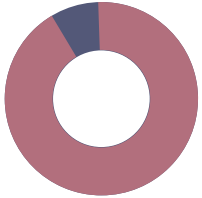
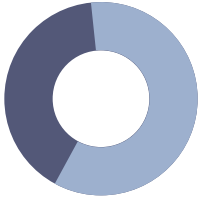
Forecasting Rainfall to Optimise Smart Water Sprinklers

AARON TOH & SHAO QI | SC1015 GROUP PROJECT

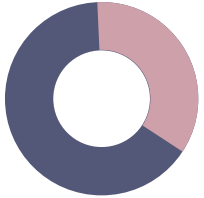
SCOPE



PROBLEM MOTIVATION



DATA PREPARATION



EXPLORATORY ANALYSIS

MACHINE LEARNING

IMPROVEMENT/ TAKEAWAYS

Singapore's Water Supply Situation

01

Only 1% of land area used for collecting rainwater

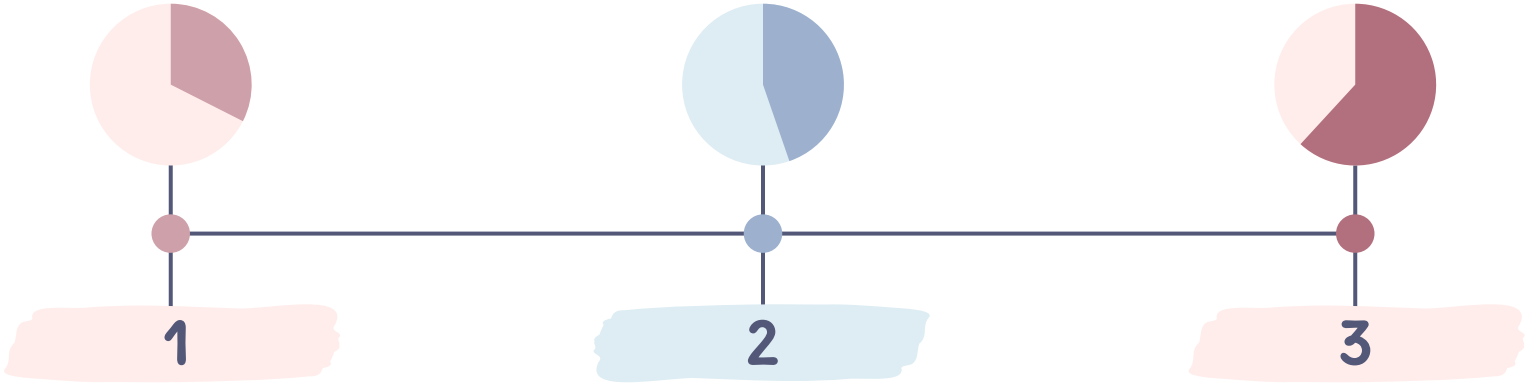
02

Heavy reliance on important water from neighboring countries

03

Expected doubling of water demand by 2060

The Urgent need for Water Conservation in Singapore



1

Devastating consequences for Singapore's economy

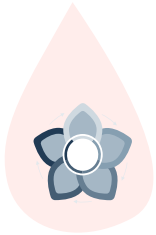
2

Climate change and population growth

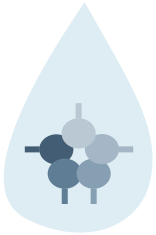
3

Water scarcity challenge by 2030 if govt does not improve water efficiency

Some water conservation measures



NEWater



Water recycling & desalination plants

Efficient irrigation systems - Smart Water Sprinklers

SMART WATER SPRINKLERS

OUR VISION

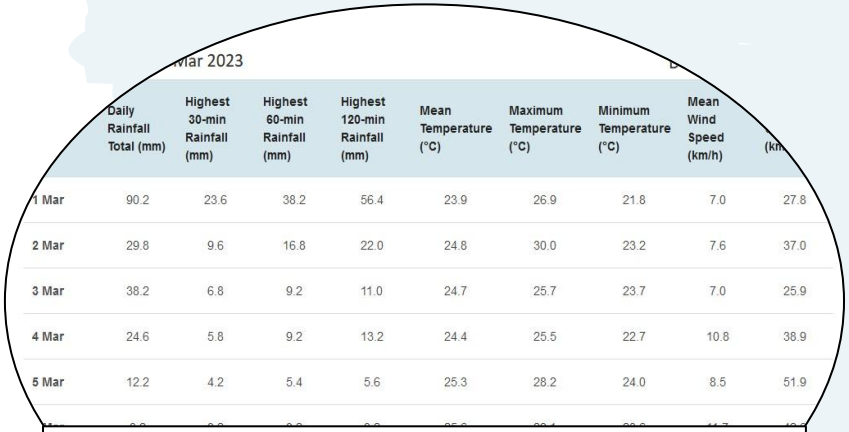
To collect past rainfall data and develop a model for predicting future rainfall. By analysing past rainfall data, we uncover patterns and trends that will help us create a more accurate model for predicting future rainfall. With this information, we can make informed decisions about water usage and conservation, and ensure that we are prepared for any weather changes that may affect our operations.

Rainfall prediction for smart water sprinklers

- Help water sprinklers conserve water
- Adjust their watering schedules based on upcoming weather patterns
- Ensures plants receive only the water they need
- Reducing our water usage
- Conserve precious resources for future generations.

DATA COLLECTION

DATA CLEANING


Mar 2023

	Daily Rainfall Total (mm)	Highest 30-min Rainfall (mm)	Highest 60-min Rainfall (mm)	Highest 120-min Rainfall (mm)	Mean Temperature (°C)	Maximum Temperature (°C)	Minimum Temperature (°C)	Mean Wind Speed (km/h)	
1 Mar	90.2	23.6	38.2	56.4	23.9	26.9	21.8	7.0	27.8
2 Mar	29.8	9.6	16.8	22.0	24.8	30.0	23.2	7.6	37.0
3 Mar	38.2	6.8	9.2	11.0	24.7	25.7	23.7	7.0	25.9
4 Mar	24.6	5.8	9.2	13.2	24.4	25.5	22.7	10.8	38.9
5 Mar	12.2	4.2	5.4	5.6	25.3	28.2	24.0	8.5	51.9

DATA PREPARATION

DATA COLLECTION

Daily Records

Location: Changi Mar 2023 Download as:CSV | PDF

Date	Daily Rainfall Total (mm)	Highest 30-min Rainfall (mm)	Highest 60-min Rainfall (mm)	Highest 120-min Rainfall (mm)	Mean Temperature (°C)	Maximum Temperature (°C)	Minimum Temperature (°C)	Mean Wind Speed (km/h)	Max Wind Speed (km/h)
1 Mar	90.2	23.6	38.2	56.4	23.9	26.9	21.8	7.0	27.8
2 Mar	29.8	9.6	16.8	22.0	24.8	30.0	23.2	7.6	37.0
3 Mar	38.2	6.8	9.2	11.0	24.7	25.7	23.7	7.0	25.9
4 Mar	24.6	5.8	9.2	13.2	24.4	25.5	22.7	10.8	38.9
5 Mar									
6 Mar									
7 Mar									
8 Mar									
9 Mar									

