

Lec-1 Intro to Time Series & forecasting

- Introduction to TS
- Mobiplus case study
- Missing value imputation
- Handling Anomalies
- Decomposition of TS

Contact details

Name : Jitin Gupta

WhatsApp : 9650204730 ✓

Email : jitin.gupta-1@scaler.com

- Lecture notes:
- ① PDF of scribble notes
 - ② Zip file → PDF + *.ipyb file
of colab notebook

Introduction

TS → Stock prices, data that varies with time, historical records, sales
Data along with a timestamp

Timestamp	Sales
1 st Mar 2024	10,000
19 th Feb 2023	8000
15 th Aug 2023	5000
23 rd Jan 2022	..

Timestamp	Sales
1 st Mar 2024	10,000
1 st Apr 2024	11,000
1 st May 2024	12,000
1 st June 2024	13,000



ordered timestamps

1st Mar, 15th Apr, 29th May

Timestamp
1 st Mar 2024
1 st Apr 2024
1 st May 2024
1 st June 2024

↓
ordered timestamps

Sales
10,000
11,000
12,000
13,000

Q) What is frequency of sampling?

A) Monthly / 30 day

Sampling frequencies:

- a) Monthly
- b) Yearly
- c) Daily
- d) weekly
- e) quarterly
- f) half-Yearly
- g) hourly
- h) second by second
- i) minute by minute

Analysis of TS: ① Forecasting

Sales → next day, next week, next month etc.

