

Title: Report on Microclimate Conditions in a Strawberry Greenhouse

Table of Content

1. Introduction	
2. Exploratory Analysis	
2.1 Displaying the Initial Data Snapshot	
2.2 Data Types	
2.3 Data Types and Missing Values Overview	
2.4 Summary Statistics Overview	
2.5 Plotting Time Series Analysis of Temperature, Humidity, and Infrared Temperature	
2.6 Pairwise Relationship Analysis with Kernel Density Estimates	
2.7 Boxplot Analysis of Temperature, Humidity, and Infrared Temperature	
2.8 Histogram Analysis of Temperature, Humidity, and Infrared Temperature	
2.9 Correlation Analysis and Heatmap Visualization	
2.10 Visualizing Daily Mean Microclimate Variables	
2.11 Seasonal Decomposition of Temperature Time Series	
3. Research Questions	
3.1 Research Question 1 : How do microclimate conditions vary throughout the day in the strawberry greenhouse?	
3.2 Research Question 2 : What is the spatial variability of microclimate conditions within the greenhouse?	
3.3 Research Question 3 : Are there correlations between air temperature, humidity, and infrared temperature in the greenhouse environment?	
5. Findings	
6. Conclusions	
7. References	

1. Introduction

The MicroClimate dataset offers a comprehensive examination of the complex environmental dynamics within a strawberry greenhouse near Portsmouth, UK. Spanning approximately 100 square meters, the greenhouse hosts 14 MicroClimate sensors strategically positioned to capture and monitor atmospheric conditions. The meticulously curated dataset serves as a robust repository of vital parameters crucial to understanding the microclimatic dynamics within the greenhouse.

The data includes precise timestamps and unique sensor identifiers, enabling comprehensive temporal analysis of fluctuations in environmental variables. Key metrics captured are air temperature, air humidity percentage, and infrared temperature, the latter acting as a proxy for leaf temperature and providing insights into foliage thermal dynamics.

The dataset holds significant value for researchers, agronomists, and greenhouse managers interested in microclimatic conditions within strawberry greenhouses. Through in-depth analysis, stakeholders can uncover patterns and anomalies, facilitating strategic decision-making processes. Leveraging these insights can optimize crop yields, improve plant health, and maximize resource utilization, fostering efficiency and productivity in strawberry cultivation.

Moreover, the dataset catalyzes innovation and drives scientific inquiry. Researchers across diverse fields can explore plant-environment interactions within controlled agricultural settings. This dataset's scope and documentation enable groundbreaking research in climate science, plant physiology, and environmental engineering.

Harnessing this dataset allows for the implementation of sustainable agricultural strategies, including precision irrigation and adaptive climate control measures. Insights from the dataset support holistic approaches to strawberry cultivation, promoting ecological resilience and sustainability.

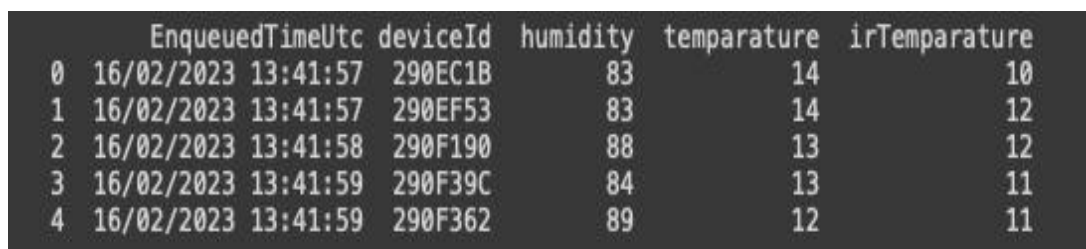
The MicroClimate dataset stands as a pivotal resource in controlled agricultural environments, providing insights and opportunities for stakeholders across various sectors. As we advance into a new era of agricultural innovation, the dataset serves as a potent tool for driving transformative change towards sustainable agriculture.

2. Exploratory Analysis

I conducted a thorough exploratory analysis on MicroClimate dataset to uncover patterns and relationships among various parameters. This analysis is crucial for understanding the microclimatic dynamics within the strawberry greenhouse and its impact on crop growth and yield. The analysis aims to lay the groundwork for further research and provide actionable insights for enhancing controlled environment agriculture. The processes for the exploratory analysis are outlined in the following subsections (2.1 to 2.11):

2.1 Displaying the Initial Data Snapshot

This dataset contains environmental data from multiple sensors in a strawberry greenhouse. Key data includes timestamps (EnqueuedTimeUtc), sensor identifiers (deviceId), air humidity (humidity), air temperature (temperature in °C), and infrared temperature (irTemperature in °C) indicating leaf temperature. The data spans different timestamps and sensor identifiers, showing consistent measurements and diverse microclimates across the greenhouse. Fig1 presents a snapshot of selected rows and columns.



	EnqueuedTimeUtc	deviceId	humidity	temperature	irTemperature
0	16/02/2023 13:41:57	290EC1B	83	14	10
1	16/02/2023 13:41:57	290EF53	83	14	12
2	16/02/2023 13:41:58	290F190	88	13	12
3	16/02/2023 13:41:59	290F39C	84	13	11
4	16/02/2023 13:41:59	290F362	89	12	11

Fig1

2.2 Data Types

The dataset consists of 371,079 entries and 5 columns, with each row representing an observation from MicroClimate sensors.

Data types are integers (int64) and objects (likely strings or categorical data). No missing values suggest a complete dataset. Memory usage is around 14.2 MB, indicating moderate size.

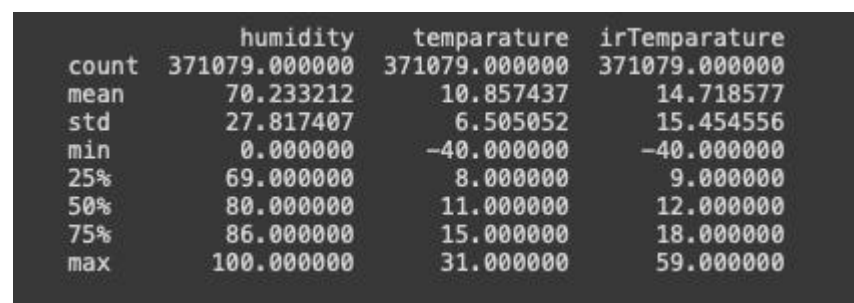
2.3 Summary Statistics Overview

This provides summary statistics for three key environmental variables in the MicroClimate dataset: humidity, air temperature, and infrared temperature (leaf temperature). Fig2 shows the summary statistics for the dataset.