Daily Observations

Time	Temperature (°F)			Dew Point (°F)			Humidity (%)			Wind Speed (mph)			Pressure (in)			Precipitation
Mar	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Total
1	79	74.7	72	75	72.2	70	100	91.7	83	9	5.2	1	29.9	29.9	29.8	0.00
2	84	76.3	73	75	72.7	72	100	88.5	66	15	5.4	0	29.9	29.9	29.8	0.00
3	79	76.3	75	75	74.1	73	94	93.5	89	16	5.6	2	29.9	29.9	29.8	0.00
4	77	75.4	73	75	73.0	72	94	92.4	89	17	8.5	0	30.0	29.9	29.9	0.00
5	82	77.3	75	75	73.5	72	94	89.1	79	15	6.2	2	30.0	29.9	29.9	0.00
6	82	77.8	73	75	72.8	72	94	84.8	74	17	9.3	3	29.9	29.9	29.8	0.00
7	84	78.7	75	77	73.3	72	94	84.4	74	16	10.6	5	29.9	29.9	29.8	0.00
8	86	80.2	75	77	74.4	73	94	83.8	70	18	10.6	3	29.9	29.9	29.8	0.00
9	88	80.9	77	77	74.4	73	94	82.0	66	17	11.5	7	29.9	29.8	29.7	0.00
10	86	80.9	77	77	75.1	73	94	83.5	70	17	12.0	7	29.9	29.8	29.7	0.00

Our Data Sources:

- **Rainfall**
- **Temperature**

(http://www.weather.gov.sg)

- **Humidity**

(https://www.wunderground.com)

Period of Collection:
Apr 2013 to Mar 2023

DATA CLEANING

Daily Records

Location: Changi Mar 2023

Download as: CSV | PDF

Date	Daily Rainfall Total (mm)	Highest 30-min Rainfall (mm)	Highest 60-min Rainfall (mm)	Highest 120-min Rainfall (mm)	Mean Temperature (°C)	Maximum Temperature (°C)	Minimum Temperature (°C)	Mean Wind Speed (km/h)	Max Wind Speed (km/h)
1 Mar	90.2	23.6	38.2	56.4	23.9	26.9	21.8	7.0	27.8
2 Mar	29.8	9.6	16.8	22.0	24.8	30.0	23.2	7.6	37.0
3 Mar	38.2	6.8	9.2	11.0	24.7	25.7	23.7	7.0	25.9
4 Mar	24.6	5.8	9.2	13.2	24.4	25.5	22.7	10.8	38.9
5 Mar	12.2	4.2	5.4	5.6	25.3	28.2	24.0	8.5	51.9
6 Mar	0.2	0.2	0.2	0.2	25.6	29.1	23.6	11.7	42.6
7 Mar	0.2	0.2	0.2	0.2	26.5	29.7	24.2	11.7	37.0
8 Mar	0.0	0.0	0.0	0.0	27.1	30.9	24.8	14.9	42.6
9 Mar	0.0	0.0	0.0	0.0	27.6	31.2	25.3	16.1	40.7

Daily Observations

Time	Temperature (°F)			Dew Point (°F)			Humidity (%)			Wind Speed (mph)			Pressure (in)			Precipitation
Mar	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg	Min	Total
1	79	74.7	72	75	72.2	70	100	91.7	83	9	5.2	1	29.9	29.9	29.8	0.00
2	84	76.3	73	75	72.7	72	100	88.5	66	15	5.4	0	29.9	29.9	29.8	0.00
3	79	76.3	75	75	74.1	73	94	93.5	89	16	5.6	2	29.9	29.9	29.8	0.00
4	77	75.4	73	75	73.0	72	94	92.4	89	17	8.5	0	30.0	29.9	29.9	0.00
5	82	77.3	75	75	73.5	72	94	89.1	79	15	6.2	2	30.0	29.9	29.9	0.00
6	82	77.8	73	75	72.8	72	94	84.8	74	17	9.3	3	29.9	29.9	29.8	0.00
7	84	78.7	75	77	73.3	72	94	84.4	74	16	10.6	5	29.9	29.9	29.8	0.00
8	86	80.2	75	77	74.4	73	94	83.8	70	18	10.6	3	29.9	29.9	29.8	0.00
9	88	80.9	77	77	74.4	73	94	82.0	66	17	11.5	7	29.9	29.8	29.7	0.00
10	86	80.9	77	77	75.1	73	94	83.5	70	17	12.0	7	29.9	29.8	29.7	0.00
11	88	80.8	77	77	74.9	73	94	83.5	70	15	10.9	8	29.9	29.8	29.8	0.00
12	88	80.8	75	77	73.8	73	94	81.0	66	15	9.9	6	29.9	29.8	29.7	0.00
13	86	80.2	77	73	72.4	70	89	78.1	62	21	13.3	7	29.9	29.8	29.7	0.00
14	84	79.5	75	73	72.4	72	[illegible]	[illegible]	[illegible]	17	11.6	6	29.8	29.8	29.7	0.00

Columns of Concern:

- Date
- Daily Rainfall Total (mm)
- Mean Temperature (°C)
- Avg Humidity (%)

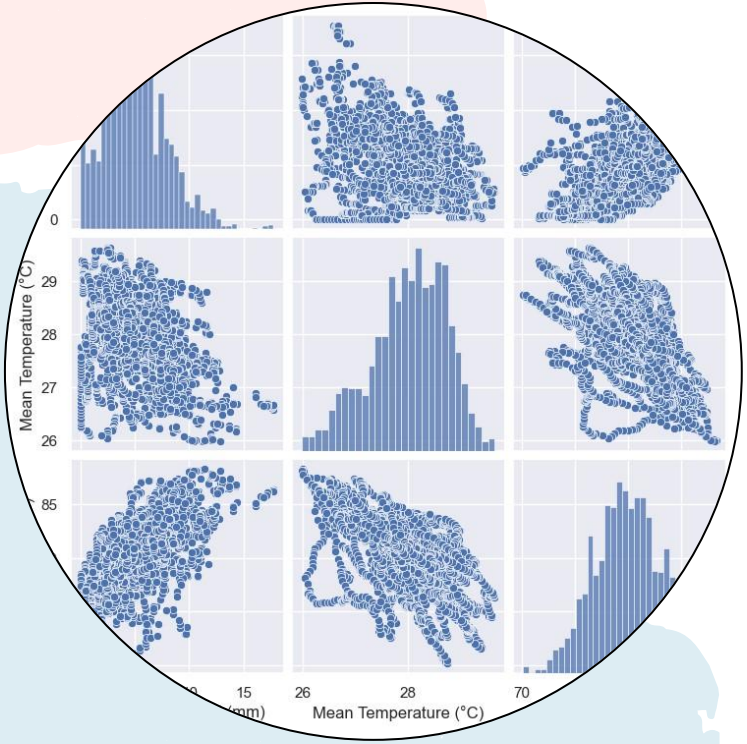
DATA CLEANING

```
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Station                         3652 non-null   object
 1   Year                            3652 non-null   int64
 2   Month                           3652 non-null   int64
 3   Day                             3652 non-null   int64
 4   Daily Rainfall Total (mm)       3652 non-null   float64
 5   Highest 30 min Rainfall (mm)    3652 non-null   object
 6   Highest 60 min Rainfall (mm)    3652 non-null   object
 7   Highest 120 min Rainfall (mm)   3652 non-null   object
 8   Mean Temperature (°C)           3652 non-null   float64
 9   Maximum Temperature (°C)        3652 non-null   float64
 10  Minimum Temperature (°C)        3652 non-null   float64
 11  Mean Wind Speed (km/h)          3652 non-null   object
 12  Max Wind Speed (km/h)           3652 non-null   object
 13  Mean Humidity (%)               3652 non-null   float64
dtypes: float64(5), int64(3), object(6)
memory usage: 428.0+ KB
```

