1. Introduction

1.1. Project overviews

XYZ Company, a leader in agricultural innovation, has launched a project to optimize plant growth using advanced data analytics and visualization with Power BI. This initiative focuses on analyzing a comprehensive dataset that includes critical environmental and management factors such as soil type, sunlight hours, water frequency, fertilizer type, temperature, and humidity. By leveraging this data, the company's primary goal is to predict the growth milestones of plants, thereby identifying the optimal conditions for growth. The project involves creating interactive dashboards and predictive models to reveal key patterns and insights that will inform and enhance agricultural practices and greenhouse management. The analysis will utilize a decomposition tree to break down growth milestone counts by various factors, providing a clear view of each variable's impact.

1.2. Objectives

The project aims to achieve the following specific objectives:

- To create a robust data model in Power BI for analyzing plant growth.
- To develop several calculated columns and measures to enhance the dataset's analytical depth.
- To create interactive dashboards using visualizations such as clustered bar charts, pie charts, scatter plots, and column charts to present findings effectively.
- To identify the optimal combination of environmental and management factors (e.g., soil type, sunlight, watering frequency) that lead to the highest plant growth milestones.
- To provide data-driven insights that can help improve crop yields, optimize resource allocation, and promote sustainable agricultural practices.

2. Project Initialization and Planning Phase

2.1. Define Problem Statement

The project is focused on assisting farmers, agronomists, and greenhouse managers who need to predict plant growth stages accurately. These individuals are trying to optimize watering, fertilizer use, and harvesting time, but they are unable to see clear patterns between growth stages and environmental factors. This is because their data is scattered across different logs and devices, and their current tracking methods are manual and inconsistent. This problem makes them feel frustrated, uncertain, and less confident in their decision-making. The solution aims to visualize and analyze this data in Power BI, enabling them to make timely decisions to improve crop yield and resource use.

2.2. Project Proposal (Proposed Solution)

This project proposal outlines a solution to address inconsistent plant growth and sub-optimal yields.

- **Objective:** The main objective is to help farmers and greenhouse managers make timely decisions to improve yield and resource efficiency by visualizing and predicting plant growth stages.

- **Scope:** The project's scope covers the collection, integration, and visualization of environmental data (temperature, humidity, sunlight hours) and management data (watering frequency, fertilizer type, soil type) to predict plant growth stages using Power BI.
- **Approach:** The approach involves collecting sample data, cleaning and transforming it in Power BI, creating relationships between datasets, developing predictive visuals, and designing a user-friendly dashboard.
- **Key Features:** The solution will integrate environmental and management data, clean and transform data for accuracy, predict plant growth stages, and provide an interactive dashboard with filters and visuals to give insights for resource optimization.

2.3. Initial Project Planning

This section details the initial planning for the project, "Predicting Plant Growth Stages Using Environmental & Management Data in Power BI". The project is structured using a product backlog and sprint schedule.

3. Data Collection and Preprocessing Phase

3.1. Data Collection Plan and Raw Data Sources Identified

This project aims to create an interactive dashboard to analyze plant growth under various conditions. The data used includes key environmental and management factors like soil type, sunlight hours, water frequency, fertilizer type, temperature, and humidity. The goal of the dashboard is to visualize trends, identify patterns, and assist with decision-making for optimal plant care.

Data Collection Plan:

- **Sources:** Data is gathered from agricultural field experiments, greenhouse monitoring systems, IoT sensors, and research datasets.
- **Parameters:** Parameters collected include soil type, sunlight hours, water frequency, fertilizer type, temperature, humidity, growth milestones, growth status, and watering level.
- **Frequency:** Environmental factors are logged daily, while growth milestones are logged weekly.
- **Format:** The data is in CSV/Excel format, exported from sensors and manual records, using standardized templates.
- **Validation:** Real-time checks are performed for missing or abnormal values before saving the data.

Raw Data Sources Identified:

- **Government Portals:** Datasets from sources like the Indian Council of Agricultural Research (ICAR).
- **Open Data Platforms:** Platforms such as Kaggle and the Food and Agriculture Organization (FAO).
- **Research Publications:** Experimental data from agricultural universities.
- **IoT Sensor Feeds:** Readings for temperature, humidity, and soil moisture from greenhouse devices.

- **Manual Farm Logs:** Records for watering schedules and fertilizer application.

The primary dataset is from a public Kaggle link. It is in CSV format, with a size of 12 KB, and contains 13 key attributes, including plant ID, soil type, sunlight hours, and various calculated measures like growth rate.

