Lettuce Dataset Analysis

Prateek Singh/ 10st June 2025

Agenda

Introduction
Data Source
Data Cleaning
Summary of the Analysis
Detailed Summary

Introduction I

- The Lettuce dataset offers a unique perspective into the life cycle of a Lettuce plant.
- The project aims to investigate the relationship between the temperature, humidity, TDS value, pH level as the plant progress throughout its life and how they effect the plant growth.
- The project also aims to calculate the statistical summary such as mean, median, quartiles, min/max for each variables to gain insights into the distribution of the data.
- The dataset consists of the following variables:
  - Plant_Id: A distinctive identifier assigned to each individual plant.
  - Date: The timestamp of the observation, marking key milestones in the growth process.
  - Temperature(°C): The recorded temperature in degree Celsius, a crucial environmental variable.
  - Humidity(%): The percentage representing the humidity level, influencing the plant's water uptake.
  - Total Dissolved Solvents(TDS)(ppm): A measurement of dissolved solids in parts per million, reflecting nutrient availability.
  - pH level: The environmental pH level, another crucial factor impacting nutrient absorption.
  - Growth Days: The duration, in days, from the initial growth stage to the plant's full maturity.
- The dataset was made available under the Apache 2.0 licence.

Introduction II

Plant_ID	Date	Temperature (°C)	Humidity (%)	TDS Value (ppm)	pH Level	Growth Days
1	8/3/2023	33.4	53	582	6.4	1
1	8/4/2023	33.5	53	451	6.1	2
1	8/5/2023	33.4	59	678	6.4	3
1	8/6/2023	33.4	68	420	6.4	4
1	8/7/2023	33.4	74	637	6.5	5
1	8/8/2023	32.3	77	478	6.8	6
1	8/9/2023	32.3	75	682	6	7
1	8/10/2023	22.7	63	576	6.3	8
1	8/11/2023	31.9	69	662	6.1	9
1	8/12/2023	30.2	59	607	6.2	10
1	8/13/2023	30.1	77	670	6.5	11
1	8/14/2023	30.1	54	535	6.4	12
1	8/15/2023	30.1	78	480	6	13
1	8/16/2023	29.8	56	688	6.4	14
1	8/17/2023	29.6	62	653	6.6	15