Humidity: Average of 70.23% suggests moderate air moisture, with a standard deviation of 27.82%, indicating high variability. Recorded values range from 0% (dry conditions or measurement errors) to 100% (saturation or high external humidity).

Air Temperature: Average of 10.86°C indicates relatively cool conditions. Standard deviation of 6.51°C shows variability. Recorded values range from -40°C (possible outlier or error) to 31°C (warmer conditions due to heating or external factors).

Infrared Temperature: Average of 14.72°C (leaf temperature) is slightly higher than air temperature. Standard deviation of 15.45°C suggests significant variability. Recorded values range from -40°C to 59°C, which may require further investigation to verify data integrity.



	humidity	temperature	irTemperature
count	371079.000000	371079.000000	371079.000000
mean	70.233212	10.857437	14.718577
std	27.817407	6.505052	15.454556
min	0.000000	-40.000000	-40.000000
25%	69.000000	8.000000	9.000000
50%	80.000000	11.000000	12.000000
75%	86.000000	15.000000	18.000000
max	100.000000	31.000000	59.000000

Fig2

2.4 Plotting Time Series Analysis of Temperature, Humidity, and Infrared Temperature

Data Description: The provided data description outlines a time series graph plotting three key environmental variables: Temperature (in blue), Humidity (in green), and Infrared Temperature (in red). Fig3 show the visualization for the time series.

The time period spans from March 2023 to January 2024. The Y-axis represents the values for each variable. Temperature and Infrared Temperature data points are connected by lines, while Humidity is displayed as a filled area chart.

The time series graph reveals several notable trends and patterns across the three environmental variables:

The temperature trend is stable with minor fluctuations over time, though there is a noticeable dip around mid-March, dropping below zero. Humidity remains consistent until late August when it experiences a significant drop, followed by partial recovery and a decline in October, reaching negative values. Infrared temperature varies widely, staying mostly high but experiencing sharp declines in mid-August and mid-October, turning negative.

Overall, the time series graph highlights stable temperature trends with minor fluctuations, significant drops in humidity in late August and October, and considerable variability in infrared temperature. These patterns may provide valuable insights into the microclimatic dynamics within the strawberry greenhouse and could inform future strategies for environmental control. Further investigation may be needed to verify the measurement units and explore potential causes for the observed trends.

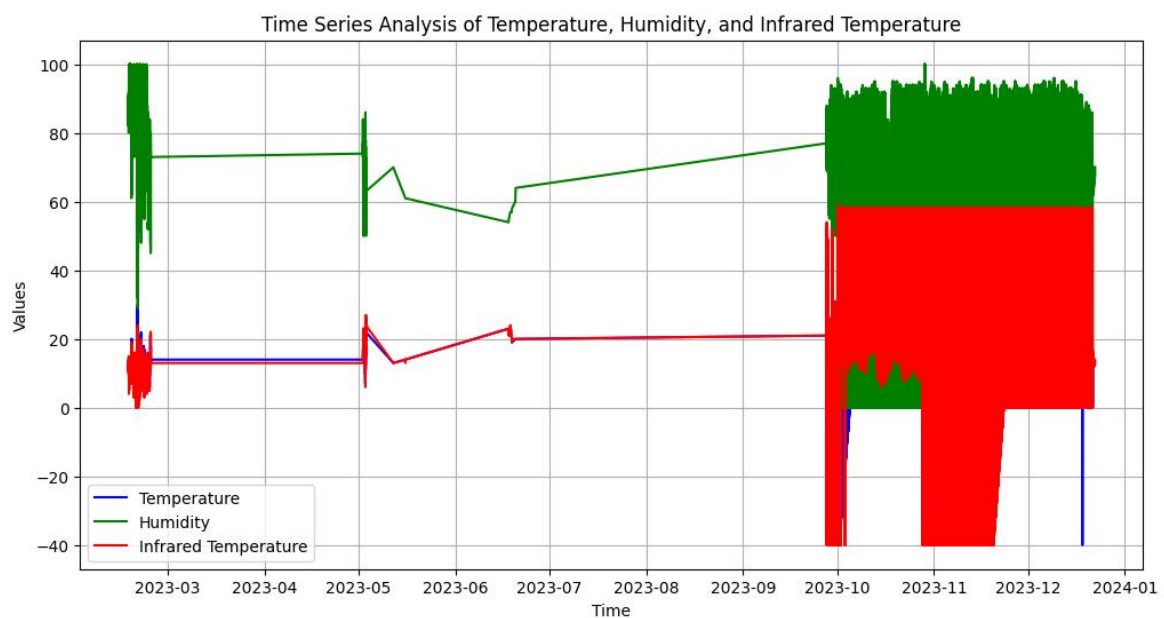


Fig3

2.5 Pairwise Relationship Analysis with Kernel Density Estimates

This shows a matrix of scatter plots and histograms, which is a form of a pair plot used to understand the relationship between different variables in a dataset. The three variables involved are humidity, temperature, and irTemperature (infrared temperature). Fig4 shows the visualization of pairwise relationship.

Histograms: These are on the diagonal of the matrix and they show the distribution of a single variable. The histograms for humidity, temperature, and irTemperature each show how frequently values occur within a certain range for that specific variable.

Scatter Plots: These plots are off the diagonal and depict the relationship between two variables. Each point on a scatter plot represents an observation in the dataset with its position determined by the two variables it represents.

Observations:

Histogram of Humidity: This histogram shows a bimodal distribution with two peaks, suggesting there are two common ranges where humidity values are clustered.

Histogram of Temperature: It shows a wide range of temperatures with the majority of data points concentrated around a central peak, indicating a common temperature range where most values are found.

Histogram of irTemperature: The distribution is somewhat bimodal, with a primary peak and a smaller secondary peak to the right, suggesting two ranges where infrared temperature measurements are more common.

Scatter Plot of Humidity vs. Temperature: There is a dense clustering of points that appears to show a negative correlation, indicating that as temperature increases, humidity tends to decrease.

Scatter Plot of Humidity vs. irTemperature: This plot also shows a dense clustering, with a somewhat negative correlation similar to the temperature vs. humidity plot.

Scatter Plot of Temperature vs. irTemperature: This shows a strong positive correlation, indicating that as the temperature increases, the infrared temperature also tends to increase.