3.2. Data Quality Report

This report summarizes data quality issues and their resolution plans.

Data Quality Issues & Resolution Plan:

- **Missing values:** Missing data for some environmental parameters like temperature and humidity. The plan is to use mean/median imputation for numeric data and forward-fill for time-series gaps.
- **Duplicate records:** Caused by repeated data logging. The resolution is to identify and delete duplicate rows using unique plant IDs and timestamps.
- **Inconsistent units:** Temperature is recorded in both °C and °F. All measurements will be standardized to a single unit, such as °C for temperature.
- **Incorrect data types:** Measurement columns may contain text instead of numeric values. The resolution is to convert text-based numbers to numeric format and ensure dates are in the proper format.
- **Misaligned table headers:** This can occur during data import. The plan is to re-import the data with proper delimiter settings or manually adjust headers.
- **Outliers:** Extreme values in growth rate can affect prediction accuracy. The resolution is to use statistical methods (e.g., IQR or Z-score) to detect and cap/replace these values.

3.3. Data Exploration and Preprocessing

This process involves assessing data quality, identifying issues, and implementing resolution plans to ensure accurate and reliable analysis.

Process Description:

- **Data Overview:** The dataset contains approximately 10,000 records with environmental and management parameters.
- **Data Cleaning:** Tasks include handling missing values, removing duplicates, and correcting inconsistent entries.
- **Data Transformation:** Power Query will be used for filtering, sorting, pivoting, and creating calculated columns like average growth rate.
- **Data Type Conversion:** Rectifying incorrect data types by converting text-based numbers to a numeric format and dates to a date format.
- **Data Modeling:** Defining relationships between environmental and management data tables and creating measures in Power BI for trend analysis.
- **Save Processed Data:** The cleaned and processed dataset will be stored in .xlsx and .csv formats for use in prediction dashboards.

4. Data Visualization

4.1. Framing Business Questions

The key business questions for this project were framed to guide the visualization process and extract meaningful insights from the data. These questions focused on understanding the relationships between environmental and management factors and their impact on plant growth.

4.2. Developing Visualizations

To answer the framed business questions, the following visualizations were developed:

1. **Which soil type has the highest average growth rate?**
   - **Visualization:** A bar chart was created to compare the growth rate (%) across different soil types.
2. **How does watering frequency impact plant growth rate?**
   - **Visualization:** A stacked bar chart was used to show the growth rate (%) by water frequency and fertilizer type.
3. **What is the distribution of soil types used in the dataset?**
   - **Visualization:** A pie chart was developed to show the proportion of clay, sandy, and loam soil types.
4. **Which fertilizer type is most effective in increasing growth rate?**
   - **Visualization:** A horizontal bar chart was used to display the average growth rate (%) by fertilizer type.
5. **How does sunlight category (medium/high) affect growth milestones?**
   - **Visualization:** A clustered bar chart was created to show the growth rate (%) by sunlight category.
6. **What is the temperature variation with respect to sunlight hours?**
   - **Visualization:** A scatter plot was used to show temperature (°C) versus sunlight hours.
7. **Which watering level is most common in the dataset?**
   - **Visualization:** A column chart was developed to show the count of observations for high, medium, and low watering levels.
8. **What percentage of plants are treated with each type of fertilizer?**
   - **Visualization:** A bar chart was used to show the percentage share of organic, chemical, and no fertilizer.

5. Dashboard

Creating an effective dashboard for this project involved organizing charts, KPIs, and visuals in a clear, interactive, and relevant manner to meet the project's objectives. The primary goal was to help users quickly interpret plant growth trends, environmental conditions, and productivity insights. The dashboard was designed with a logical layout, clear categories, and color coding for easy interpretation, and it includes features like drill-down capability.

The dashboard incorporates several key elements to fulfill these goals:

- **Plant growth trend charts:** Visualizations that show the patterns of plant growth over time.
- **Environmental parameter graphs:** Graphs illustrating key factors such as temperature, humidity, and sunlight.
- **Comparative analysis icons:** Icons that allow for easy comparison of different variables.

Based on the dashboard image provided, here are some major outcomes:

- The dashboard helps users understand which soil type has the highest average growth rate. For example, loam soil appears to have a higher growth rate than sandy or clay.
- It visualizes the impact of watering frequency and fertilizer type on growth rate.
- The dashboard shows how sunlight category (medium vs. high) affects growth milestones.
- It illustrates the temperature variation with respect to sunlight hours, revealing potential patterns.
- The dashboard provides a clear distribution of soil types, watering levels, and fertilizer usage within the dataset.