Year	Month	Day	Daily Rainfall Total (mm)	Mean Temperature (°C)	Mean Humidity (%)
2023	3	27	19.400000	26.000000	90.700000
2023	3	28	0.000000	28.800000	80.600000
2023	3	29	0.000000	28.800000	79.700000
2023	3	30	9.200000	28.400000	80.300000
2023	3	31	0.000000	28.700000	80.100000

```
print("Rainfall with Negative Value: ", len(rainData[rainData['Daily Rainfall Total (mm)'] < 0]))
print("Temperature with Negative Value: ", len(rainData[rainData['Mean Temperature (°C)'] < 0]))
print("Humidity with Negative Value: ", len(rainData[rainData['Mean Humidity (%)'] < 0]))
```

```
Rainfall with Negative Value:  0
Temperature with Negative Value:  0
Humidity with Negative Value:  0
```



EXPLORATORY DATA ANALYSIS

01 BACKGROUND INFORMATION

02 REMOVING OUTLIERS

03 SMOOTHING NOISE

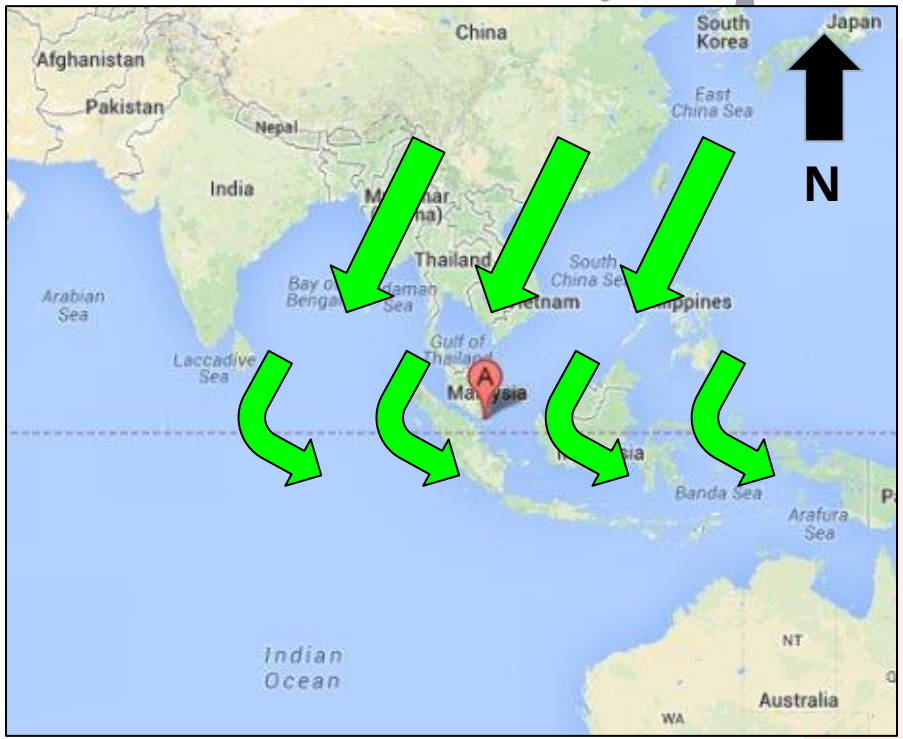
04 FINDING TREND

05 OTHER VARIABLES (FINDING CORRELATION)

BACKGROUND INFORMATION

High Rainfall During the following periods:

- **Nov to Jan (NE Monsoon)**
- **May to Jul (SW Monsoon)**



BACKGROUND INFORMATION

High Rainfall During the following periods:

- Nov to Jan (NE Monsoon)
- May to Jul (SW Monsoon)