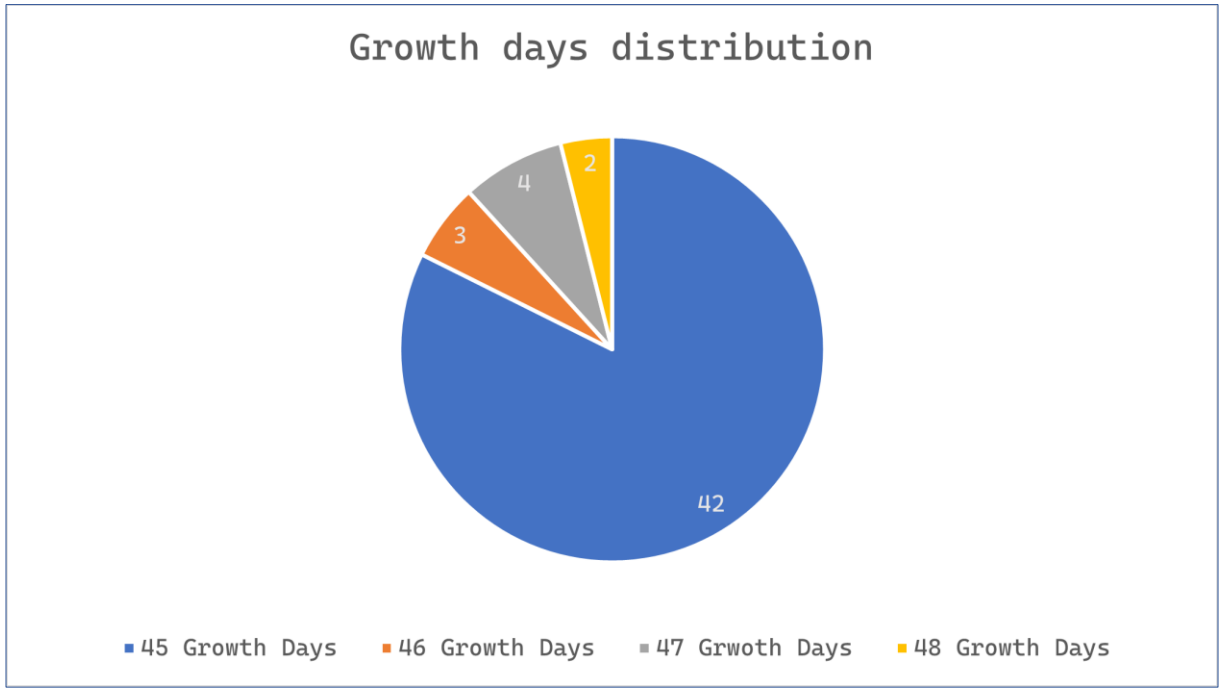
Data Cleaning and Manipulation

- The date column had some dates as text format and some in the date format. So, there was no direct method to make them consistent.
- First replaced all the delimiters present in the date column to '-' from '/'. So that it can be divided into multiple columns, using the text to column option.
- Then used the $DATE()$ function to concate the years, month, and date to make the dates in a consistent format.
- Converted the column types of $Plant_Id$, TDS_Value, $Humidity$, pH_Level and $Growth\ Days$ from General to Number, also changed the column type of $Temperature$ from General to Temperature.
- Under extensive inspection of the dataset we observe some incoherent values for the $Plant_Id$ grater than 50; hence we decided to exclude those values for accurate analysis.
- Using the Z-Score Test, I was able to find out the outliers in the dataset and removed them, hence increasing the accuracy of the analysis.

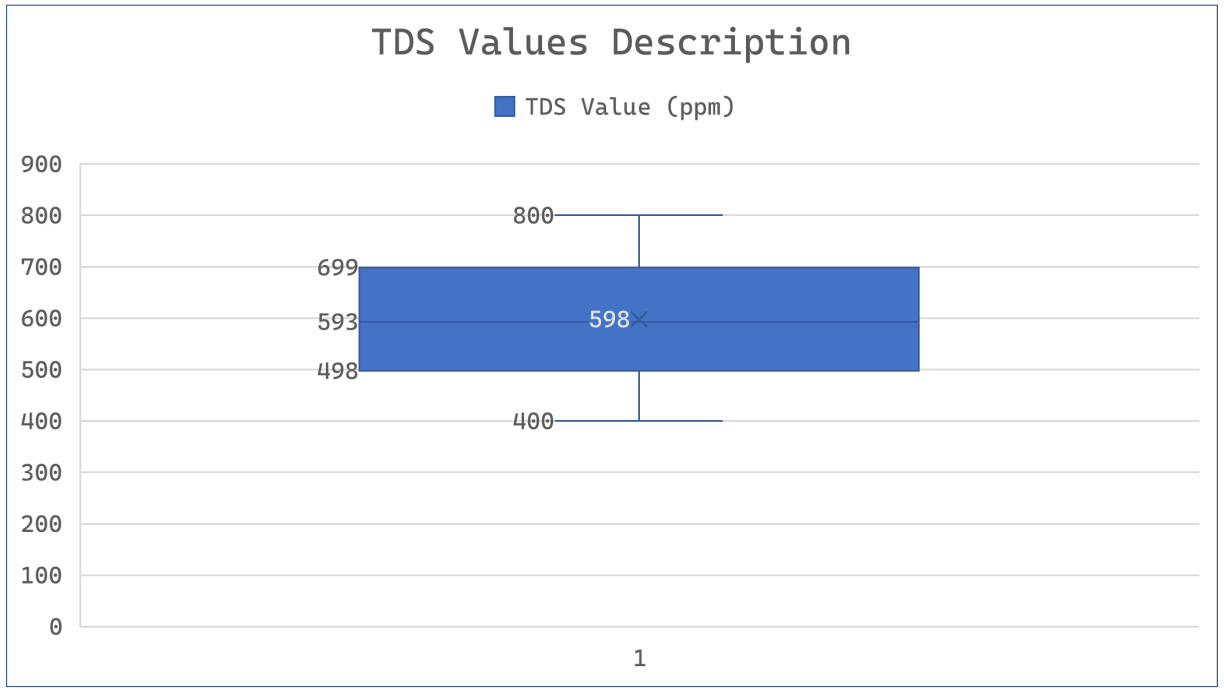
Summary of the Analysis

- For the project, 70 randomly selected Lettuce plants were taken into consideration; which had an average growth day of 45.25 days.
- The maximum growth days was 48 days, while the minimum growth days for plant was 45 days, for the plants present in the dataset.
- The pH of the soil remains between 6.3 and 6.4 throughout the Lettuce plant’s life cycle. Though we see a sudden drop in the pH after day 46, where it reaches the value 6.2.
- We see no general trend in the humidity of the environment with respect to the growth days, but it generally stays stable in the range of 60s%.
- The temperature during the Growth Days, in the life cycle of the plant, usually stays stable about 28°C, with severe drop on day 8 and day 38 where it reaches 22° and 20°C respectively.
- The TDS displayed a relatively stable profile, fluctuating within the range 580 to 620 ppm. However, a significant deviation was recorded on the day 33 and 48.

