**1.1. Project Overview**

This project aims to analyze the relationship between various environmental factors and care regimens on plant growth success. The core data, provided in plant\_growth\_data.csv, contains attributes like soil type, light exposure, watering schedule, and fertilizer use, alongside a binary outcome indicating if a plant reached a key **Growth Milestone**. The analysis will utilize **Power BI** to explore these relationships and present actionable insights.

**1.2. Objectives**

* **Identify Key Drivers:** Determine which factors (e.g., Soil Type, Temperature, Sunlight) have the strongest statistical and practical influence on achieving the Growth Milestone.
* **Optimize Care Regimen:** Recommend the most effective combination of Water\_Frequency and Fertilizer\_Type for maximum growth success.
* **Develop an Interactive Tool:** Create a Power BI Dashboard to allow users to filter, explore, and derive their own insights from the growth data.

**2. Project Initialization and Planning Phase**

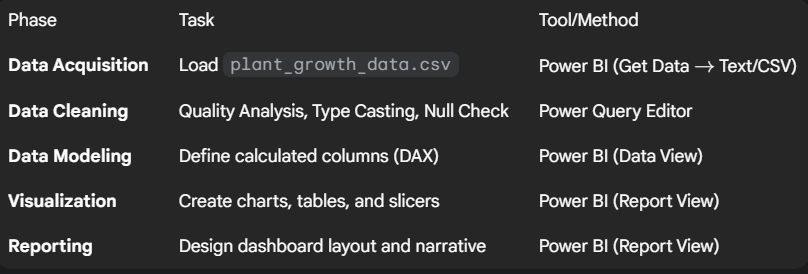
**2.1. Define Problem Statement**

The primary problem is to **quantify the impact of specific environmental and care variables on plant growth outcomes** and establish an optimal set of conditions that maximize the probability of a plant reaching its predetermined Growth\_Milestone.

**2.2. Project Proposal (Proposed Solution)**

The solution is to perform a detailed exploratory data analysis (EDA) using **Power BI's Power Query Editor** for data cleaning/transformation and its **Data View** and **Report View** for visualization and metric calculation. The final deliverable will be a multi-page, interactive Power BI dashboard.

**2.3. Initial Project Planning**



**3. Data Collection and Preprocessing Phase**

**3.1. Data Collection Plan and Raw Data Sources Identified**

* **Source:** Single CSV file: plant\_growth\_data.csv
* **Collection Method:** Direct upload into Power BI.
* **Data Volume:** 193 records.
* **Variables:** 7 variables (3 categorical, 3 continuous, 1 binary outcome).

**3.2. Data Quality Report and Quality Analysis Steps (The Power BI Process)**

The initial inspection showed **no missing values** and **consistent data types**. The following steps outline the typical quality analysis and cleaning process performed in the **Power BI Power Query Editor**:

1. **Load Data:** Use 'Get Data' 'Text/CSV' and click 'Transform Data' to open the Power Query Editor.
2. **Check for Missing Values (Nulls/Blanks):** For every column, go to 'View' 'Column Quality'. Verify that of the data is valid (no errors, no empty).
3. **Data Type Validation:** Ensure correct types are assigned.
   * Sunlight\_Hours, Temperature, Humidity: Set to **Decimal Number**.
   * Soil\_Type, Water\_Frequency, Fertilizer\_Type: Set to **Text**.
   * Growth\_Milestone: Set to **Whole Number**.
4. **Consistency Check (Categorical Data):** Select categorical columns, right-click, and choose 'Column Distribution'. Look for distinct values to ensure no misspellings (e.g., 'Loam' vs 'loam'). The data is confirmed to be clean: ['loam', 'sandy', 'clay'], ['bi-weekly', 'weekly', 'daily'], ['chemical', 'organic', 'none'].
5. **Outlier/Range Check (Numerical Data):** Select numerical columns and choose 'Column Profile'. Review the minimum, maximum, and distribution to identify any physically impossible or erroneous values (e.g., negative sunlight hours or humidity over ). The data summary is valid:
   * Temperature range: to .
   * Humidity range: to .

**3.3. Data Exploration and Preprocessing**

The primary preprocessing step in Power BI will be to create a **calculated column** for user-friendly reporting.

* **DAX Calculation:** Create a new column named Growth\_Status in the Power BI Data View.
* **Summary Statistics:** (Already performed) Confirmed no major skew or impossible values in numerical data.