REMOVING OUTLIERS

1

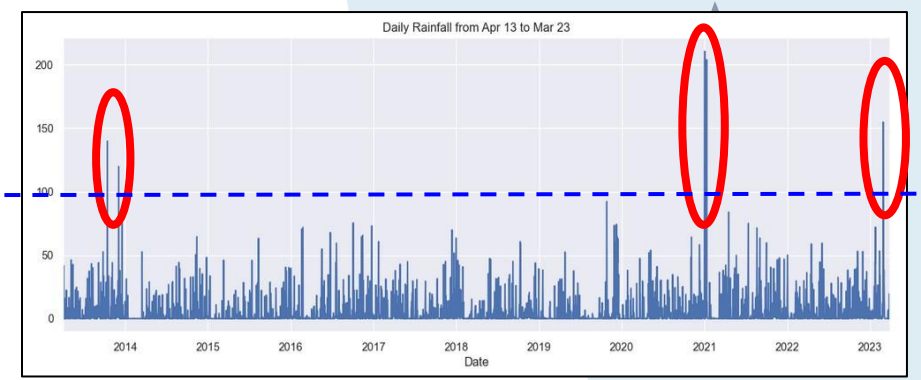
BETTER DATA ANALYSIS

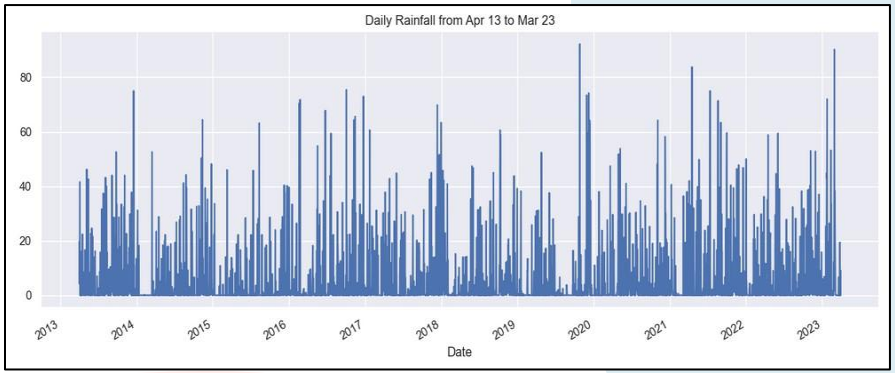
As they may alter the seasonal trend

2

BETTER MACHINE LEARNING

Lowering our RMSE





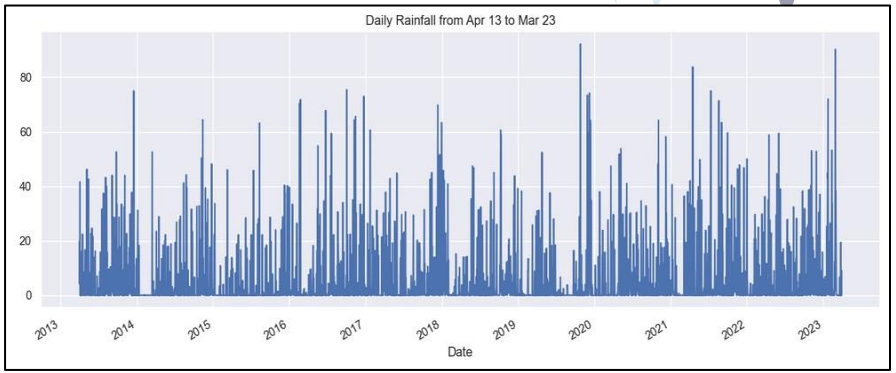
SMOOTHING NOISE

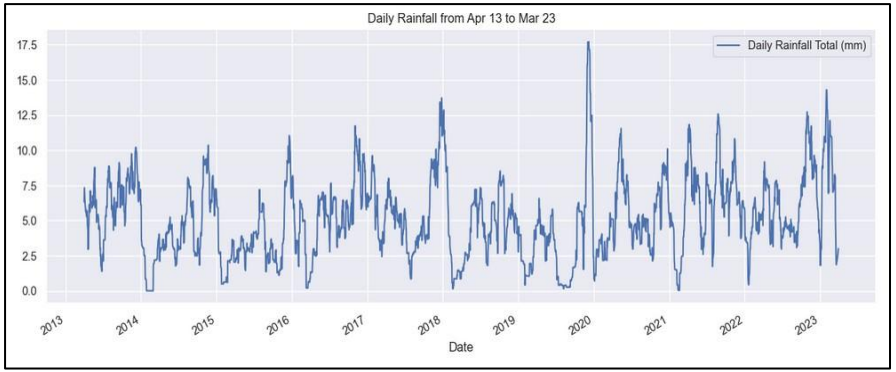
We want to find the annual trend across different months

Comparing daily observations only tells us nothing

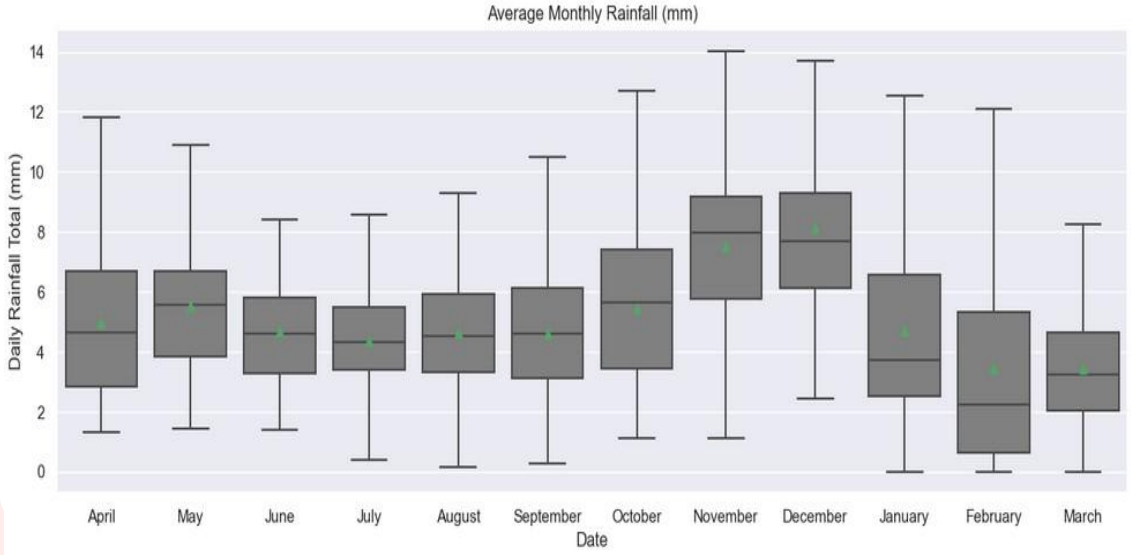
- No daily trend
- Data fluctuates erratically

A moving average of 30 days give us a better picture of the rainfall pattern across the year





FINDING TREND

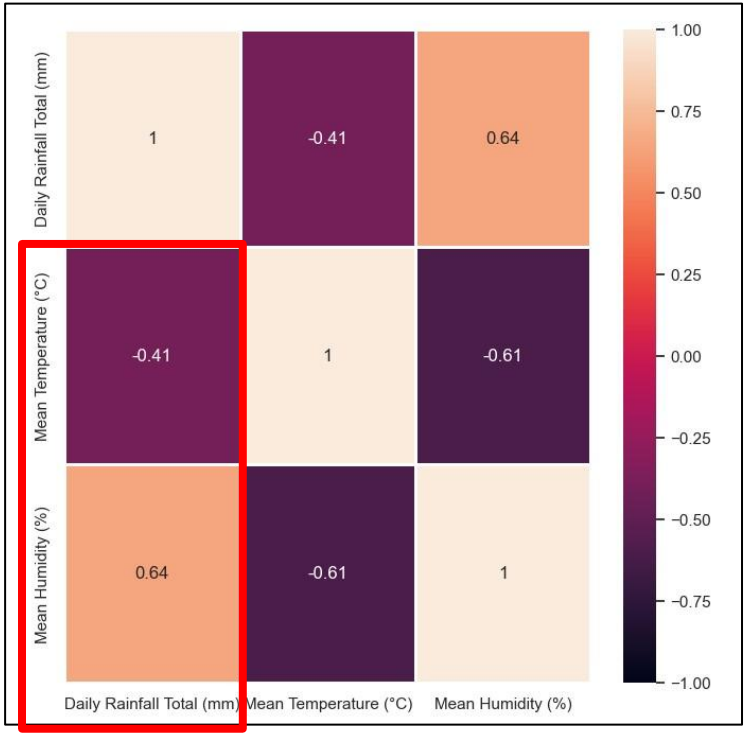


Using boxplots to represent our monthly rainfall across the year, we see that

- **Apr to Jun**
- **Oct to Dec**

are months where rainfall are higher than average.

