FROM SEED TO HARVEST: A DATA-DRIVEN JOURNEY THROUGH LETTUCE GROWTH OPTIMIZATION

BY

OYIBO OGHENENYERHOVWO DORCAS

INTRODUCTION

In the production of one of the most nutritious and popular vegetables — lettuce — every degree of temperature, every percentage of humidity, and every pH fluctuation tells a story. This is the story of how data analytics transformed raw numbers into actionable insights for optimizing lettuce production at Vertical Acres Farm in August and September 2023.

Image I: Hydroponic image of a lettuce farm land

Stakeholders: Suppliers, farmers, pre-harvest contractors, collection agents, processors, traders, retailers, and exporters.

Objective:

To investigate the relationships among temperature, humidity, TDS value, pH level, and growth days to understand how these environmental factors influence lettuce growth.

Problem Statement:

As global food security challenges intensify and urban agriculture expands, understanding the complex interactions between environmental variables and plant growth has become critical. This study analyzes 3,169 lettuce growth records to uncover patterns that distinguish thriving crops from underperforming ones, with the goal of improving yield and profitability through data-driven insights.

Key Datasets and Methodologies:

- **Dataset**: Customer purchase and growth records segmented by gender and age.
- **Tools:** Microsoft Excel (Pivot Tables, Pivot Charts, Slicers, Filters).
- **Visuals:** Bar, Column, Pie, and area charts for analytical visualization.

STORY OF DATA

Data Source:

Obtained from Kaggle — *Lettuce Growth Days Analysis Dataset* — a public dataset for educational and research purposes.

Data Collection:

Collected in **August and September 2023**, anonymized to maintain confidentiality and prevent redistribution of proprietary content. The data was processed and cleaned prior to publication.

Data Structure:

The dataset contains **12 columns** and **3,169 records**, including:

- *Plant_ID, Date, Temperature (°C), Humidity (%), TDS (ppm), pH Level, Growth Days, Temperature (°F), Humidity,* and additional derived columns (*Day Name, Growth Stage, Nutrient Adequacy*).

Feature Overview:

- **pH:** Determines nutrient absorption efficiency.
- **TDS:** Measures dissolved solids, indicating nutrient strength.
- **Humidity:** Influences transpiration and water uptake.
- **Temperature:** Affects growth rate and leaf development.
- **Growth Days:** Tracks progression from seedling to maturity.

Limitations:

1. Data covers only two months (August–September), not all seasons.
2. The geographical location of the dataset is unspecified.
3. The specific lettuce species are not identified.

Why This Analysis Matters: The Farmer's Perspective

According to research from the University of Arizona Controlled Environment Agriculture Center, precision agriculture through data analytics can increase crop yields by 20–30% while reducing resource waste by up to 40%. For vertical farms operating on tight margins, these improvements can mean the difference between profitability and failure.

The global indoor farming market size was estimated at USD 42.08 billion in 2024 and is projected to reach USD 88.48 billion by 2030 whereby lettuce under the Crop Category Outlook (Revenue, USD Million, 2018–2030) section is the second of the most profitable crops due to its short growth cycle and high demand. However, the controlled environment advantage only materializes when growers can fine-tune conditions with precision — exactly what this analysis aims to achieve.

Key benefits for farmers:

- Predictable harvests through optimized growing conditions.
- Reduced crop loss by identifying stress factors early.
- Resource efficiency via targeted nutrient and climate management.
- Quality consistency that commands premium pricing.
- Competitive advantage in an increasingly data-driven industry.

This analysis also serves multiple stakeholders in the lettuce supply chain:

- Farmers/Growers: Direct operational insights for yield optimization.
- Pre-harvest Contractors: Quality prediction and timing optimization.
- Processors: Supply consistency for production planning.
- Retailers: Predictable inventory and quality standards.
- Suppliers: Demand forecasting for nutrients and equipment.
- Traders/Exporters: Market timing and volume projections.

DATA PREPARATION AND PREPROCESSING

Data Cleaning:

- Ensured all numeric fields were standardized to two decimal places.
- Verified consistency and converted data into an Excel table format.
- Confirmed no duplicates or missing values.