→ Inventory management

Dangers : Overproduction , under production

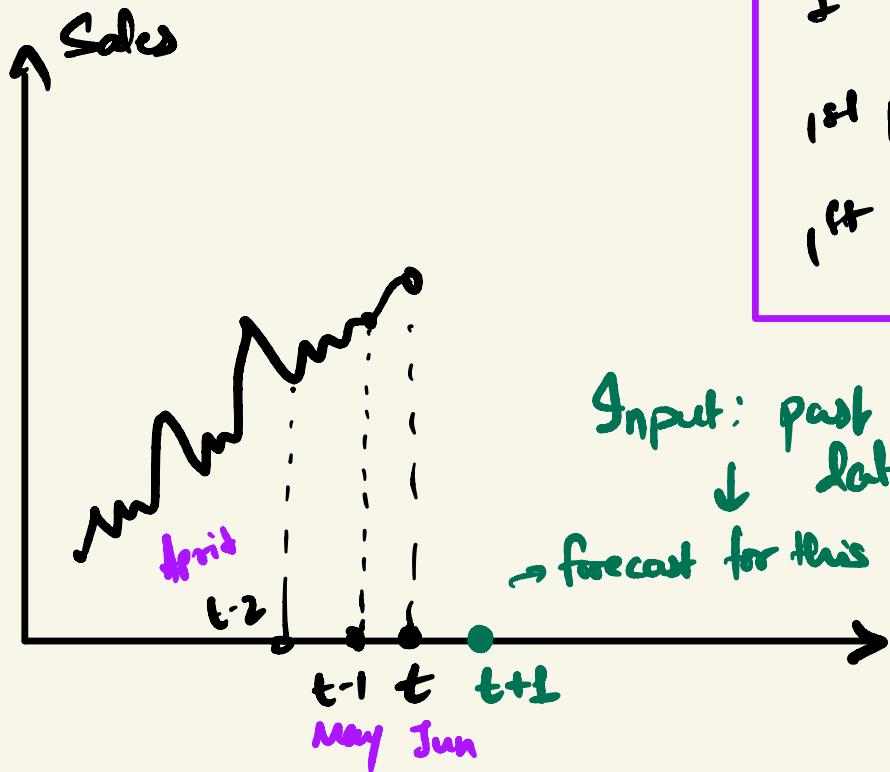
② Identifying seasonal patterns

→ plan promotional strategies

Basic challenges in TS

- ① Missing values
- ② Outliers/Anomalies

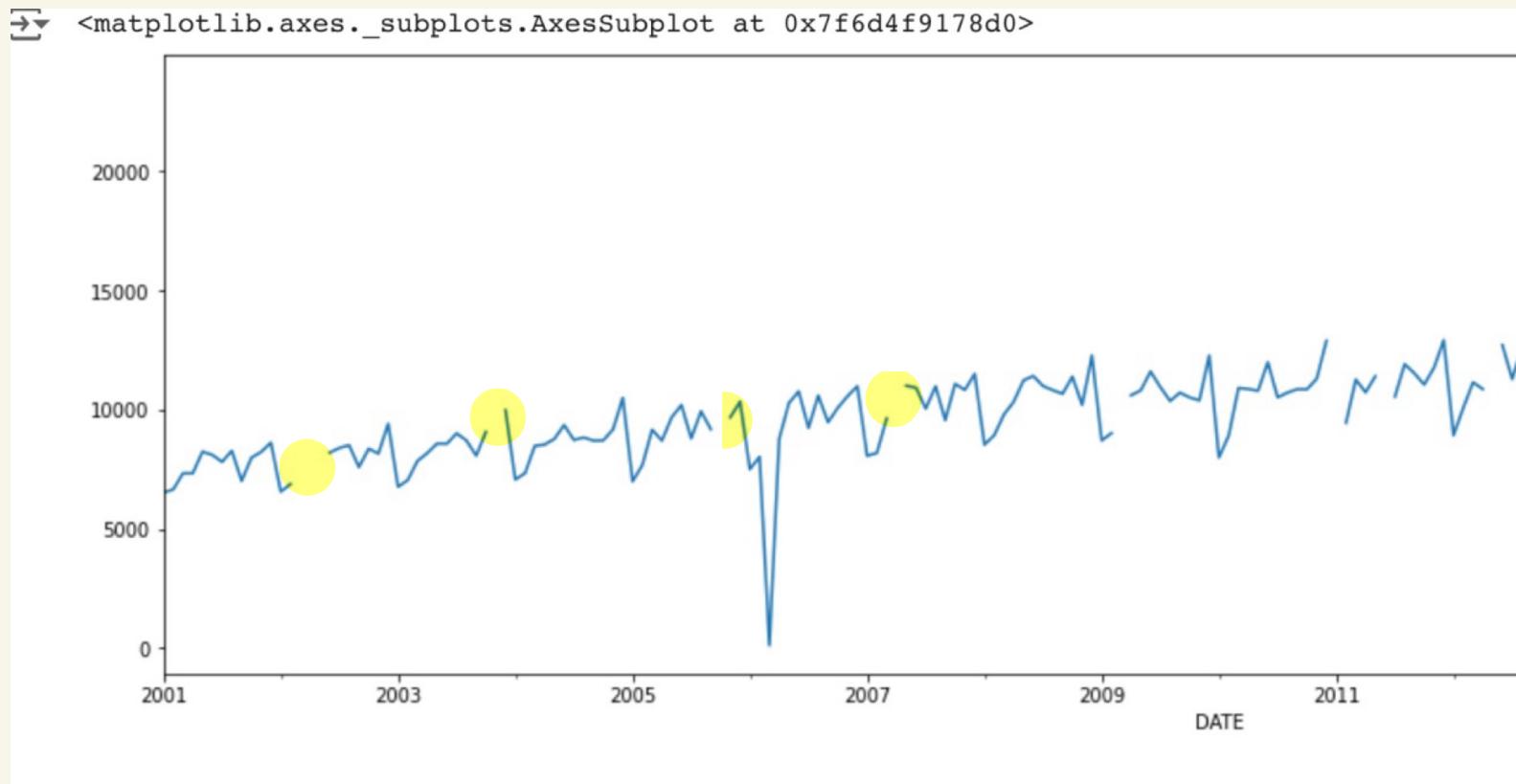
Forecasting