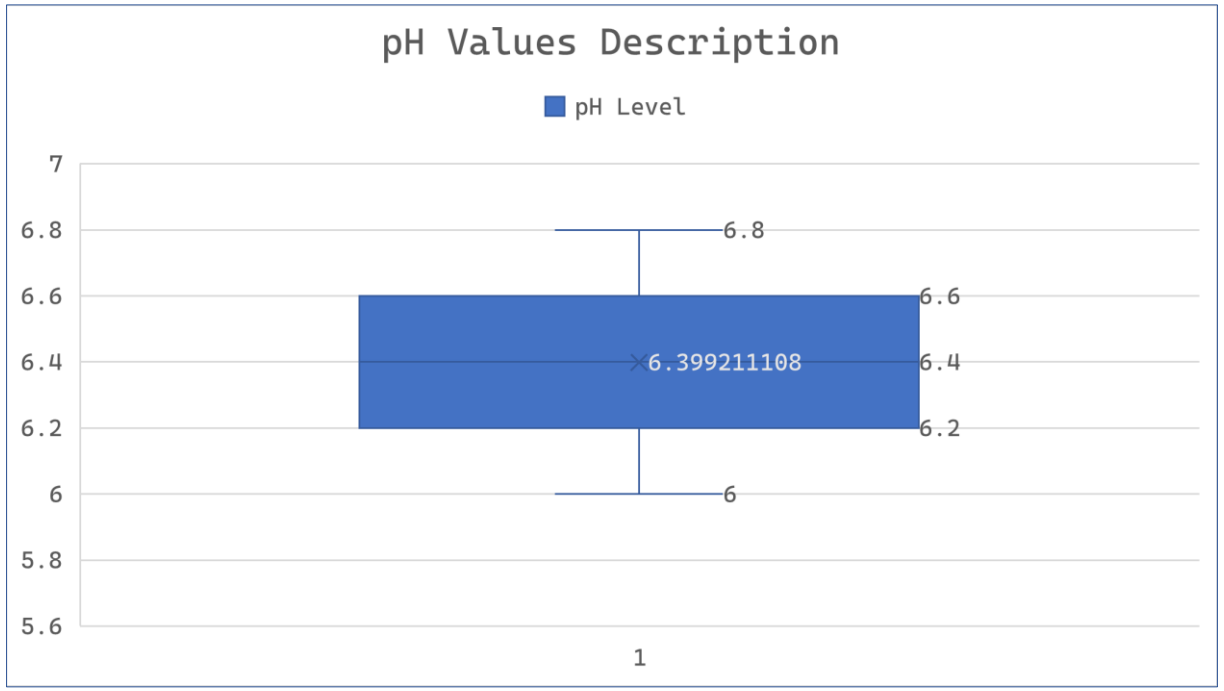
Analysis


Growth days distribution
42
3
4
2
45 Growth Days
46 Growth Days
47 Grwoth Days
48 Growth Days

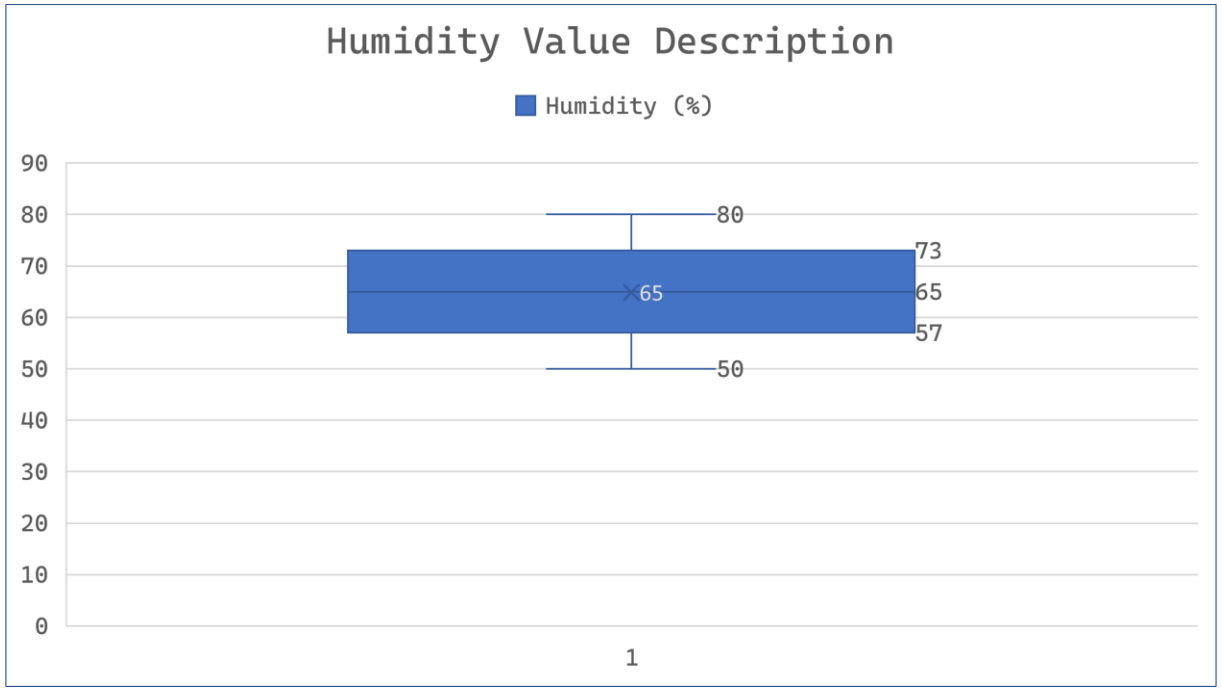
Analysis


TDS Values Description
TDS Value (ppm)
900
800
700
600
500
400
300
200
100
0
800
699
593
598
498
400
1