Data Transformations:

Three calculated columns were created to enhance interpretability:

a. Day Name Classification — Formula: =TEXT([@Date],"dddd")

- *Purpose:* Identify weekly performance trends.

b. Growth Stage Categorization — Formula:
=IFS([@[Growth Days]]<=14,"Seedling",[@[Growth Days]]<=35,"Vegetative",[@[Growth Days]]<=50,"Maturity")

- *Purpose:* Classify plant growth phases.
- Reference: [How Many Days Lettuce Grow: Tips for Growth]

c. Nutrient Adequacy Assessment — Formula:
=IF([@[TDS Value (ppm)]]<560,"Low",IF([@[TDS Value (ppm)]]<=800,"Optimal","High"))

- *Purpose:* Evaluate nutrient sufficiency levels.
- Reference: [pH and TDS Guidelines for Lettuce — Green Packs]

Data Splitting:

- **Independent Variables:** Date, Day Name, Growth Stage, Nutrient Adequacy.
- **Dependent Variables:** Temperature (°C), Humidity (%), TDS (ppm), pH, Growth Days.

Industry Context: Agriculture — specifically indoor and hydroponic lettuce farming. **Value to Stakeholders:** Enables better control of environmental factors to increase yield and revenue.

These data transformations allowed me to segment the dataset by meaningful biological phases rather than arbitrary date ranges. A snapshot of the dataset before data cleaning and after data cleaning as seen below:

Plant_ID	Date	Temperature (°C)	Humidity (%)	TDS Value (ppm)	pH Level	Growth Days	Temperature (F)	Humidity
1	03/08/2023	33.4	53	582	6.4	1	92.12	0.53
1	04/08/2023	33.5	53	451	6.1	2	92.3	0.53
1	05/08/2023	33.4	59	678	6.4	3	92.12	0.59
1	06/08/2023	33.4	68	420	6.4	4	92.12	0.68
1	07/08/2023	33.4	74	637	6.5	5	92.12	0.74
1	08/08/2023	32.3	77	478	6.8	6	90.14	0.77
1	09/08/2023	32.3	75	682	6	7	90.14	0.75
1	10/08/2023	22.7	63	576	6.3	8	72.86	0.63
1	11/08/2023	31.9	69	662	6.1	9	89.42	0.69
1	12/08/2023	30.2	59	607	6.2	10	86.36	0.59
1	13/08/2023	30.1	77	670	6.5	11	86.18	0.77
1	14/08/2023	30.1	54	535	6.4	12	86.18	0.54
1	15/08/2023	30.1	78	480	6	13	86.18	0.78
1	16/08/2023	29.8	56	688	6.4	14	85.64	0.56
1	17/08/2023	29.6	62	653	6.6	15	85.28	0.62
1	18/08/2023	30.9	70	742	6.1	16	87.62	0.7
1	19/08/2023	30.8	57	580	6.2	17	87.44	0.57
1	20/08/2023	30.9	78	789	6.5	18	87.62	0.78
1	21/08/2023	30.2	63	516	6	19	86.36	0.63
1	22/08/2023	30.5	51	527	6.4	20	86.9	0.51
1	23/08/2023	30.8	57	669	6.2	21	87.44	0.57
1	24/08/2023	31.3	73	645	6.6	22	88.34	0.73
1	25/08/2023	31.4	52	744	6.8	23	88.52	0.52
1	26/08/2023	31.4	61	719	6.4	24	88.52	0.61