OTHER VARIABLES

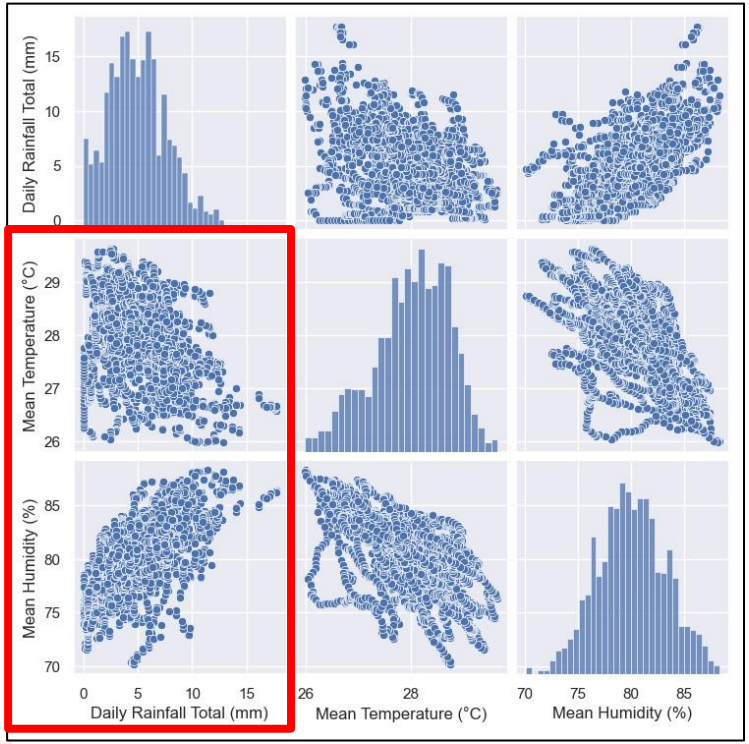


Generally, we see that rainfall increases when:

- **Mean Temperature (°C) ↓**
- **Avg Humidity (%) ↑**

***Note that this may not always be the case!**

OTHER VARIABLES



Generally, we see that rainfall increases when:

- **Mean Temperature (°C) ↓**
- **Avg Humidity (%) ↑**

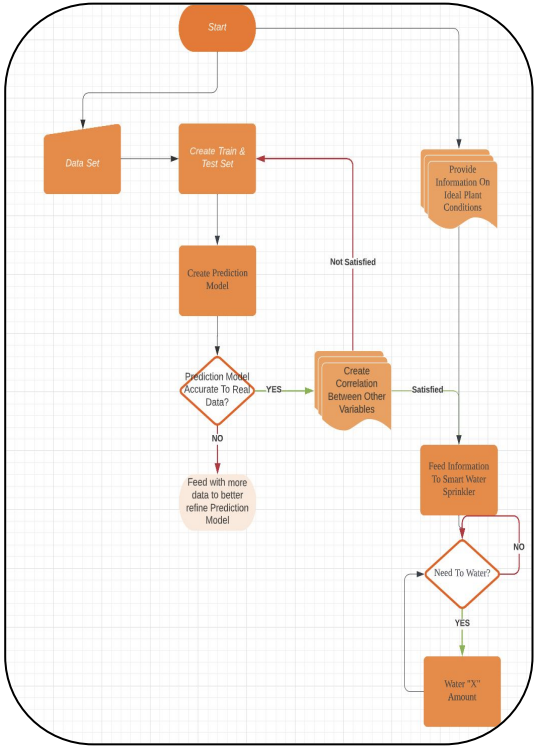
***Note that this may not always be the case!**

MACHINE LEARNING

A **TIME SERIES SPLIT**

B **CHOICE OF MODEL**

C **SMOOTHING NOISE**



D **LAGGED FEATURES**

E **TEMPERATURE & HUMIDI**

F **FORECAST WATER REQUIREMENTS**

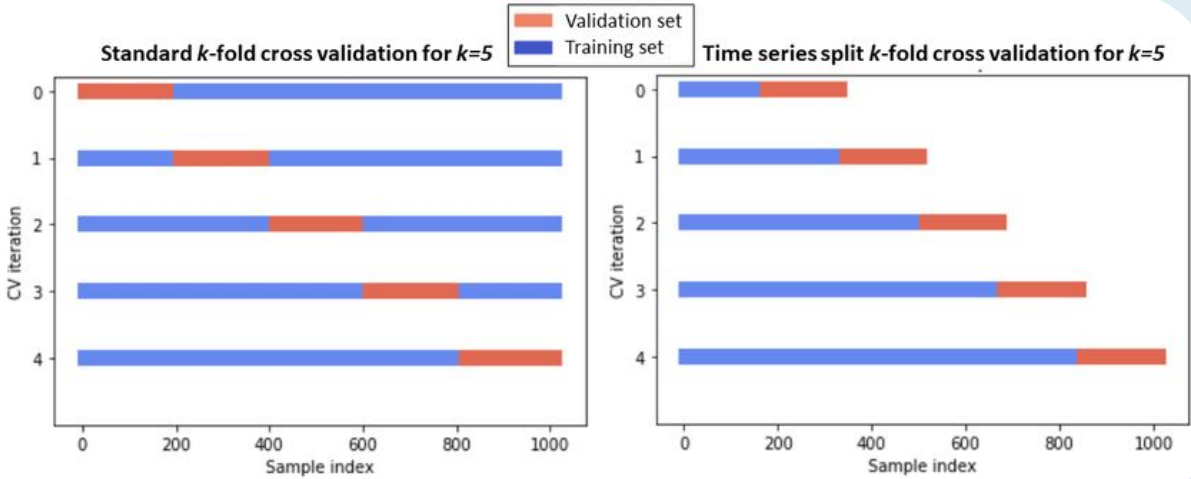
TIME SERIES SPLIT

Time Series Split

- Splits Dataset into N Folds
- Train Data of Increasing Time Interval
- Learns Test Data of Previous Fold
- Learns Dependence of Past Observations