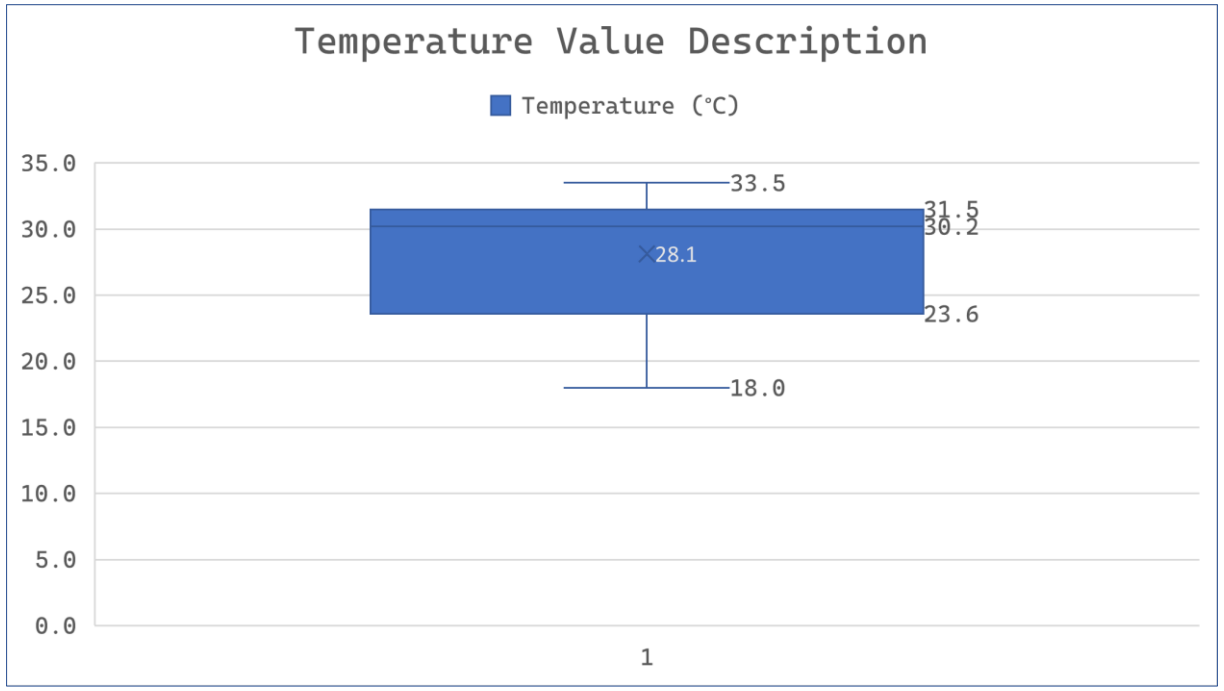
Analysis


pH Values Description
pH Level
7
6.8
6.6
6.4
6.2
6
5.8
5.6
6.8
6.6
6.399211108
6.4
6.2
6
1