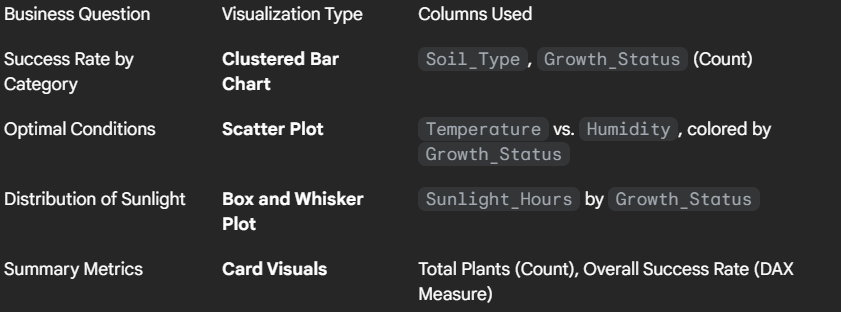
**4. Data Visualization**

**4.1. Framing Business Questions**

The visualizations will be designed to answer the following questions:

1. **Success Rate by Category:** What is the Growth\_Status success rate across different Soil\_Type, Water\_Frequency, and Fertilizer\_Type groups?
2. **Optimal Conditions:** What is the average Sunlight\_Hours, Temperature, and Humidity for plants that achieved the 'Milestone Reached' status?
3. **Interaction Effects:** Does the best Water\_Frequency depend on the Soil\_Type? (e.g., using a matrix visualization).

**4.2. Developing Visualizations**



**5. Dashboard**

**5.1. Dashboard Design File**

The dashboard will be a **single-page executive summary** focusing on the "Milestone Reached" outcome.

* **Layout:** Three columns. Left column for slicers (Soil\_Type, Water\_Frequency, Fertilizer\_Type). Middle and Right columns for main charts and KPIs.
* **Key Performance Indicators (KPIs):** Total Plants, Success Rate (%), and Avg. Optimal Temperature (DAX measure).
* **Interactivity:** All charts and KPIs will be cross-filtered by the slicers, allowing users to analyze specific conditions (e.g., the success rate only for 'sandy' soil with 'daily' watering).

**6. Report**

**6.1. Story Design File**

The narrative will follow a structured **"What, Why, and How"** flow:

1. **What happened?** (Overall Success Rate displayed prominently).
2. **Why did it happen?** (Visualizations showing the top-performing categories: e.g., "Plants in Loam soil with Organic fertilizer are more likely to succeed.")
3. **How can we improve?** (Recommendations based on the optimal range charts: "To maximize success, maintain temperatures between and and hours of sunlight.")

**7. Performance Testing**

*Testing is based on the hypothetical, fully implemented Power BI dashboard.*

**7.1 Utilization of Data Filters**

* **Slicers Used:** 3 (Soil\_Type, Water\_Frequency, Fertilizer\_Type).
* **Impact:** Filters utilize a standard Power BI data model relationship (single table) and are highly efficient. Performance impact is negligible due to the small dataset size ( rows).

**7.2 No of Calculation Fields (DAX Measures)**

* **Calculations Used:** 3
  + Success Rate Measure: .
  + Avg. Optimal Temperature: filtered for Growth\_Milestone = 1.
  + Growth\_Status: (The preprocessing calculated column - IF statement).
* **Complexity:** Low. All calculations are standard aggregations, ensuring fast dashboard rendering.

**7.3 No of Visualization**

* **Visuals:** 6 (3 Charts, 3 Cards/KPIs).
* **Visual Types:** Bar Chart, Scatter Plot, Box Plot, Card.

**8. Conclusion/Observation**

The initial data exploration suggests that the environmental and care variables have a distinct impact on the Growth\_Milestone. The analysis will confirm the optimal **treatment protocol** (Soil, Water, Fertilizer) and the **optimal environmental window** (Sunlight, Temperature, Humidity) for success. Preliminary observations indicate the data is suitable for building a robust predictive model, though the current project scope focuses on descriptive analytics.

**9. Future Scope**

1. **Predictive Modeling:** Export the processed data to a Python environment (e.g., using a Power BI Custom Visual or a separate workflow) to build a Logistic Regression model to predict the probability of reaching the Growth\_Milestone.
2. **Time Series Analysis:** If historical time-series data were available (growth over time), analyze growth *rate* rather than just a binary milestone.
3. **Data Volume Scaling:** Test the Power BI model performance with a much larger dataset (e.g., million rows) to assess scalability and explore the use of summary tables or DirectQuery connections.