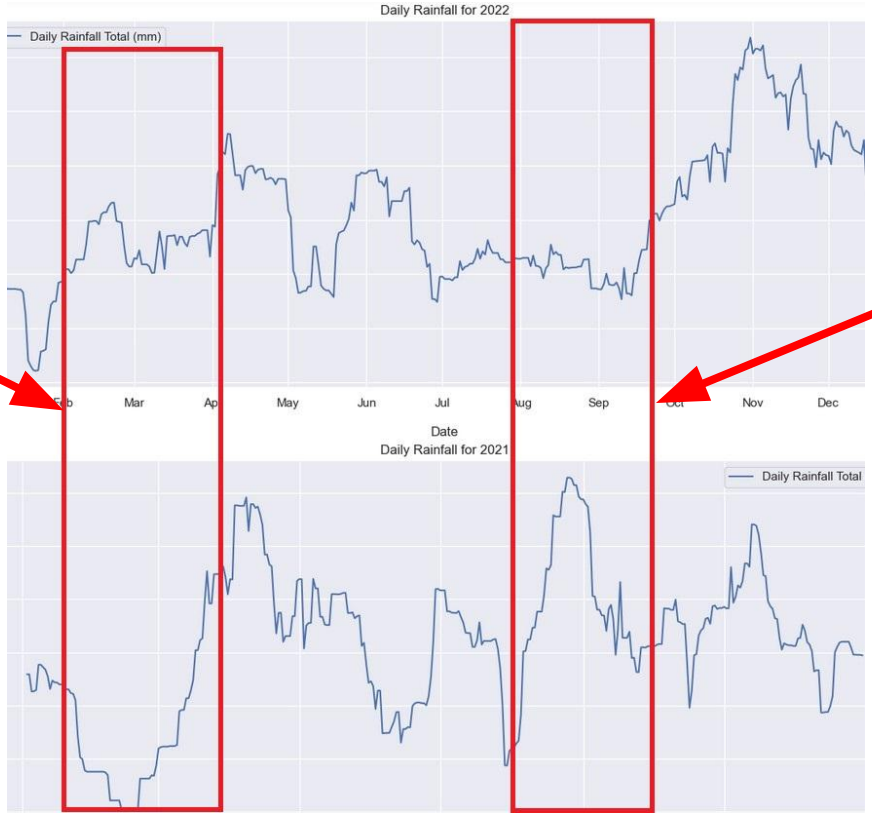
K-Fold Cross Validation

- Splits Data in K Random Folds
- Uses the (K+1) from the Kth Fold as Test Set



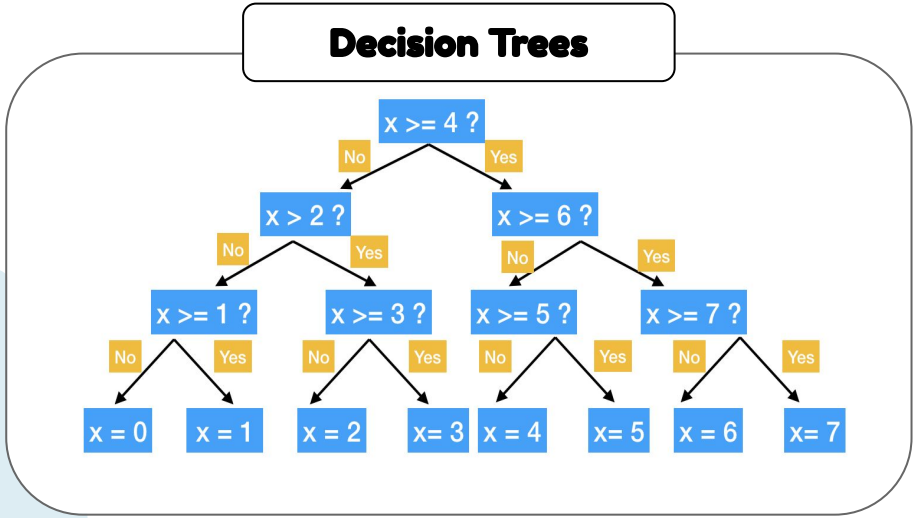
CHOICE OF MODEL: XGBOOST

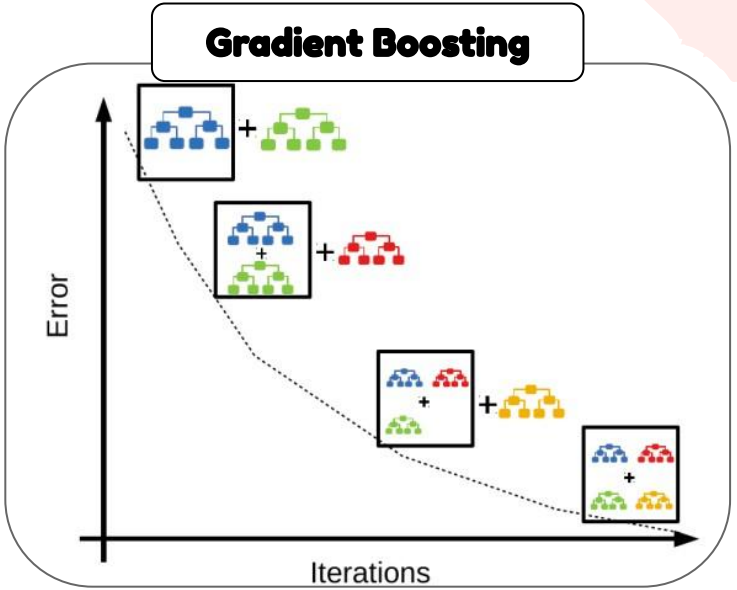
Pattern Irregularities



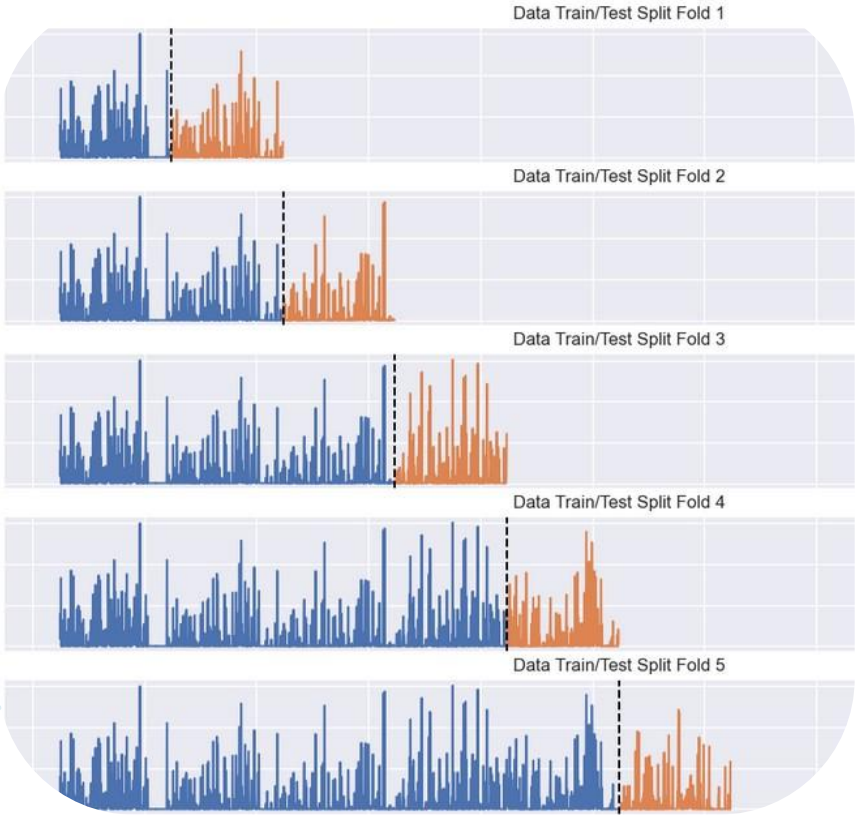
Pattern Irregularities

CHOICE OF MODEL: XGBOOST



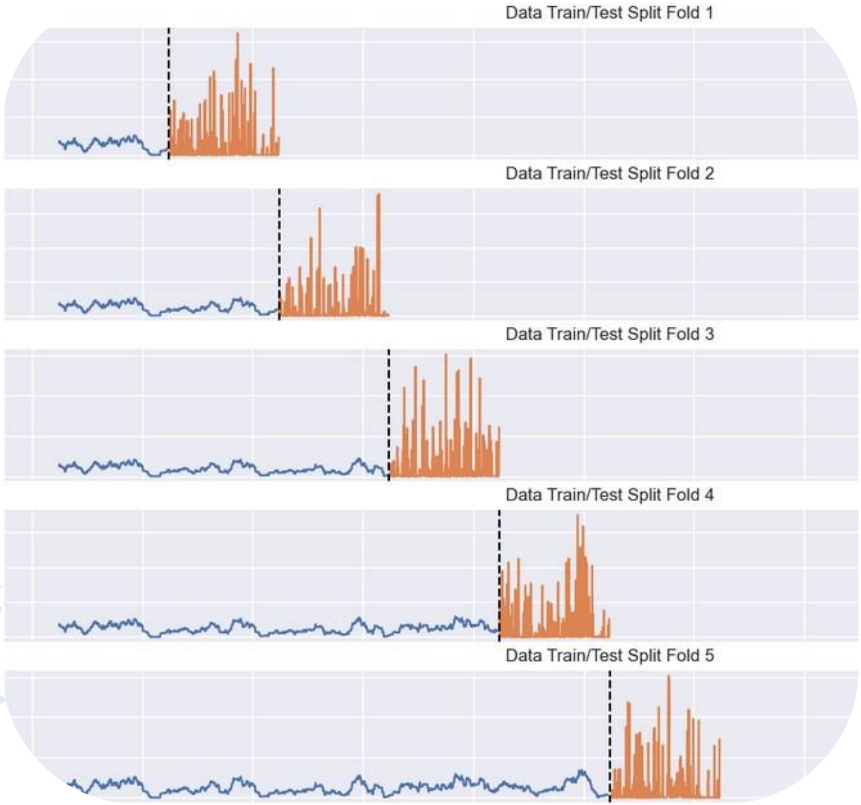


SMOOTHING NOISE



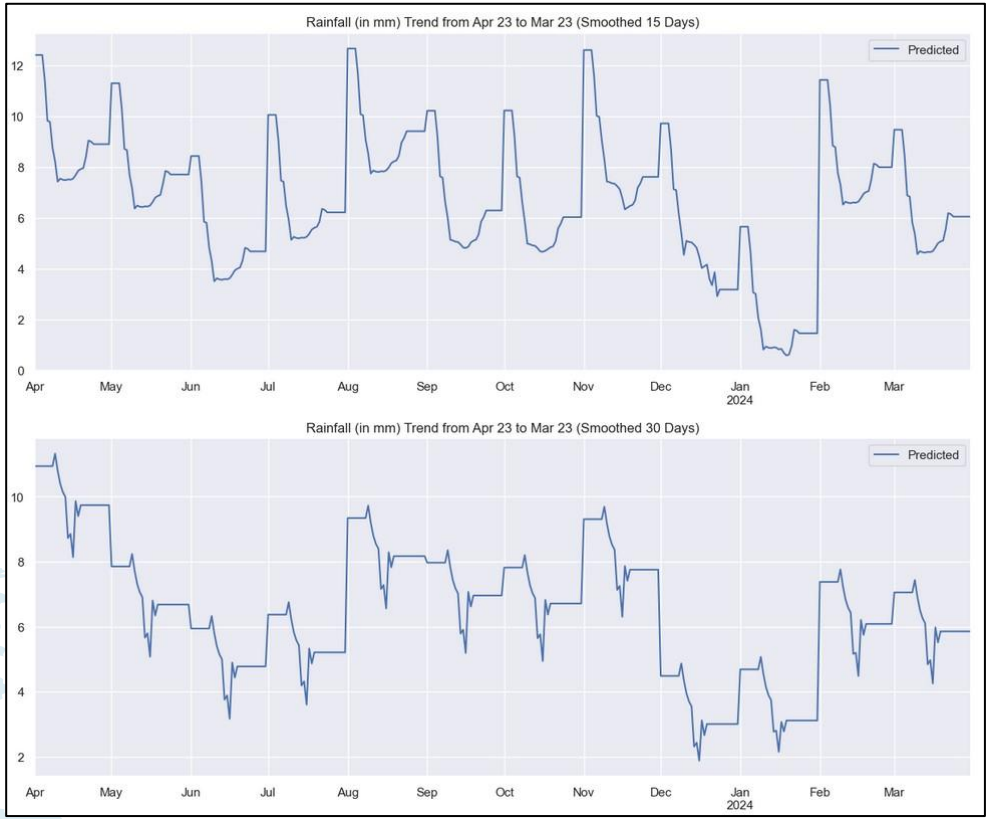
- Noisy dataset due to daily rainfall fluctuations
- No daily trend
- Moving average to smooth noise that may affect prediction
- Smooth train set only

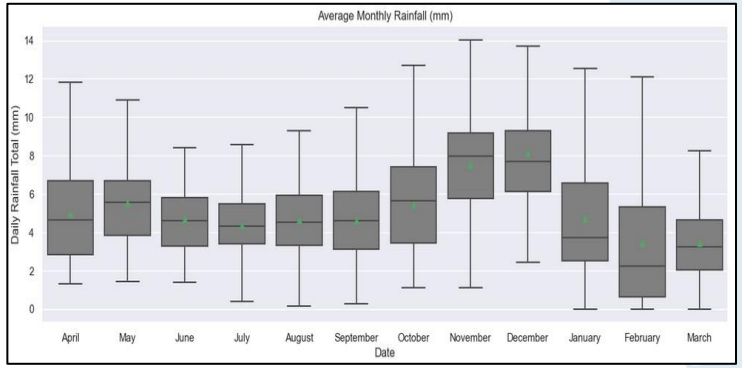
SMOOTHING NOISE



- Noisy dataset due to daily rainfall fluctuations
- No daily trend
- Moving average to smooth noise that may affect prediction
  - Unsmoothed (RMSE: 12.08)
  - 7 Days (RMSE: 11.82)
  - 15 Days (RMSE: 11.50)
  - 30 Days (RMSE: 11.35)
- Smooth train set only

SMOOTHING NOISE





Comparing Prediction Graph's Seasonal Trend
No Lag Features

LAGGED FEATURES (TIME)

Date	Daily Rainfall Total (mm)	Mean Temperature (°C)	Mean Humidity (%)	D-1_rain	D-2_rain	D-3_rain	D-4_rain	D-5_rain	D-6_rain	D-7_rain