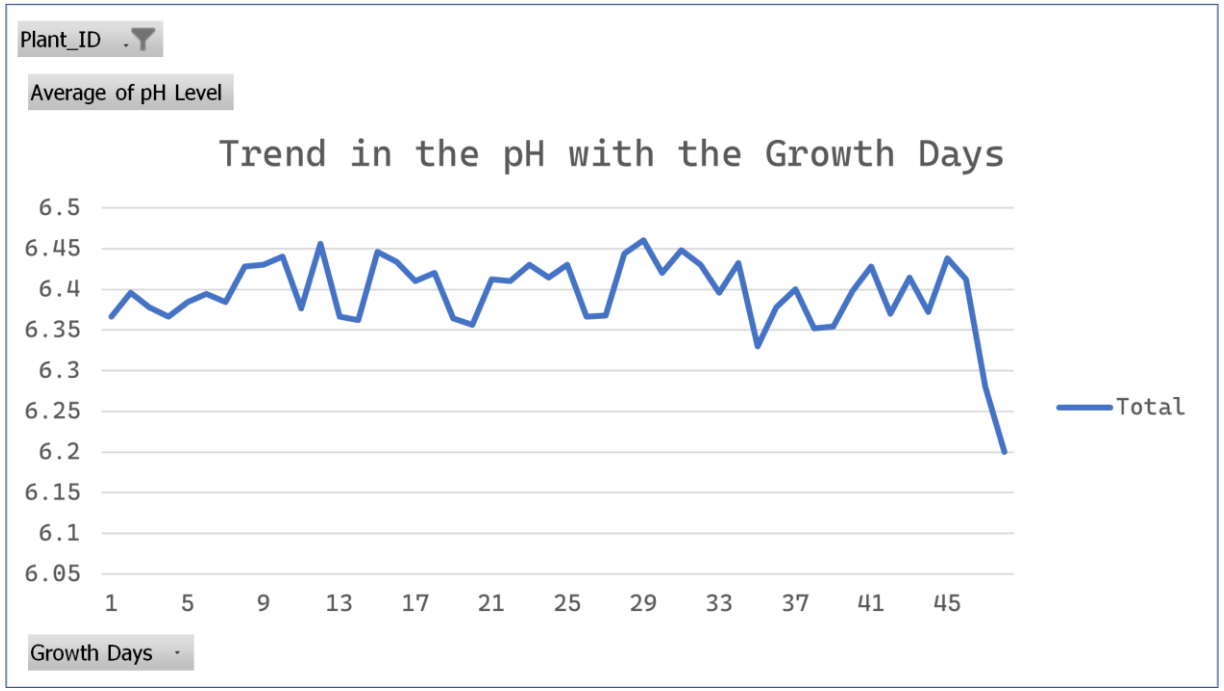
Analysis


Humidity Value Description
Humidity (%)
90
80
70
60
50
40
30
20
10
0
80
73
65
65
57
50
1

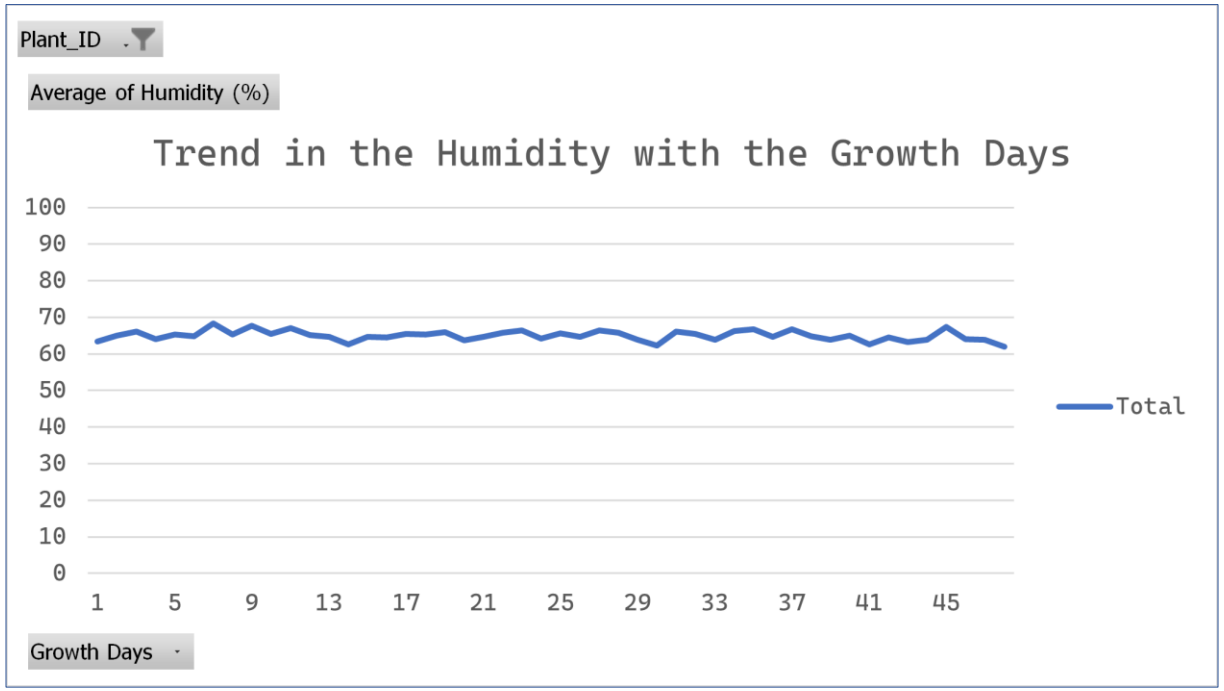
Analysis


Temperature Value Description
Temperature (°C)
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
33.5
31.5
30.2
28.1
23.6
18.0
1