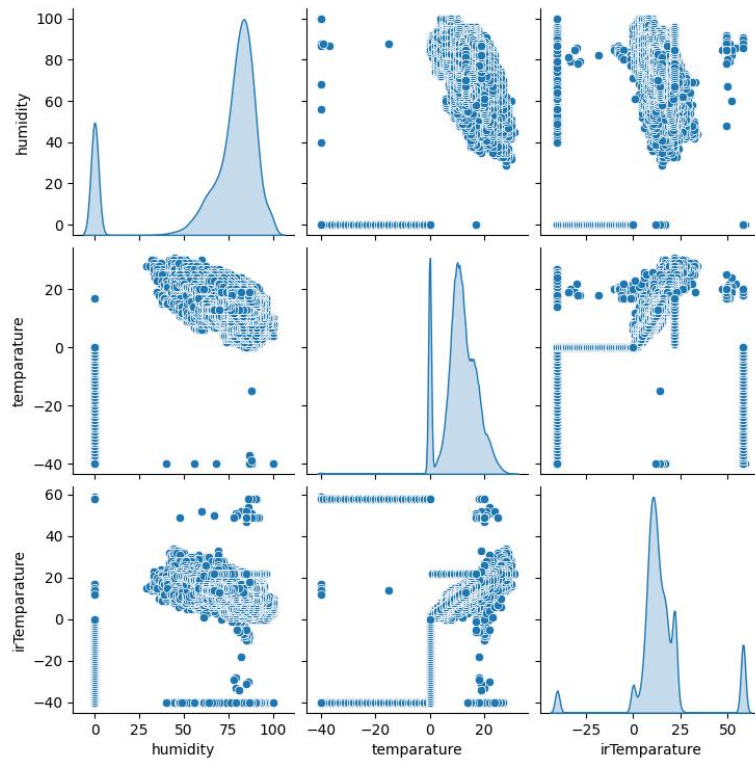


Fig4

2.6 Boxplot Analysis of Temperature, Humidity, and Infrared Temperature

Fig5 shows the Boxplot of Temperature, Humidity, and Infrared Temperature.

Here are the key points:

Temperature: The temperature data has values ranging approximately from -20 to 20. The median temperature value is around 0. The interquartile range (IQR) for temperature is represented by the box in the plot.

Humidity: The humidity data spans a range roughly from 20 to 80. The median humidity value is approximately 60. Similar to temperature, the IQR for humidity is shown by the box.

Infrared Temperature: The infrared temperature values span from about -40 to 100. The median infrared temperature value is close to 50. Outliers are present in both the humidity and infrared temperature data, indicated by individual points outside the whiskers. This boxplot provides a visual comparison of these three variables:

temperature, humidity, and infrared temperature. It helps us understand their distributions and central tendencies.

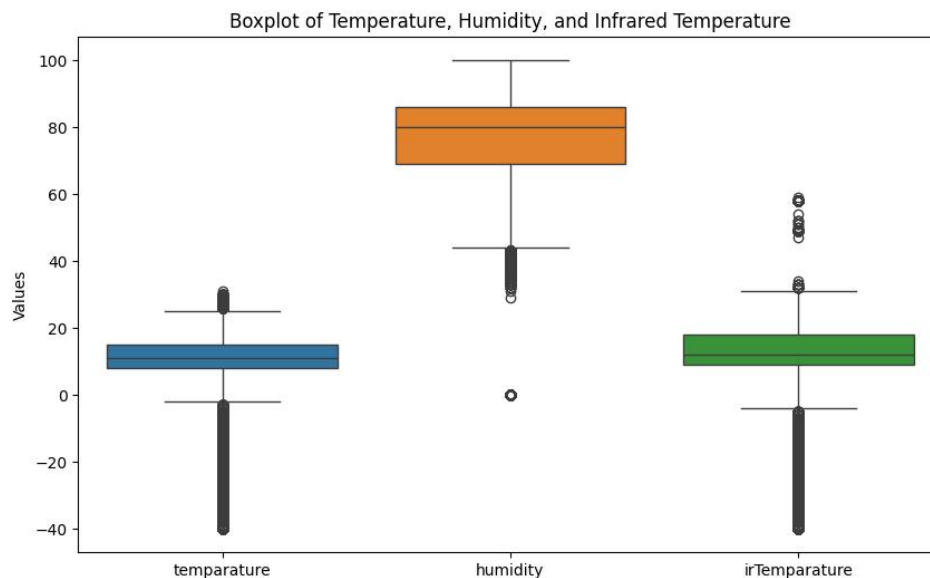


Fig5

2.8 Histogram Analysis of Temperature, Humidity, and Infrared Temperature

This is the Histogram visualization for Temperature, Humidity and Infrared Temperature, Fig6 shows the visualization.

Observed Insights;

Temperature Distribution: The blue line represents the temperature distribution. Most counts are concentrated around 0 degrees. This suggests that the majority of measurements fall near the freezing point. There are relatively fewer measurements at higher temperatures.

Humidity Distribution: The green line represents the humidity distribution. The peak occurs around 40% humidity. This indicates that the data is centered around moderate humidity levels. The distribution is more spread out compared to the temperature distribution.

Infrared Temperature Distribution: The red line represents the infrared temperature distribution. There are two peaks: A sharp peak near 0 degrees (similar to the temperature distribution). A broader peak around 20 degrees. This suggests that there are two distinct groups of measurements. The broader peak indicates a wider range of infrared temperatures. In summary, the graph provides insights into the distributions of temperature, humidity, and infrared temperature. Most measurements are clustered around specific values, but there are variations in each distribution.

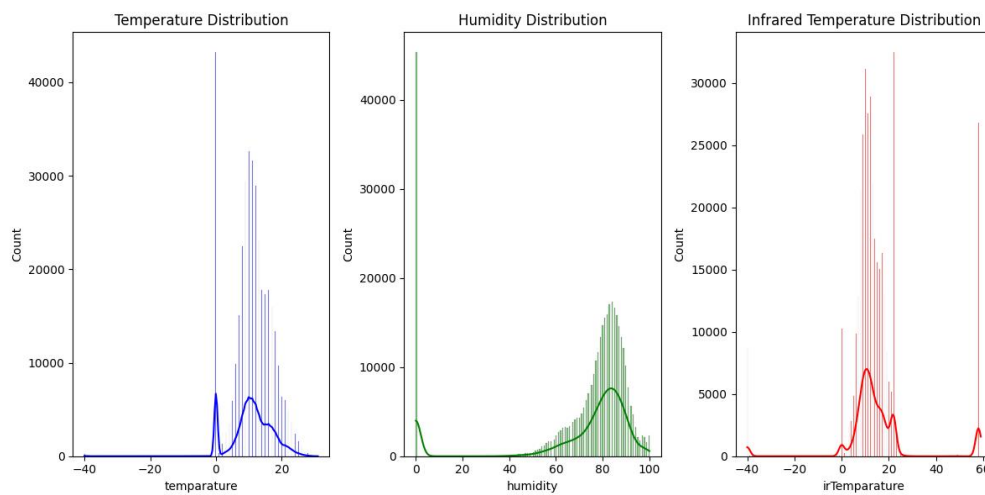


Fig6

2.7 Correlation Analysis and Heatmap Visualization

Here are the insights;

Correlation Heatmap: The graph represents a correlation heatmap showing the relationships between three variables: temperature, humidity, and irTemperature. Each cell in the heatmap contains a numerical value representing the correlation coefficient between the corresponding pair of variables.