2023-03-27	2.400000	27.720000	81.585000	2.285714	2.181818	2.086957	2.000000	1.928000	1.861538	2.244444
2023-03-28	2.526316	27.726316	81.615789	2.400000	2.285714	2.181818	2.086957	2.000000	1.928000	1.861538
2023-03-29	2.666667	27.755556	81.811111	2.526316	2.400000	2.285714	2.181818	2.086957	2.000000	1.928000
2023-03-30	2.823529	27.805882	81.988235	2.666667	2.526316	2.400000	2.285714	2.181818	2.086957	2.000000
2023-03-31	3.000000	27.850000	82.193750	2.823529	2.666667	2.526316	2.400000	2.285714	2.181818	2.086957

Need for lagged features

- Singapore's Rainfall has an annual trend
- Need to use past values of the same period to capture such pattern and predict rainfall accurately

LAGGED FEATURES (TIME)

Days	No Smooth	7 Days Smooth	15 Days Smooth	30 Days Smooth
D-3	12.87	12.31	11.70	11.27
D-7	13.03	13.39	11.71	11.45
D-14	12.87	12.79	11.85	11.43
D-28	12.72	12.81	12.08	11.36

Using Days only as Lagged Features:

- D-3
- 30 Day Smoothing Window

Gave the most accurate prediction

LAGGED FEATURES (TIME)

Months	No Smooth	7 Days Smooth	15 Days Smooth	30 Days Smooth
M-3	12.98	12.24	11.62	11.21
M-6	13.01	12.16	11.58	11.18
M-9	13.05	12.23	11.52	11.15
M-12	13.31	12.19	11.57	11.17