Image II: Dataset Before Data Cleaning

Plant_ID	Date	Days Name	Temperature (°C)	Humidity (%)	TDS Value (ppm)	pH Level	Growth Days	Temperature (F)	Humidity	Growth Stage	Nutrient Adequacy
1	07/08/2023	Monday	33.40	74.00	637.00	6.50	5	92.12	0.74	Seedling	Optimal
1	14/08/2023	Monday	30.10	54.00	535.00	6.40	12	86.18	0.54	Seedling	Low
1	21/08/2023	Monday	30.20	63.00	516.00	6.00	19	86.36	0.63	Vegetative	Low
1	28/08/2023	Monday	29.40	71.00	654.00	6.40	26	84.92	0.71	Vegetative	Optimal
1	04/09/2023	Monday	31.70	55.00	517.00	6.00	33	89.06	0.55	Vegetative	Low
1	11/09/2023	Monday	31.20	79.00	672.00	6.30	40	88.16	0.79	Maturity	Optimal
2	07/08/2023	Monday	33.40	50.00	488.00	6.40	5	92.12	0.5	Seedling	Low
2	14/08/2023	Monday	30.10	65.00	746.00	6.80	12	86.18	0.65	Seedling	Optimal
2	21/08/2023	Monday	30.20	51.00	704.00	6.30	19	86.36	0.51	Vegetative	Optimal
2	28/08/2023	Monday	29.40	57.00	780.00	6.40	26	84.92	0.57	Vegetative	Optimal
2	04/09/2023	Monday	31.70	61.00	498.00	6.00	33	89.06	0.61	Vegetative	Low
2	11/09/2023	Monday	31.20	56.00	797.00	6.80	40	88.16	0.56	Maturity	Optimal
3	07/08/2023	Monday	33.40	65.00	735.00	6.00	5	92.12	0.65	Seedling	Optimal
3	14/08/2023	Monday	30.10	80.00	569.00	6.30	12	86.18	0.8	Seedling	Optimal
3	21/08/2023	Monday	30.20	58.00	714.00	6.70	19	86.36	0.58	Vegetative	Optimal
3	28/08/2023	Monday	29.40	75.00	535.00	6.10	26	84.92	0.75	Vegetative	Low
3	04/09/2023	Monday	31.70	50.00	516.00	6.50	33	89.06	0.5	Vegetative	Low
3	11/09/2023	Monday	31.20	73.00	460.00	6.70	40	88.16	0.73	Maturity	Low
3	18/09/2023	Monday	29.70	72.00	549.00	6.60	47	85.46	0.72	Maturity	Low
4	07/08/2023	Monday	33.40	57.00	726.00	6.10	5	92.12	0.57	Seedling	Optimal
4	14/08/2023	Monday	30.10	64.00	525.00	6.20	12	86.18	0.64	Seedling	Low
4	21/08/2023	Monday	30.20	71.00	502.00	6.70	19	86.36	0.71	Vegetative	Low
4	28/08/2023	Monday	29.40	56.00	437.00	6.70	26	84.92	0.56	Vegetative	Low
4	04/09/2023	Monday	31.70	71.00	485.00	6.20	33	89.06	0.71	Vegetative	Low
4	11/09/2023	Monday	31.20	50.00	491.00	6.20	40	88.16	0.5	Maturity	Low

Image III: Dataset After Data cleaning

The following dashboard visualizes these patterns across all growth stages:

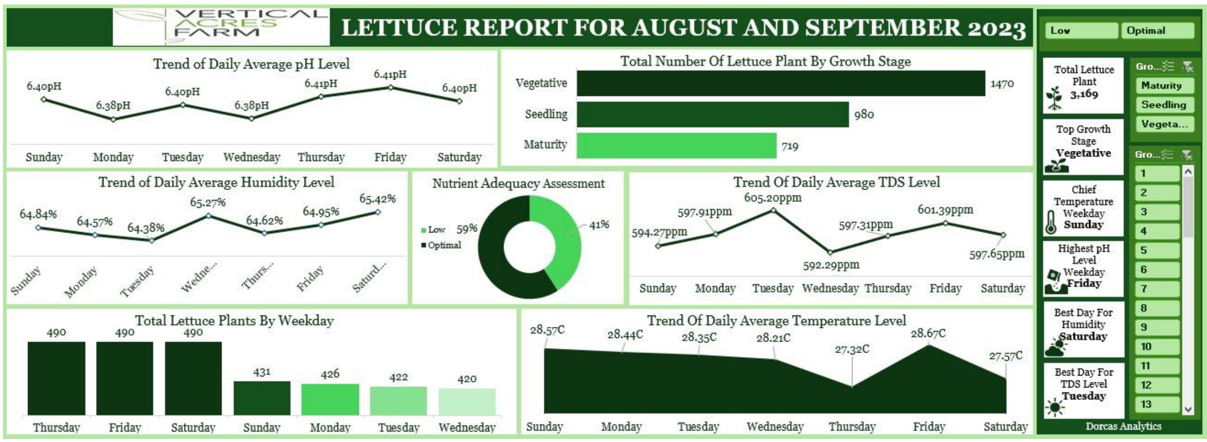

Image IV: Dashboard of Lettuce Report for August and September 2023

PRE-ANALYSIS

To uncover actionable insights, I structured the investigation around seven key questions:

1. What is the distribution of lettuce plants across different days of the week?
2. How does average pH level vary by day?
3. What are the TDS patterns throughout the week?
4. How do temperature conditions fluctuate across planting days?
5. What humidity trends emerge day-to-day?
6. What is the plant population breakdown by growth stage?
7. How does nutrient adequacy distribute across the week?

a. Key Trends

- Thursday and Friday recorded the highest average pH levels (6.41).
- The *Vegetative Stage* dominated with **1,470 plants**, showing peak growth activity.
- Friday had the highest average temperature (28.67°C).

b. Potential Correlations

- Late-week increases in temperature and pH may correlate with improved nutrient absorption.
- A midweek drop in humidity suggests environmental imbalances impacting nutrient uptake.

c. Initial Insights

- Vegetative plants appear to thrive under slightly higher pH and temperature conditions.
- Environmental consistency across the week may be influencing overall growth rates.