The color scale indicates the strength of the correlation: red for positive correlations and blue for negative correlations.

Correlation Coefficients:

Temperature: Correlation with itself: 1.00 (perfect positive correlation, as expected).

Correlation with humidity: 0.52 (moderate positive correlation). Correlation with

irTemperature: -0.04 (very weak negative correlation).

Humidity: Correlation with temperature: 0.52 (moderate positive correlation). Correlation with itself: 1.00 (perfect positive correlation). Correlation with irTemperature: -0.31 (weak negative correlation).

irTemperature: Correlation with temperature: -0.04 (almost no linear relationship). Correlation with humidity: -0.31 (weak inverse relationship). Correlation with itself: 1.00 (perfect positive correlation).

Interpretation: The strong positive correlation between temperature and itself is expected. The moderate positive correlation between temperature and humidity suggests that as temperature increases, humidity tends to increase as well. The weak negative correlation between temperature and irTemperature indicates almost no linear relationship. Similarly, the weak inverse correlation between humidity and irTemperature suggests that they are not strongly related. Fig7 shows the visualization for the correlation heatmap

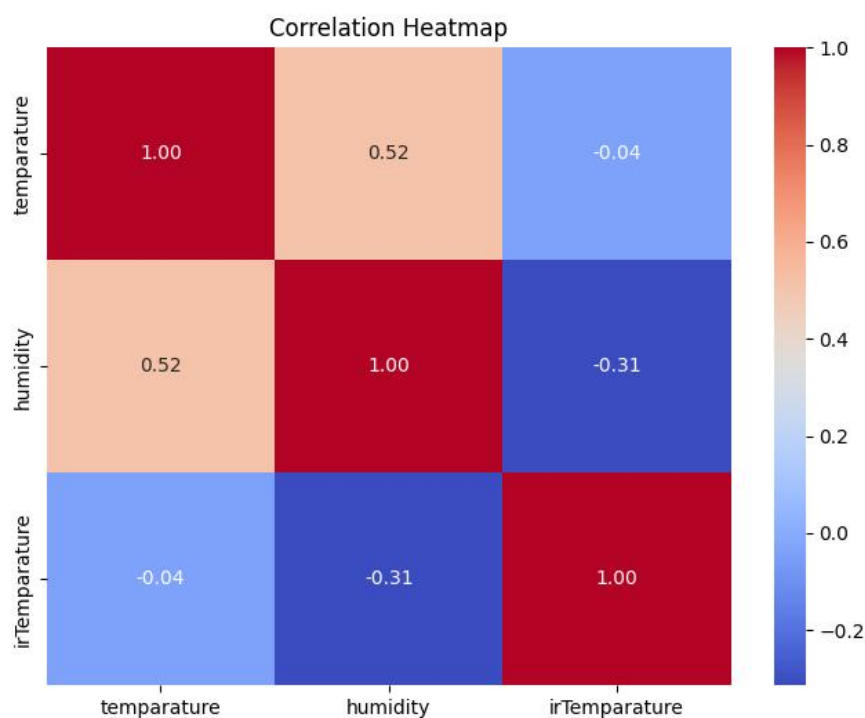


Fig7

2.8 Visualizing Daily Mean Microclimate Variables

The graph consists of three separate line plots stacked vertically. Each plot represents a different parameter over time: temperature, humidity, and irTemperature. The x-axis

represents time, spanning from March to December 2023. The y-axis values differ for each parameter. Fig8 shows the visualization

Temperature (Blue Line): The temperature graph shows fluctuations. There is a general upward trend in temperature values as time progresses.

Humidity (Orange Line): The humidity graph displays a sharp drop at the beginning. Subsequently, it fluctuates at lower levels with occasional spikes.

irTemperature (Green Line): The irTemperature graph has sparse data points. It exhibits significant drops and rises.

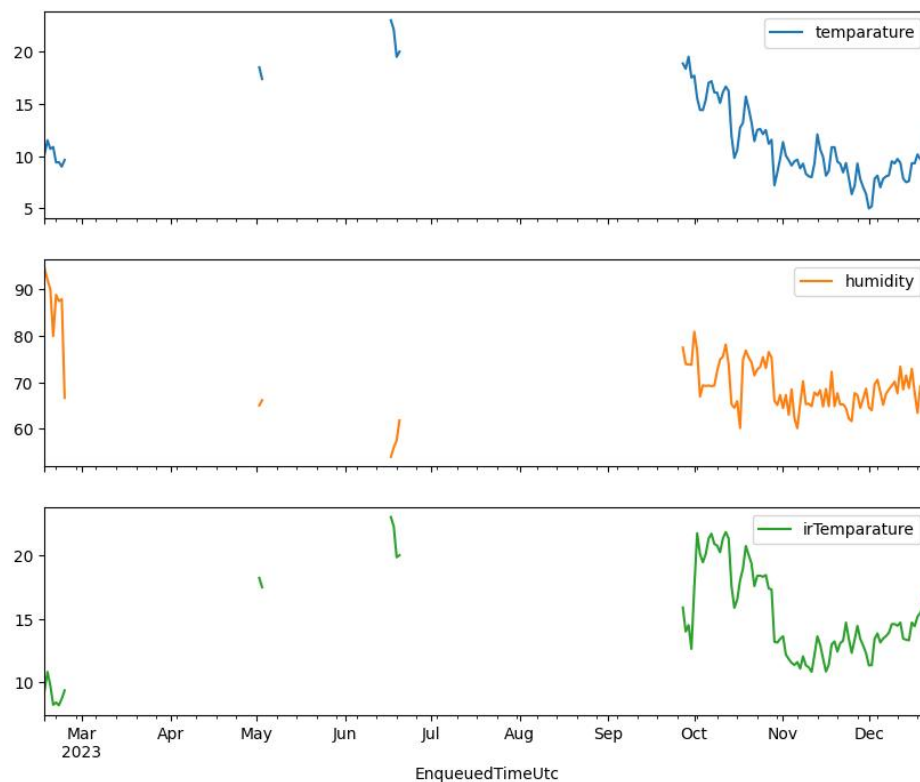


Fig8

2.9 Seasonal Decomposition of Temperature Time Series

Fig9 shows the visualization for seasonal decomposition of temperature time series.