Adding Months as Lagged Features:

- M-9
- 30 Day Smoothing Window

Improved our model's performance

LAGGED FEATURES (TIME)

Years	No Smooth	7 Days Smooth	15 Days Smooth	30 Days Smooth
Y-3	13.18	12.21	11.54	11.16
Y-5	13.17	12.19	11.50	11.16
Y-7	13.26	12.22	11.53	11.13

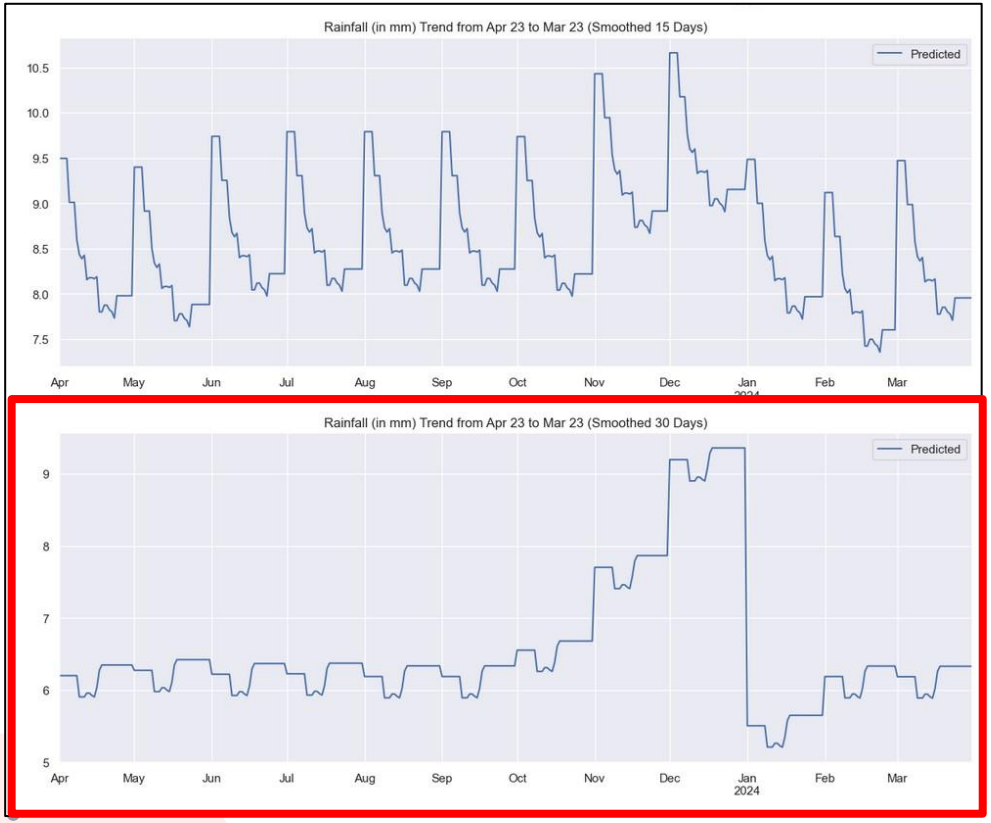
Adding Years as Lagged Features:

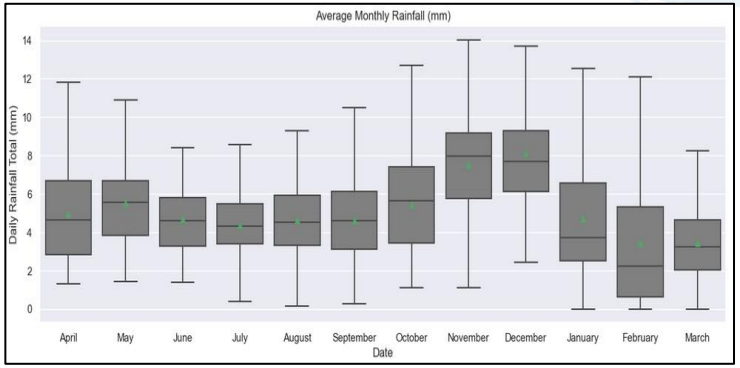
- Y-7
- 30 Day Smoothing Window

Improved our model's performance

D-3, M-9, Y-7, on a 30 Day Smoothing Window

LAGGED FEATURES (TIME)





Comparing Prediction Graph's Seasonal Trend
D-3, M-9, Y-7

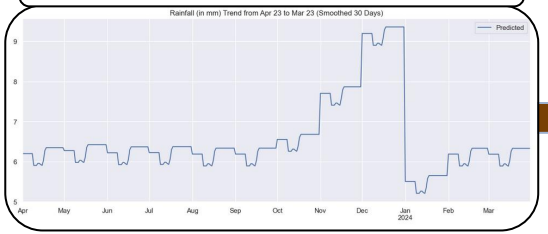
TEMPERATURE & HUMIDITY

Days	No Smooth	7 Days Smooth	15 Days Smooth	30 Days Smooth
Temperature	13.60	11.96	11.57	11.19
Humidity	12.78	12.02	11.46	11.20
Both	12.96	11.96	11.47	11.20

- We had Temperature / Humidity from our Correlation Matrix
- Adding either of them did not improved our model's performance
(Our best RMSE was **11.13**)

FORECAST WATER REQUIREMENTS

Rainfall Prediction Data

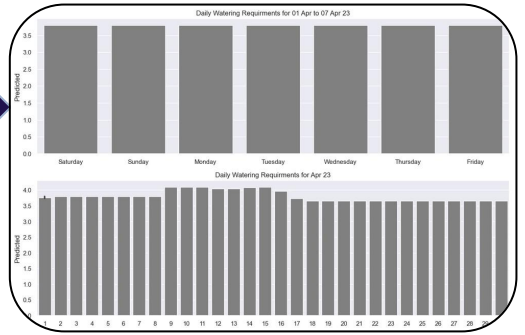


Input Crop's Watering Requirement (in mm):

Crop Water Requirements

Our Hypothetical Smart Sprinkler System

Water Required
(Adjusted for Rainfall)



TAKEAWAYS FROM THIS PROJECT

TAKEAWAYS

1. Due to its location, SG experiences higher annual rainfall compared to countries in tropical regions.
    - SG lies on the **equator** and is **surrounded by waters**
    - We expected our model to forecast **higher water requirements for crops in tropical countries**

TAKEAWAYS

2. Why did Temperature & Humidity not improve our performance?
    - **Lower Temperature does not mean Higher Rainfall** → The air may lack moisture for precipitation
    - High Humidity simply means **more moisture in air**, but **not precipitation**
    - Precipitation also affected by **Convection** and **Condensation** activities

TAKEAWAYS

3. **Wider Applications of Rainfall Forecasting.**
    - Water yield - placement of reservoirs
    - Energy production - hydroelectric power plants
    - Flood control
    - Wildlife conservation

CONCLUSION

Yield fruitful outcomes

Gain thorough comprehension of intricate predicaments

Thank You!

THIS SLIDES TEMPLATE IS PROVIDED BY: slidesgo