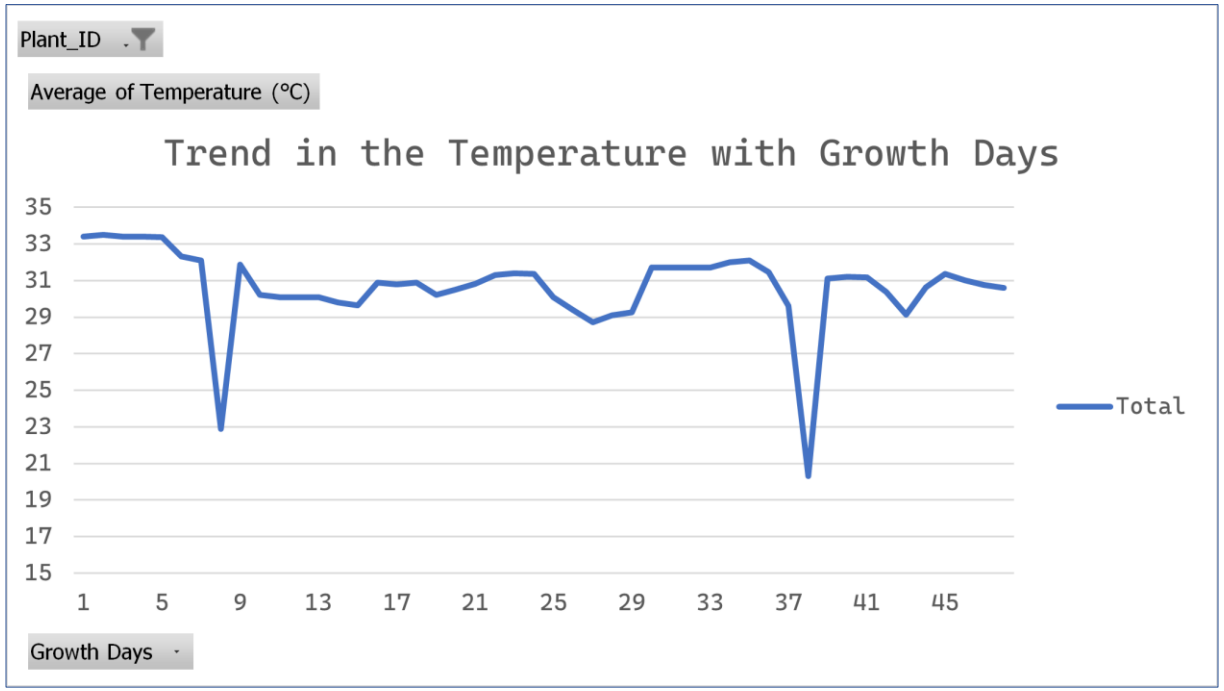
Analysis


Plant_ID
Average of pH Level
Trend in the pH with the Growth Days
6.5
6.45
6.4
6.35
6.3
6.25
6.2
6.15
6.1
6.05
1
5
9
13
17
21
25
29
33
37
41
45
Total
Growth Days

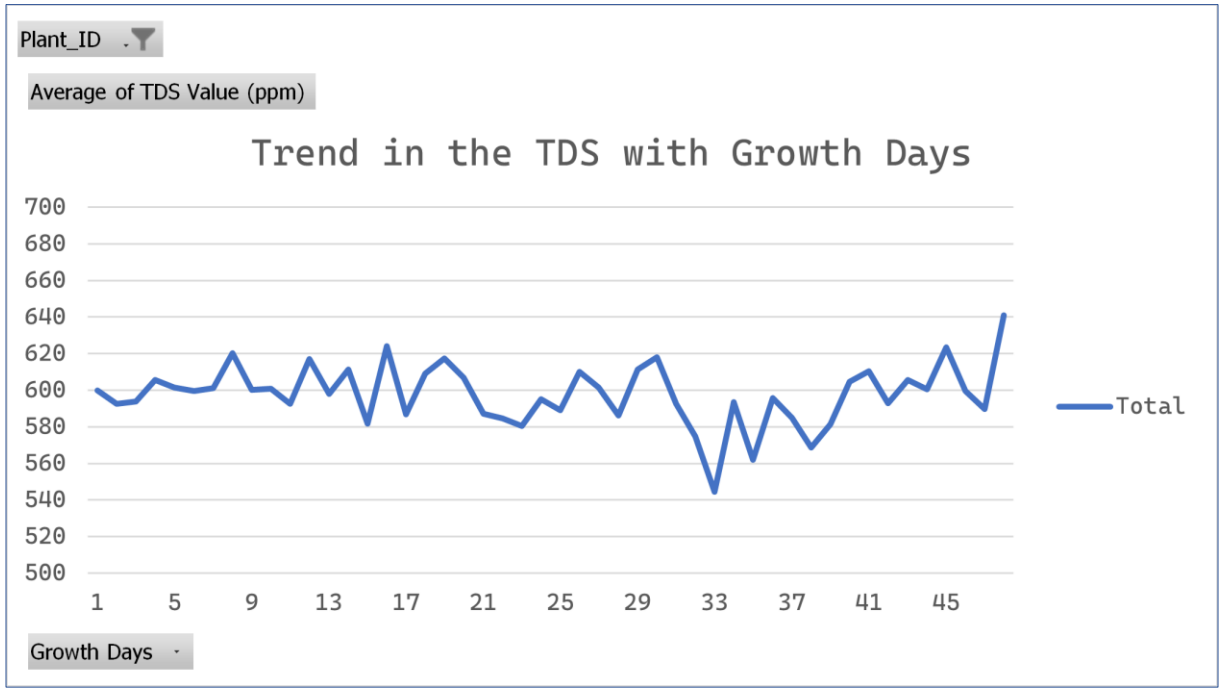
Analysis


Plant_ID
Average of Humidity (%)
Trend in the Humidity with the Growth Days
100
90
80
70
60
50
40
30
20
10
0
1
5
9
13
17
21
25
29
33
37
41
45
Total
Growth Days