IN-ANALYSIS

a. Unconfirmed Insights

- Early-week low pH levels may indicate insufficient nutrient availability.
- Friday's high temperature and nutrient levels could explain better plant health.
- The vegetative stage's strong representation suggests smooth transitions from seedlings.
- TDS levels in seedling stages may need to be lower to avoid nutrient stress.
- Equal planting frequency on Thursday–Saturday (490 plants each) might be linked to accumulated early-week farm prep work.

b. Recommendations

- Investigations should examine seasonal temperature variations, as cooler conditions often increase lettuce sweetness and potassium concentration.
- 59% of available nutrients are within optimal range.
- Wednesday recorded the second-highest humidity level (65.27%).
- Educate farm laborers on essential nutrients per growth stage.
- Conduct soil and nutrient tests weekly — before planting and after harvest — to ensure quality consistency.
- Optimal humidity per growth stage should be further tested to maintain consistent performance.

c. Analysis Techniques Used in Excel

- Pivot Tables for aggregation and ranking (count, average).
- Charts for visualization: Bar, Column, Pie, Line, and Area charts.

POST-ANALYSIS AND INSIGHTS

Key Findings

1. Seedling Stage (Days 1–14): The Foundation Phase

Population: 980 plants (31% of total)

Planting Distribution: Consistent 140 plants per day across all weekdays

The seedling stage showed the most stable environmental conditions:

- **Peak pH**: 6.42 (Friday)
- **Peak Humidity**: 66.21% (Friday)
- **Peak Temperature**: 29.55°C (Friday)
- **Peak TDS**: 612.56 ppm (Thursday)
- **Nutrient Adequacy**: 61% optimal

Insight: Fridays emerged as the optimal day for seedling growth across multiple parameters, suggesting weekend preparation protocols may be influencing early-stage success.

2. Vegetative Stage (Days 15–35): The Growth Acceleration Phase

Population: 1,470 plants (46% of total)

Planting Distribution: Uniform 210 plants daily

This critical growth phase revealed interesting patterns:

- **Peak pH**: 6.43 (Friday)
- **Peak Humidity**: 65.71% (Wednesday)
- **Peak Temperature**: 28.77°C (Friday)
- **Peak TDS**: 613.97 ppm (Tuesday)
- **Nutrient Adequacy**: 58% optimal

Concern: The vegetative stage recorded the lowest nutrient adequacy at 58%, despite being the phase when plants require maximum nutrition for leaf development. Additionally, 614 plants (42%) experienced low nutrient adequacy, primarily on Saturday (198 plants) and Thursday (193 plants).

3. Maturity Stage (Days 36–50): The Harvest-Ready Phase

Population: 719 plants (23% of total)

Planting Distribution: Highly variable (70–140 plants per day)

The final stage showed more balanced conditions but uneven production:

- **Peak pH**: 6.42 (Tuesday)
- **Peak Humidity**: 65.91% (Saturday)
- **Peak Temperature**: 28.42°C (Friday)
- **Peak TDS**: 606.72 ppm (Tuesday)
- **Nutrient Adequacy**: 58% optimal

Distribution Pattern:

- High volume: Thursday–Saturday (140 plants each)
- Low volume: Wednesday (70), Tuesday (72), Monday (76), Sunday (81)

Insight: The uneven distribution suggests misalignment between planting schedules and harvest demand, potentially causing market supply gaps or storage issues.

4. Nutrient Adequacy Deep Dive

The nutrient adequacy analysis revealed critical optimization opportunities:

Low Adequacy Scenario (614 plants):

- Saturday showed the highest low-nutrient count (198 plants)
- Thursday followed with 193 plants
- Peak environmental stressors varied by day (pH on Friday, humidity on Sunday, TDS on Monday, temperature on Sunday)

Optimal Adequacy Scenario (856 plants):

- Thursday (297 plants) and Friday (302 plants) dominated optimal nutrition
- Peak conditions: pH (Thursday), humidity (Saturday), TDS (Tuesday), temperature (Friday)

Insight: The midweek dip in nutrient adequacy (Tuesday–Wednesday) suggests that weekend maintenance protocols may be creating inconsistencies that propagate through the early week.

