature in °C.  7. Humidity – Average relative humidity percentage.  8. Growth Milestone – Stages of plant development (e.g., seedling, flowering, mature).  9. Growth Status – Current condition of the plant (e.g., healthy, stressed)  10. Sunlight Category – Classified range of sunlight exposure (e.g., low, medium, high).  11. Watering Level – Quantity of water applied per session.  12. Average Sunlight – Calculated measure of sunlight exposure.  13. Growth Rate – Calculated measure of weekly plant growth percentage or height increase.. | <https://www.kaggle.com/datasets/gorororororo23/plant-growth-data-classification> | CSV | 12 KB | Public |