Analysis


Plant_ID
Average of Temperature (°C)
Trend in the Temperature with Growth Days
35
33
31
29
27
25
23
21
19
17
15
1
5
9
13
17
21
25
29
33
37
41
45
Total
Growth Days

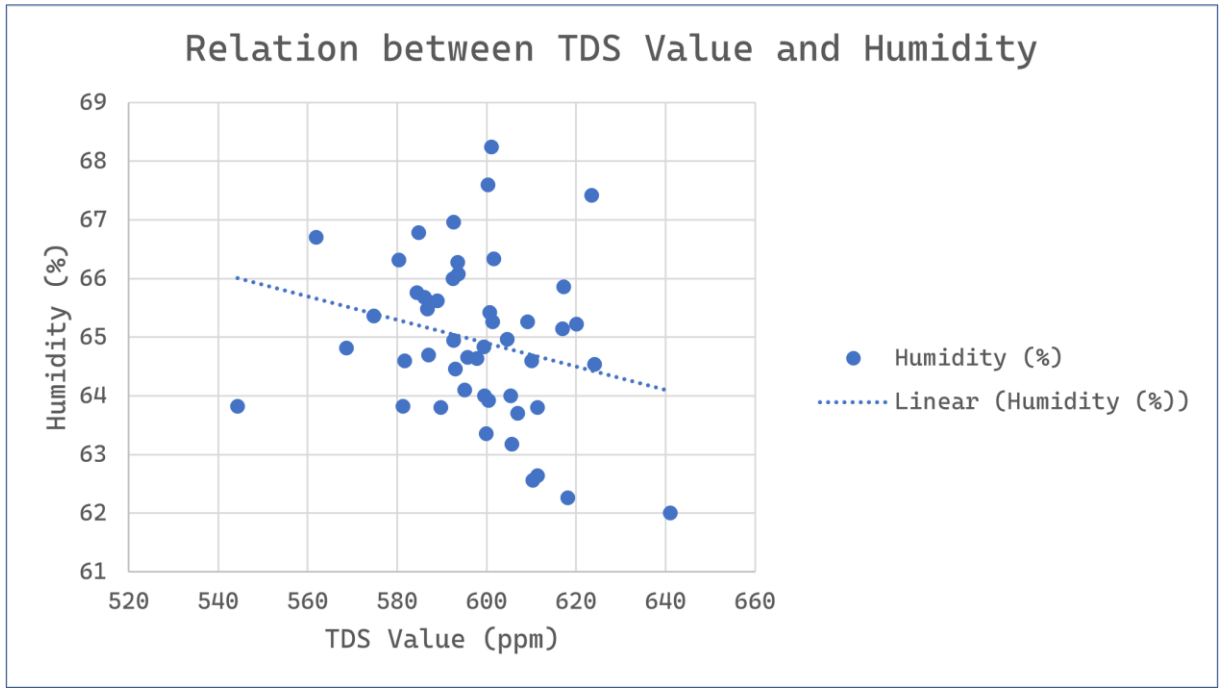
Analysis


Plant_ID
Average of TDS Value (ppm)
Trend in the TDS with Growth Days
700
680
660
640
620
600
580
560
540
520
500
1
5
9
13
17
21
25
29
33
37
41
45
Total
Growth Days