Comparison with Initial Findings

- Initial assumptions about weekday environmental patterns were confirmed.
- The expected midweek nutrient dip was validated by quantitative evidence.

- Environmental stability (pH, temperature) emerged as a stronger yield predictor than previously assumed.

DATA VISUALIZATIONS & CHARTS

Visual summaries included:

- **Bar Chart:** Lettuce count by growth stage.
- **Column Chart:** Total lettuce plants per weekday.
- **Pie Chart:** Nutrient adequacy distribution.
- **Line/Area Charts:** Trends in temperature, pH, humidity, and TDS across weekdays.

Visual Highlights:

- Vegetative Stage dominated (1,470 plants).
- Friday = highest pH (6.41).
- Tuesday = best TDS (605.20 ppm).
- Saturday = highest humidity (65.42%).

RECOMMENDATIONS AND OBSERVATIONS

Based on these findings, here are ten data-driven strategies to enhance lettuce production efficiency:

a. Actionable Insights

1. Maintain pH stability between 6.2–6.5 for optimal nutrient uptake.
2. Improve nutrient adequacy from 58–61% to at least 65–70%.
3. Adjust TDS levels by growth stage (Seedling: 560–620 ppm; Vegetative: 650–750 ppm; Maturity: 600–700 ppm).
4. Reduce temperature from 28–29°C to the ideal 24–26°C range.
5. Maintain humidity between 60–70% using automated controls.

b. Optimizations or Business Decisions

1. Balance planting schedules evenly throughout the week to avoid overproduction on weekends.
2. Focus on vegetative-stage nutrition and improve nutrient replenishment frequency.
3. Strengthen midweek monitoring (Tuesday–Thursday) to reduce variability in pH and TDS.

4. Leverage high-performance days (Friday–Saturday) for critical operations.
5. Implement IoT sensor-based automation for continuous environmental tracking.

c. Unexpected Outcomes

- Midweek inconsistencies, likely from manual operations, affected nutrient stability.
- Nutrient adequacy dropped sharply despite favorable humidity — indicating possible fertilizer mixing or timing issues.
- Temperature peaks at week's end aligned with improved growth, contrary to expectations of heat stress.

CONCLUSION

This analysis of 3,169 lettuce plants demonstrates that successful farming isn't about perfecting one factor — it's about harmonizing many. The data revealed how subtle environmental shifts can have outsized impacts on yield and quality.

By optimizing nutrient adequacy, stabilizing temperatures, and maintaining consistent midweek operations, farms can potentially boost yields by 15–20% while enhancing flavor and freshness.

In both farming and data science, small, consistent adjustments compound into remarkable results. A 0.1 pH change or a 2°C reduction may seem small — but multiplied across thousands of plants, it's the difference between *good* and *great.*

REFERENCES & APPRENDENCIES

1. Kaggle Dataset: Lettuce Growth Days Analysis
   Available at: https://www.kaggle.com/datasets/jurijsruko/lettuce (Accessed: 7th October 2025).
2. Lucy Stone, (2025). Rural Sprout: How Many Days Lettuce Grow: Tips for Accelerating Growth (Accessed: 7th October 2025).
3. Matt Gallagher, (2025). *What Is The PH And TDS For Lettuce?* Available at: https://greenpacks.org/what-is-the-ph-and-tds-for-lettuce/ (Accessed: 7th October 2025).
4. University of Arizona: Controlled Environment Agriculture Center (Accessed: 7th October 2025).

5. Grand View Research, (2025). *Indoor Farming Market (2025 - 2030).* Available at: https://www.grandviewresearch.com/industry-analysis/indoor-farming-market (Accessed: 7th October 2025).

6. Image !: Hydroponic Lettuce photos & images.
Available at: https://www.vecteezy.com/photo/17725924-hydroponic-vegetables-from-hydroponic-farms-fresh-green-cos-lettuce-growing-in-the-garden-hydroponic-plants-on-water-without-soil-agriculture-organic-health-food-nature-chlorophyll-leaf-crop-bio (Accessed: 7th October 2025).