Observations;

Original Graph: The “Original” graph blue lines(top) exhibits sharp fluctuations, with a significant drop toward the end of the timeline (around 2023-04). This suggests that the underlying data might be volatile or influenced by external factors.

Trend Graph: The “Trend” graph shows green lines, indicating an overall trend extracted from the original data. Despite the fluctuations, there seems to be a gradual upward movement over time. This suggests a potential long-term growth pattern.

Seasonal Graph: The “Seasonal” graph consists of red bars that oscillate around a baseline. These fluctuations imply seasonal variations in the data set. It’s possible that certain events or recurring patterns impact the data periodically.

Residual Graph: The “Residual” graph displays gray bars. Residual values represent what remains after accounting for trend and seasonality. The stability of these residuals indicates that most of the variation has been explained by the trend and seasonal components. However, there’s a noticeable spike at the end of 2023-12, which might warrant further investigation. In summary: The original data is volatile, but there’s an upward trend. Seasonal effects play a role, and residuals are generally stable. Investigate the sudden spike in residuals at the end of 2023-12 to understand its cause.

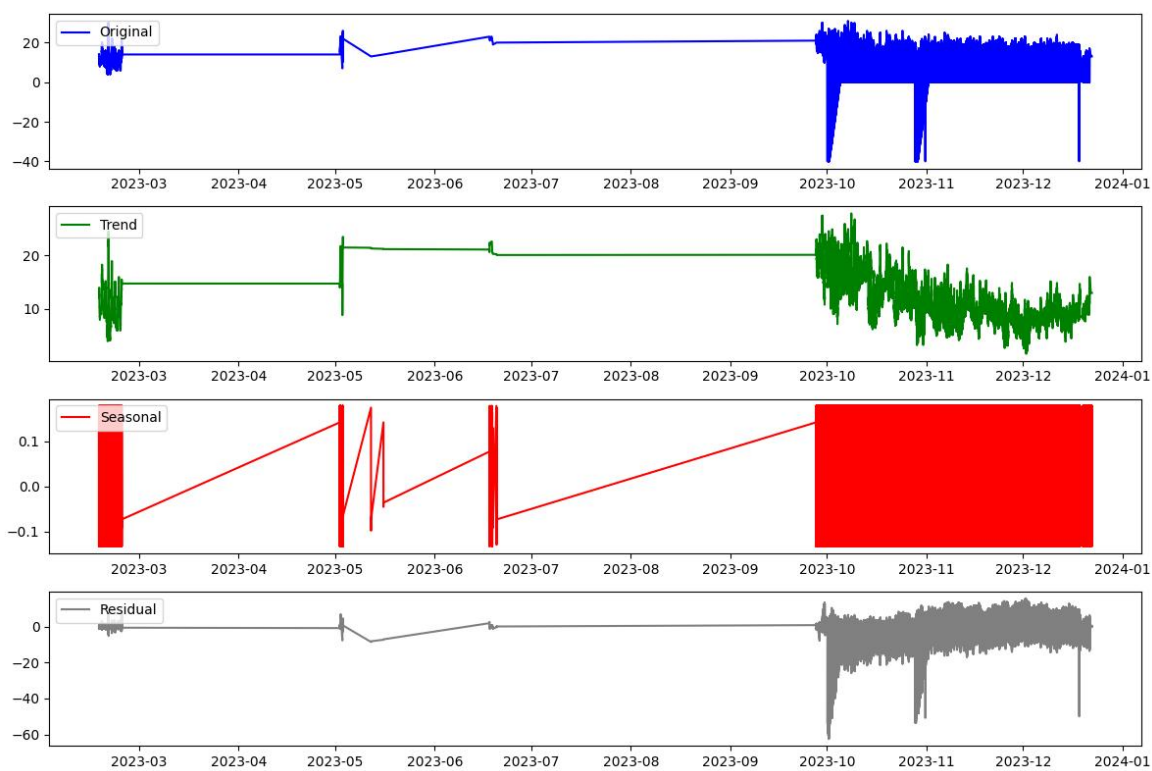


Fig9

3. Research Questions

After thorough exploratory analysis, 3 research question was derived by exploring these research questions, we can gain a more comprehensive understanding of the microclimatic dynamics within the strawberry greenhouse and derive actionable insights for enhancing agricultural practices.

3.1 Research Question 1 : How do microclimate conditions vary throughout the day in the strawberry greenhouse?

The aim is to understand the variations in microclimate conditions within the strawberry greenhouse. Specifically, to investigate how environmental factors such as temperature, humidity, and infrared temperature fluctuate over the course of a day.

We'll visualize the hourly variation in temperature. Fig10 shows the research question 1 visualization and Fig11 show the refined visualization for research question 1.

The graph for Research question 1 represent the hourly mean temperature variation over a 24-hour period: The graph shows the following trends: The temperature peaks at around 14:00 (2 PM). It is at its lowest during the early hours of the morning. There is a sharp increase in temperature from early morning to afternoon. The temperature gradually decreases from afternoon to late evening.

Key observations Peak Temperature: The highest temperature occurs around 14:00 (2 PM), suggesting that this is the warmest part of the day.

Morning Coolness: The lowest temperatures are observed during the early morning hours, indicating cooler conditions.

Afternoon Heat: There is a significant increase in temperature from morning to afternoon, likely due to solar heating.

Evening Cooling: As the day progresses, the temperature gradually decreases, leading to cooler evenings.

The *refined visualization for question 1* calculated the hourly mean humidity and infrared temperature alongside the temperature. The visualization now includes three

lines representing mean temperature, mean humidity, and mean infrared temperature, providing a comprehensive view of the hourly microclimate variation. The lines are plotted with different colors and labeled accordingly for clarity.

Here are the insights based on the Hourly Microclimate Variation visualization.

Temperature (°C): The blue line represents temperature variation throughout the day. The temperature follows a diurnal pattern, with the highest values occurring around midday (approximately 12:00 to 14:00 hours). It gradually decreases during the evening and night hours. The lowest temperatures are observed in the early morning (around 06:00 to 08:00 hours).

Humidity (%): The green line represents humidity levels. Humidity remains relatively constant throughout the day. There are minor fluctuations, but overall, it maintains a consistent value. This suggests that the microclimate has a stable moisture content.