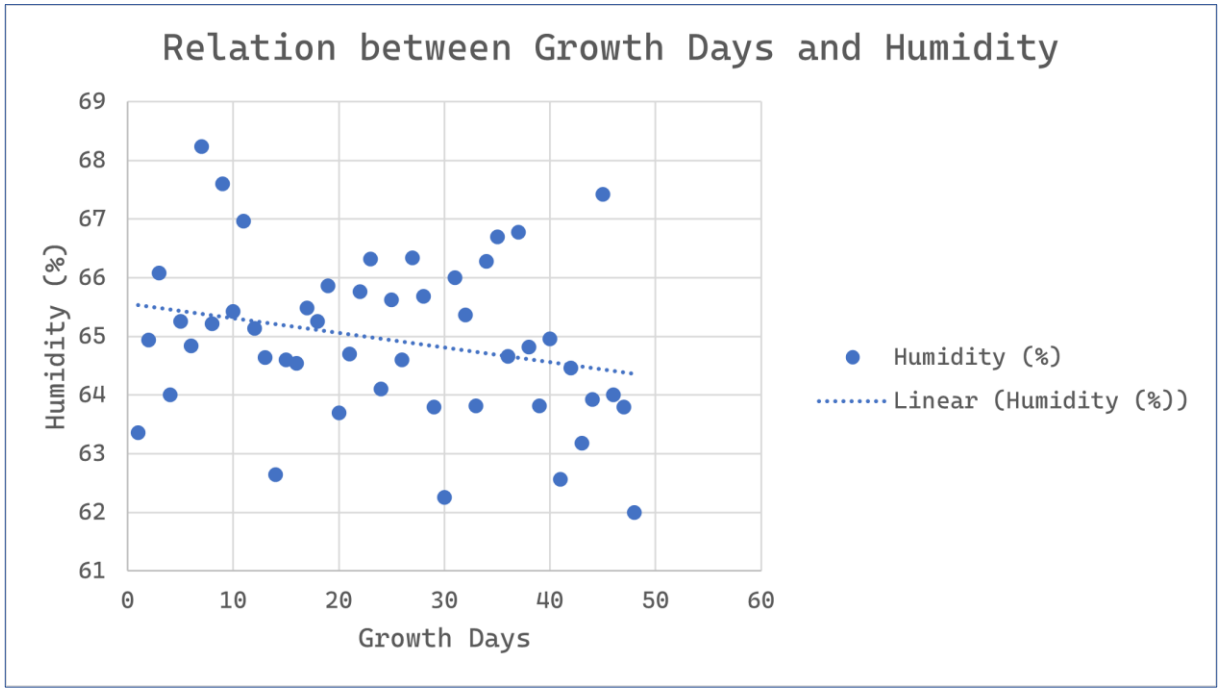
Analysis

Fields	Correlation
Growth Days/pH	-0.222094003
Growth Days/TDS	-0.085829256
Growth Days/Humidity	0.2544166
Growth Days/Temperature	-0.198258478
pH/TDS	-0.079982341
pH/Humidity	0.304372321
pH/Temperature	-0.001541007
TDS/Humidity	-0.243289718
TDS/Temperature	-0.002434058
Humidity/Temperature	0.040068941

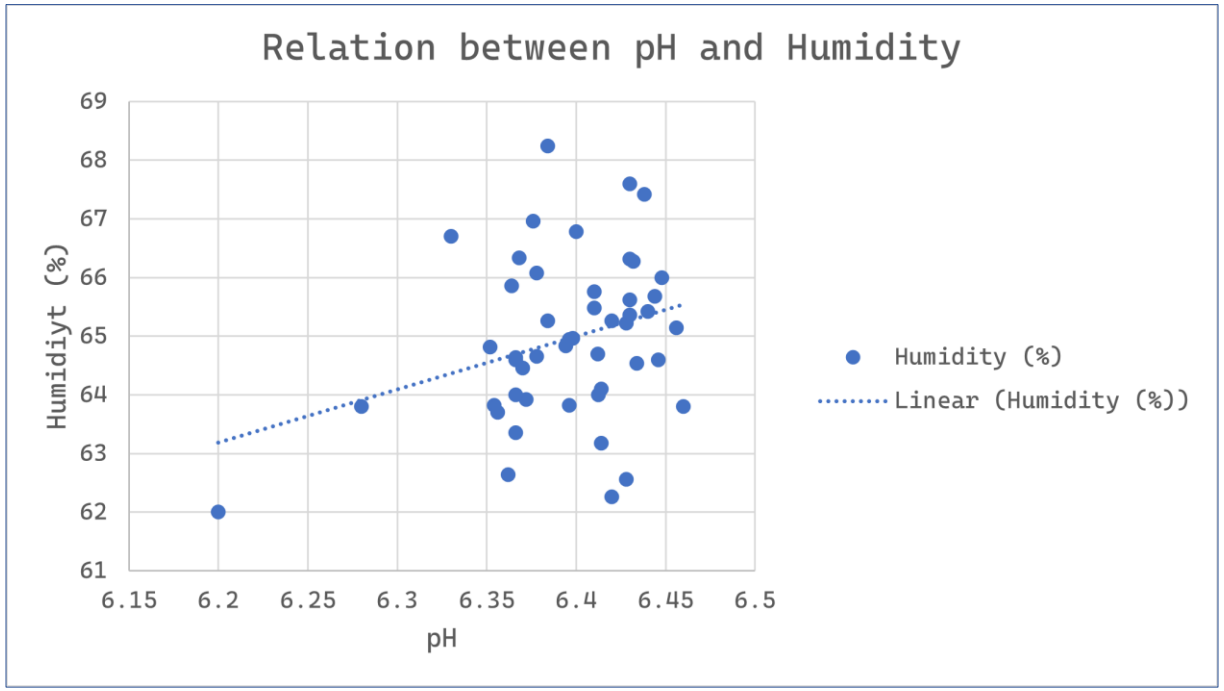
Analysis


Relation between TDS Value and Humidity
Humidity (%)
69
68
67
66
65
64
63
62
61
520
540
560
580
600
620
640
660
TDS Value (ppm)
Humidity (%)
Linear (Humidity (%))

Analysis


Relation between Growth Days and Humidity
Humidity (%)
69
68
67
66
65
64
63
62
61
0
10
20
30
40
50
60
Growth Days
Humidity (%)
Linear (Humidity (%))

Analysis


Relation between pH and Humidity
Humidiyt (%)
69
68
67
66
65
64
63
62
61
6.15
6.2
6.25
6.3
6.35
6.4
6.45
6.5
pH
Humidity (%)
Linear (Humidity (%))

Thank You!