CDS6324-DATA VISUALIZATION

TUTORIAL: TT3L

Group 17

Final Project

GROUP MEMBERS

Student Name	Student ID
TAN HUI JEEN	1211101196
TEH YVONNE	1211100571
TAN XIN YI	1211100903
TANG YU XUAN	1211103236

Dataset: Advanced IoT Agriculture 2024

Link to dataset:

- https://www.kaggle.com/datasets/wisam1985/advanced-iot-agriculture-2024/data

1. Introduction

Precision monitoring of plant growth in controlled environments is critical for optimizing crop yield and resource use in both traditional and IoT-enabled greenhouses. This proposal utilizes a comprehensive dataset gathered in 2023–2024 from Tikrit University's Agriculture Lab to develop predictive models and conduct exploratory analyses of key growth indicators under varying greenhouse conditions.

2. Data Source and Authorship

Data was collected by M. I. Lifta in the Master's thesis research at the Department of Computer Science, College of Computer Science and Mathematics, Tikrit University (2023–2024), under the supervision of Assistant Professor Wisam Dawood Abdullah, Administrator of the Cisco Networking Academy, Tikrit University. Measurements originate from two greenhouse environments—an IoT-enabled facility and a traditional (non-IoT) greenhouse—allowing comparative analysis of modern sensing technologies.

3. Dataset Overview

The dataset comprises time-stamped records of plant physiological and morphological metrics, captured at regular intervals throughout the growth cycle. Each record is uniquely identified by the categorical field **Random** (e.g., R1, R2, R3), denoting a specific sampling batch. A final categorical variable, **Class**, labels each sample according to its treatment group or observed phenotype.

This rich dataset supports a variety of analyses, including:

- **Predictive modeling** of chlorophyll content and biomass accumulation
- **Time-series analysis** of root and shoot development rates

- **Comparative studies** between IoT-monitored and traditional greenhouse performance

4. Data Dictionary

Variable	Type	Description
Random	Categorical	**Sample identifier (e.g., R1, R2, R3),** denoting random sampling batches.
ACHP	Float	**Average chlorophyll content in the plant** (μg/cm²). Indicator of photosynthetic capacity and physiological health.
PHR	Float	**Plant height rate (cm/day).** Reflects vertical growth dynamics over the measurement interval.
AWWGV	Float	**Average wet weight of vegetative growth (grams).** Proxy for above-ground biomass and water content.
ALAP	Float	**Average leaf area (cm²).** Determines photosynthetic surface area available for light capture.
ANPL	Float	**Average number of leaves per plant.** Correlates with photosynthetic capacity and overall vigor.
ARD	Float	**Average root diameter (mm).** Impacts water and nutrient uptake efficiency.
ADWR	Float	**Average dry weight of roots (grams).** Biomass measure after water removal; indicates structural and storage capacity.
PDMVG	Float	**Percentage of dry matter in vegetative growth (%).** Denotes proportion of non-water biomass above ground.
ARL	Float	**Average root length (cm)**. Influences soil exploration for resources.

AWWR	Float	**Average wet weight of roots (grams).** Total root biomass including water content.
ADWV	Float	**Average dry weight of vegetative parts (grams).** Structural biomass above ground after dehydration.
PDMRG	Float	**Percentage of dry matter in roots (%)**. Proportion of root biomass that is non-water, indicating root health and function.
Class	Categorical	**Treatment or phenotype category** for each sample (e.g., "IoT Greenhouse," "Traditional Greenhouse," or experimentally defined classes).

5. Project Description

The dashboard will feature ten coordinated visualizations: Average Chlorophyll Over Time, a dual‑line plot comparing ACHP in IoT vs. traditional greenhouses; Plant Height Rate by Batch, a stacked area chart of PHR for samples R1–R3; Vegetative Biomass Comparison, side‑by‑side bar charts of wet (AWWGV) vs. dry (ADWV) vegetative weight; Growth Distribution, box plots of PHR, leaf area (ALAP), and root length (ARL); Feature Scatter-Plot Matrix, a grid of scatter plots for pairs like ACHP vs. ALAP, ARL vs. AWWR, and ANPL vs. AWWGV with brushing and linking; Correlation Heatmap, a matrix of Pearson coefficients among all continuous variables; Dry Matter Violin Comparison, violin plots of vegetative (PDMVG) and root (PDMRG) dry‑matter percentages; Sample Class Proportions, a donut chart showing the frequency of each treatment or phenotype; Water-Use Efficiency Over Time, line charts of ADWV/AWWGV and ADWR/AWWR ratios; and Root-to-Shoot Ratio, box plots of ADWR/ADWV by greenhouse. All views are tied together with hover tooltips, click-to-filter, zoom/pan, and real‑time controls (date slider, greenhouse and batch selectors).

Together, these visualizations enable agronomists to rapidly spot temporal trends (e.g., chlorophyll peaks and growth spurts), compare biomass allocation and water‑use efficiency, assess variability and outliers in key growth metrics, and uncover multivariate relationships through linked brushing and correlation mapping. By turning static measurements into an interactive, exploratory environment—with on-demand tooltips for precise values and export/annotation tools for reporting—researchers can efficiently test hypotheses, isolate stress or optimal conditions, and derive actionable insights to optimize greenhouse management.