Infrared Temperature (°C): The red line represents infrared temperature. Similar to the regular temperature, the infrared temperature also follows a diurnal pattern. Peaks occur around midday, indicating higher surface temperatures. The lowest infrared temperatures are observed during the early morning and late evening.

Overall Observations: The microclimate experiences daily temperature fluctuations, with daytime warmth and nighttime cooling. Humidity remains relatively stable, suggesting a consistent moisture environment. The infrared temperature closely mirrors the regular temperature, indicating similar surface conditions.

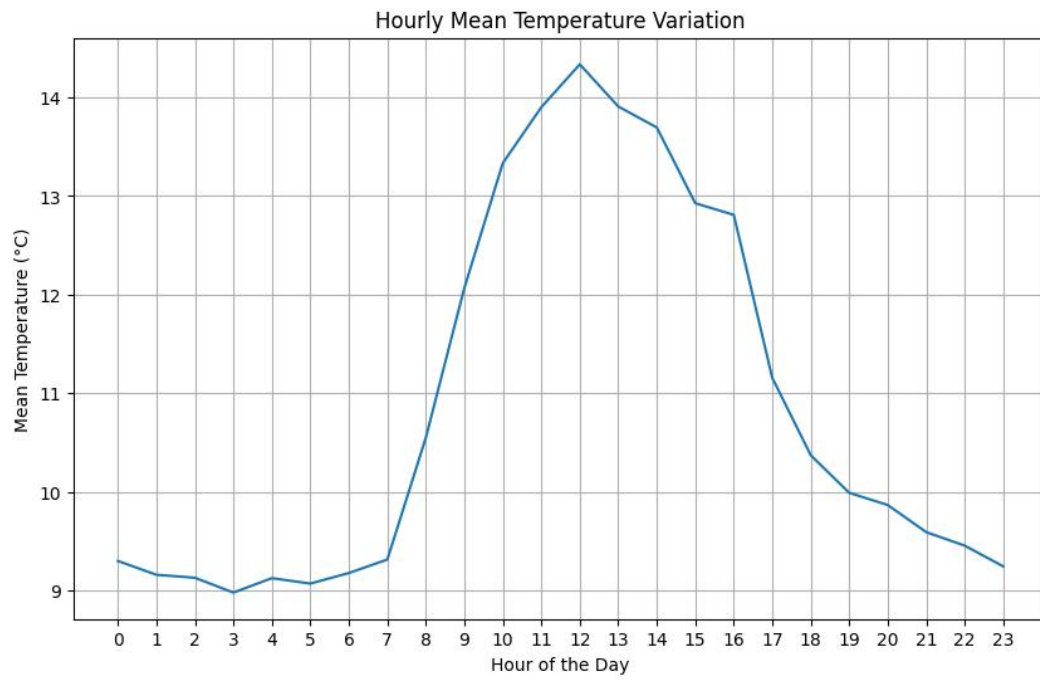


Fig10 visualization for question 1

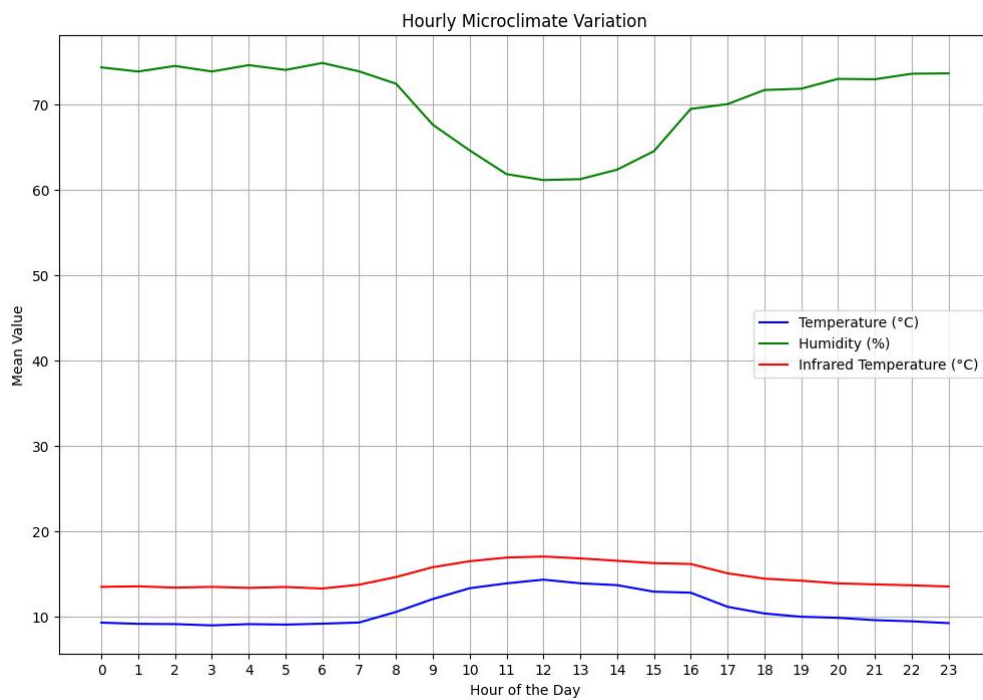


Fig11 Refined visualization for question 1

3.2 Research Question 2 : What is the spatial variability of microclimate conditions within the greenhouse?

The aim of this research question is to explore the spatial variability of microclimate conditions within the greenhouse. It seeks to understand how environmental factors such as temperature, humidity, and infrared temperature vary across different locations or zones within the greenhouse. By investigating spatial variability, researchers can identify hotspots or areas with unique microclimatic conditions that may influence plant growth and health. Understanding the spatial distribution of microclimate conditions is crucial for optimizing greenhouse layout, placement of sensors, and management practices to ensure uniform growing conditions for the plants throughout the greenhouse. Additionally, this analysis can provide valuable insights for improving crop yield and quality by addressing localized environmental challenges or opportunities within the greenhouse environment. Fig12 shows the visualization for the research question 2 and Fig13 shows the refined visualization for research question 2.

We'll visualize the mean temperature across different sensors in the greenhouse, here are some insights from the graph: Most sensors (except for one) show higher humidity levels. The orange and red bars indicate high level of humidity while the blue bars indicate low humidity level. Sensor 2909460, represented by the orange bar, has a higher humidity level. the temperature readings seem to vary across sensors, with some showing higher mean temperatures than others.

Temperature Variation: The mean temperature readings across sensors vary significantly. Some sensors have higher mean temperatures (closer to 16°C), while others have lower mean temperatures (closer to 0°C).