7. Performance Testing

7.1. Utilization of Data filters

The dashboard utilizes several data filters to allow for interactive analysis. These filters enable users to drill down into specific conditions and view performance based on their selections. The primary filters used are for:

- Soil_Type (e.g., clay, loam, sandy)
- Water_Frequency (e.g., bi-weekly, daily, weekly)
- Fertilizer_Type (e.g., chemical, none, organic)

7.2. Number of Calculated Fields

Your project uses several calculated fields (measures) to perform key analysis and provide meaningful insights. The following calculated fields are included:

- AvgSunlight
- Growth Rate (%)
- Growth_Milestone
- Humidity
- Sunlight_Hours
- Temperature

7.3. Number of Visualizations

The dashboard contains multiple visualizations to present the data in a clear and comprehensive manner. In total, there are **12** visualizations, including:

- **3** KPI cards at the top for AvgSunlight, Temperature, and Humidity.
- **9** charts that include clustered bar charts, a pie chart, a scatter plot, and a column chart.

8. Conclusion/Observation

Based on the analysis and a review of the three scenarios, the project demonstrates that Power BI can be a powerful tool for agricultural data analytics. The findings show a clear correlation between specific environmental conditions and plant growth milestones.

- **Scenario 1 (ABC Greenhouses):** The analysis confirms that a specific combination of loam soil, daily watering, and 6-8 hours of sunlight significantly improves growth milestones. This insight provides a clear path for standardizing practices and improving productivity.
- **Scenario 2 (GreenEarth Farms):** The project highlights the optimal conditions for organic farming, including a specific fertilizer type, soil, and watering frequency, as well as ideal temperature and humidity ranges. This allows for more consistent and predictable yields for organic crops.
- **Scenario 3 (FutureGrow Tech):** The analysis validates the effectiveness of smart farming technologies, specifically showing how real-time monitoring and adjustments can optimize growth. It also provides insights into which conditions are best suited for this technology, guiding future product development.

The project successfully provides a predictive and analytical framework to help companies enhance crop yields, optimize resource allocation, and promote sustainable agricultural practices.

9. Future Scope

To further enhance the project and build upon the insights gained from the current analysis, several potential areas for future development have been identified. These enhancements would move the project from a descriptive and diagnostic tool to a more predictive and prescriptive solution.

- **Real-Time Data Integration:** The dashboard could be upgraded to integrate with live IoT sensors in greenhouses. This would enable real-time monitoring and provide up-to-the-minute data on environmental conditions, allowing for immediate intervention.
- **Advanced Predictive Modeling:** The project could be extended to include advanced machine learning models. These models could be trained to predict not only plant growth stages but also potential yield, disease risks, or nutrient deficiencies based on the environmental and management data.
- **Mobile Accessibility:** Developing a dedicated mobile application would make the insights and alerts from the Power BI dashboard accessible to farmers and managers in the field, allowing for more agile and on-the-go decision-making.
- **Integration with External Data:** The analysis could be made more comprehensive by incorporating external data sources, such as local weather forecasts, crop market prices, or agricultural pest reports. This would provide a holistic view and enable more strategic planning.

- **Automated Actionable Insights:** The system could be enhanced to trigger automated alerts when environmental parameters fall outside optimal ranges. In the future, this could be integrated with automated greenhouse systems to adjust factors like watering or temperature without manual intervention.

10. Appendix

10.1. Source Code/Formulas

- AvgSunlight = AVERAGE(plant_growth_dataa[Sunlight_Hours])

- Growth Rate (%) =

  DIVIDE(

  CALCULATE(COUNTROWS(plant_growth_dataa), plant_growth_dataa[Growth_Milestone] = 1),

  COUNTROWS(plant_growth_dataa)

)

- GrowthStatus = IF(plant_growth_dataa[Growth_Milestone]=1,"Grown","Not Grown")

- SunlightCategory =

  SWITCH(TRUE(), plant_growth_dataa[Sunlight_Hours]< 4, "Low",plant_growth_dataa[Sunlight_Hours] <= 6, "Medium","High")

- WateringLevel =

  SWITCH(TRUE(),plant_growth_dataa[Water_Frequency] = "daily", "High",plant_growth_dataa[Water_Frequency] = "weekly", "Medium",plant_growth_dataa[Water_Frequency] = "bi-Weekly", "Low","Unknown"

)

10.2. GitHub & Project Demo Link

https://github.com/adesh1616/Predicting-Plant-Growth-Stages-Using-Environmental-Management-Data-in-Power-BI