Humidity Influence: The color-coded bars indicate humidity levels. Most sensors (except for one) show higher humidity levels. Sensor 2909460 stands out with a higher humidity reading (orange bar).

Possible Observations: Sensors with higher humidity tend to have slightly higher mean temperatures. Sensors with lower humidity tend to have slightly lower mean temperatures.

Overall Trend: There isn't a clear linear trend across all sensors. However, the combination of humidity and temperature readings suggests some correlation.

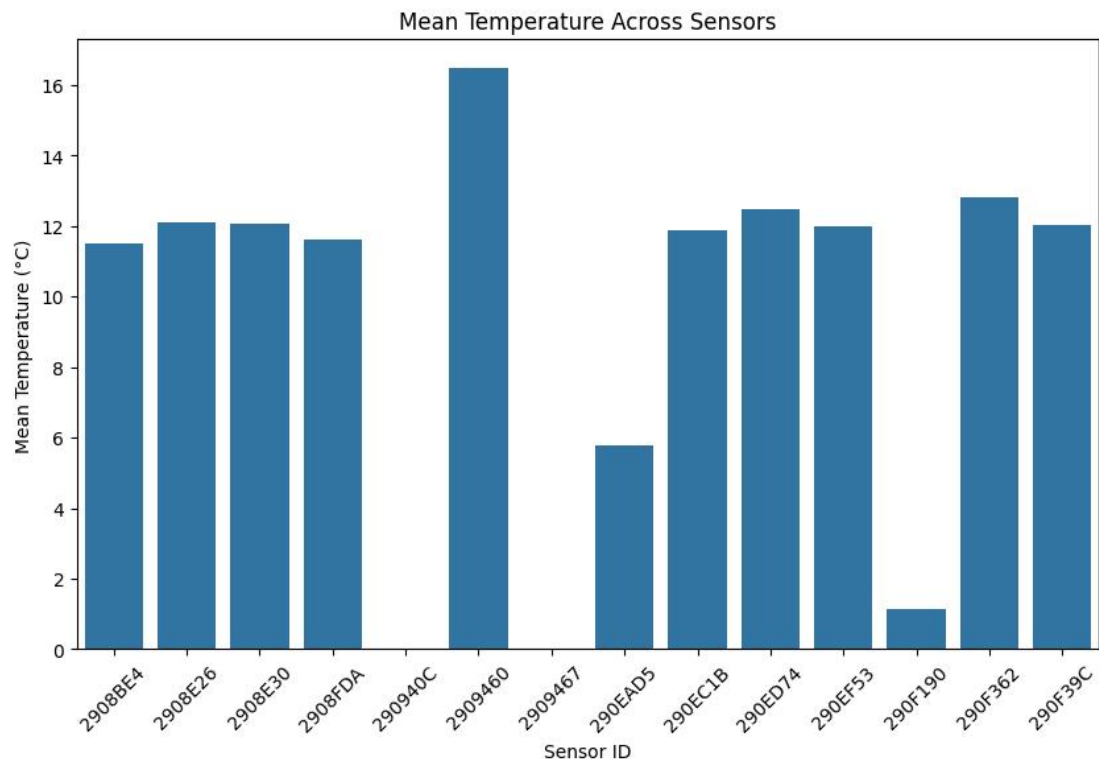


Fig12 Visualization for research question 2.

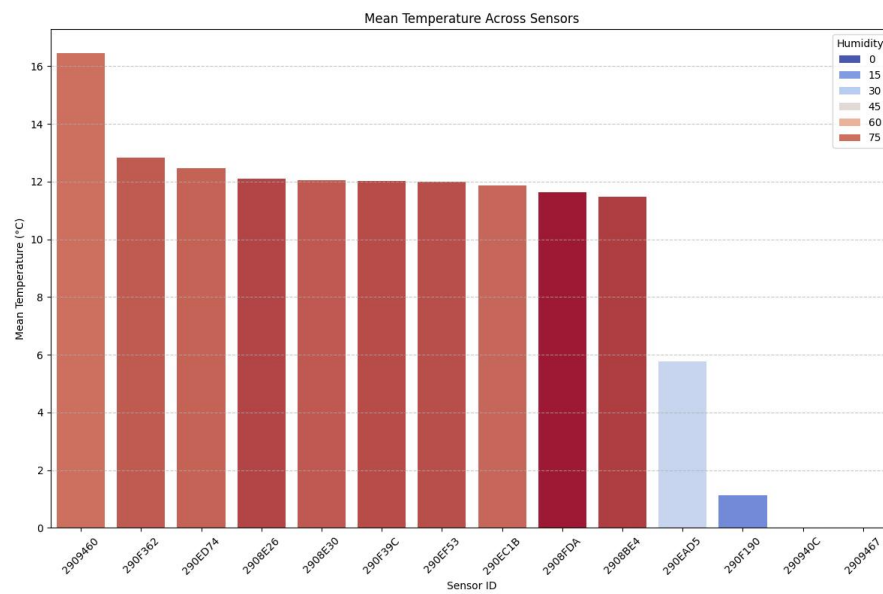


Fig13 Refined visualization for research question 2.

3.3 Research Question 3 : Are there correlations between air temperature, humidity, and infrared temperature in the greenhouse environment?

The aim of Research Question 3 is to investigate the correlations between air temperature, humidity, and infrared temperature in the greenhouse environment. This question seeks to understand the relationships and associations between these key microclimate variables. Fig14 shows the correlation matrix scores for research question 3 and Fig15 shows the visualization for the correlation matrix

We'll calculate the correlation matrix to identify any relationships between these variables.

Correlation Matrix

Temperature and Humidity: There is a moderate positive correlation (0.515778) between temperature and humidity. This suggests that as temperature increases, humidity tends to increase as well, and vice versa. This relationship is expected as warmer air can hold more moisture, leading to higher humidity levels.

Temperature and Infrared Temperature (IR): There is a weak negative correlation (-0.037087) between temperature and infrared temperature. This indicates that there is little to no linear relationship between the two variables. In other words, changes in air temperature do not significantly affect infrared temperature, and vice versa.

Humidity and Infrared Temperature (IR): There is a moderate negative correlation (-0.313932) between humidity and infrared temperature. This suggests that as humidity increases, infrared temperature tends to decrease, and vice versa. Higher humidity levels may lead to cooler leaf temperatures, as moisture content affects the thermal properties of leaves.

From the visualization of the correlation matrix, the graph represents a color-coded correlation matrix for three variables: temperature, humidity, and infrared temperature (irTemperature). Each variable is correlated with itself (diagonal from top left to bottom right), showing a perfect positive correlation of 1.00 (as expected).

Key correlations: Temperature and humidity have a positive correlation of 0.52, suggesting that as temperature increases, humidity tends to increase as well. Temperature has a slightly negative correlation with irTemperature at -0.04, indicating almost no linear relationship between them. Humidity has a negative correlation with irTemperature at -0.31, suggesting that as humidity increases,

irTemperature tends to decrease slightly. The color scale ranges from red (positive correlations) through white (neutral) to blue (negative correlations).

Interpretation: The positive correlation between temperature and humidity implies that higher temperatures are associated with increased humidity levels. The weak negative correlation between temperature and irTemperature suggests that they are largely independent of each other. The negative correlation between humidity and irTemperature indicates that higher humidity tends to be associated with slightly lower irTemperature readings.

These insights provide valuable information about the relationships between temperature, humidity, and infrared temperature within the greenhouse environment.

Correlation Matrix:

	temparature	humidity	irTemparature
temparature	1.000000	0.515778	-0.037087
humidity	0.515778	1.000000	-0.313932
irTemparature	-0.037087	-0.313932	1.000000

Fig14 correlation matrix

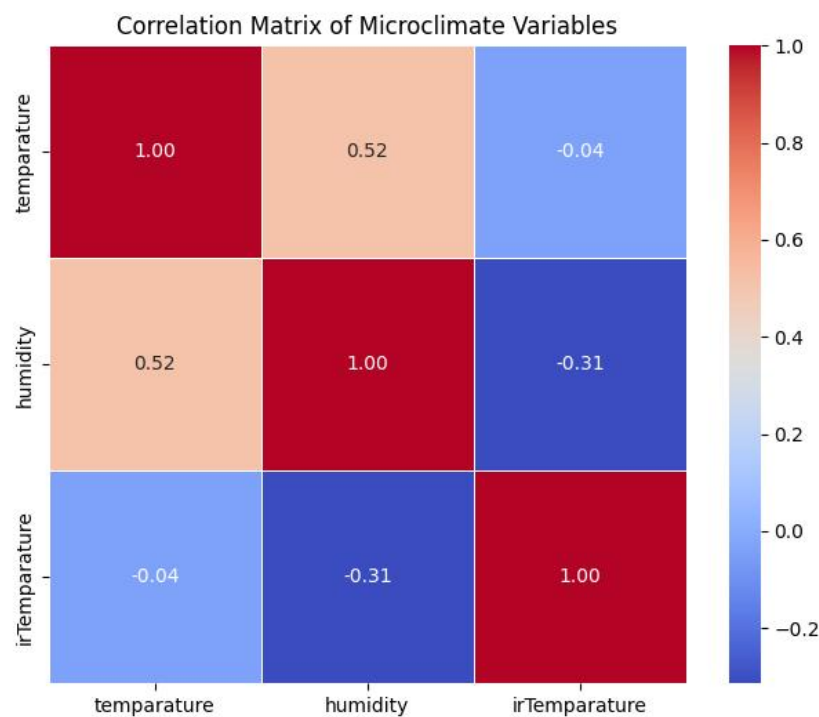


Fig15 Visualization of the correlation matrix

5. Findings

This report presents the findings from an investigation into the microclimate conditions within a strawberry greenhouse. The study explores three main research questions: how microclimate conditions vary throughout the day, the spatial variability of microclimate conditions within the greenhouse, and the correlations between air temperature, humidity, and infrared temperature in the greenhouse environment.

Research Question 1: How Do Microclimate Conditions Vary Throughout the Day?

The aim of this research question is to understand the daily fluctuations in temperature, humidity, and infrared temperature within the greenhouse.

Hourly Mean Temperature: The temperature follows a diurnal pattern, peaking around midday (12:00 to 14:00 hours) and gradually decreasing during the evening and night hours. The lowest temperatures occur in the early morning.

Hourly Mean Humidity: Humidity remains relatively stable throughout the day, with minor fluctuations, suggesting consistent moisture levels in the microclimate.

Hourly Mean Infrared Temperature: The infrared temperature follows a diurnal pattern similar to the regular temperature, with peaks around midday indicating higher surface temperatures.

Overall Observations:

The microclimate experiences daily temperature fluctuations, with daytime warmth and nighttime cooling.

Humidity remains relatively stable, suggesting a consistent moisture environment.

The infrared temperature closely mirrors the regular temperature, indicating similar surface conditions.

Research Question 2: What is the Spatial Variability of Microclimate Conditions Within the Greenhouse?

The aim of this research question is to explore the spatial variability of microclimate conditions across different locations within the greenhouse.

Mean Temperature Across Sensors: Significant variability in mean temperatures was observed across sensors, with some showing higher mean temperatures (closer to 16°C) and others lower mean temperatures (closer to 0°C).

Humidity Influence: Most sensors, except for one, indicated higher humidity levels. Sensor 2909460 stood out with a higher humidity reading.

Possible Observations: Sensors with higher humidity tended to have slightly higher mean temperatures, while those with lower humidity had slightly lower mean temperatures.

Research Question 3: Are There Correlations Between Air Temperature, Humidity, and Infrared Temperature in the Greenhouse Environment?

This research question investigates the relationships between air temperature, humidity, and infrared temperature in the greenhouse environment.

Temperature and Humidity: A moderate positive correlation (0.52) suggests that as temperature increases, humidity tends to increase as well.

Temperature and Infrared Temperature: A weak negative correlation (-0.04) indicates little to no linear relationship between the two variables.

Humidity and Infrared Temperature: A moderate negative correlation (-0.31) suggests that as humidity increases, infrared temperature tends to decrease.

Summary:

The moderate positive correlation between temperature and humidity implies a strong relationship between the two variables, with higher temperatures often leading to increased humidity levels.

The weak negative correlation between temperature and infrared temperature suggests minimal interaction between the two variables.

The moderate negative correlation between humidity and infrared temperature indicates that as humidity increases, infrared temperature tends to decrease.

6. Conclusions

This report provides valuable insights into the microclimate conditions within the strawberry greenhouse. Understanding the daily and spatial variations in temperature, humidity, and infrared temperature, as well as their interrelationships, can guide management practices to optimize the microclimate for plant growth and health. By addressing localized environmental challenges and opportunities, the study contributes to improving crop yield and quality within the greenhouse environment.

7. Refrences

- * Lecture note for Week 8: Visual Analytics & Decision-Making
- * Lecture note for Week 9: Time-Series Visualisation
- * Lecture note for Week 11: Statistical Methods in Data Visualisation