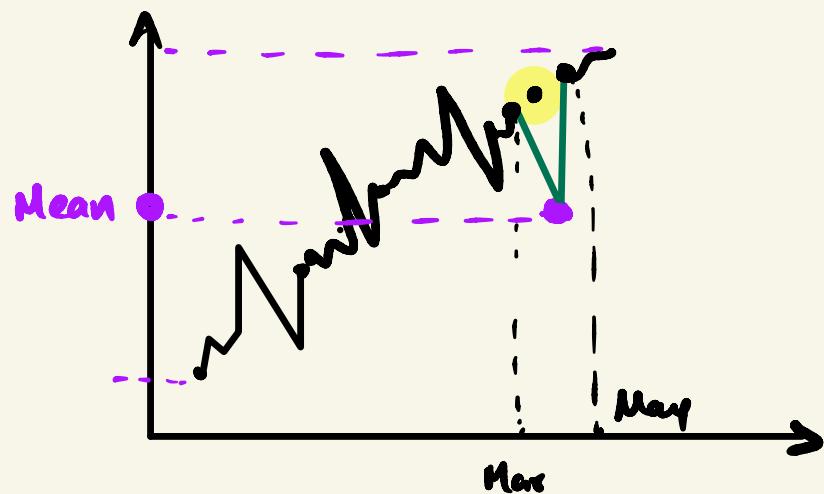
Timestamp	Sales
1 st Mar 2024	10,000
1 st Apr 2024	11,000
1 st May 2024	12,000
1 st June 2024	13,000

$$y_{t+1} = f(y_t, y_{t-1}, y_{t-2}, \dots)$$

Handling missing values

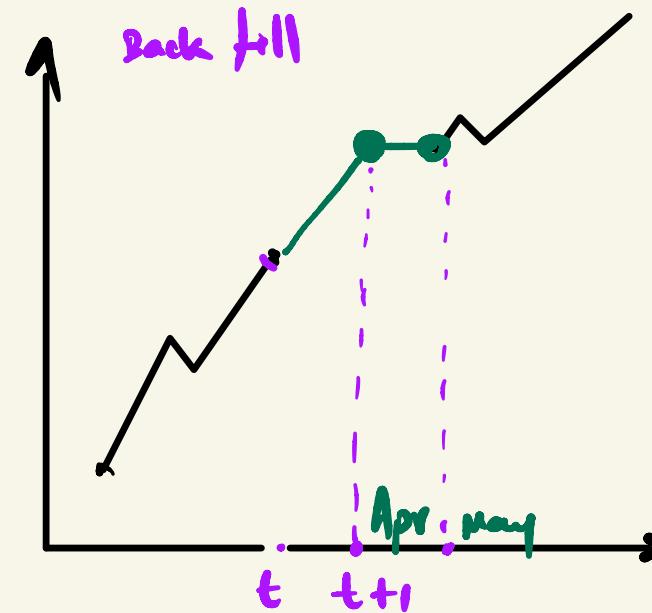
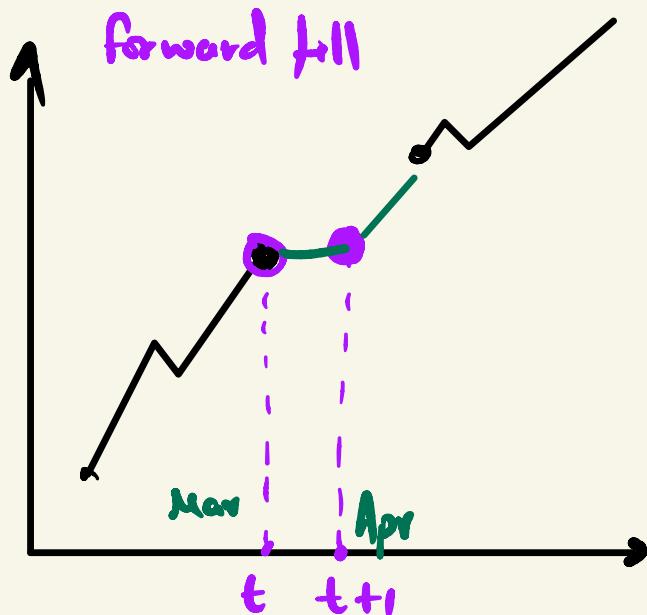


① Mean / Median imputation

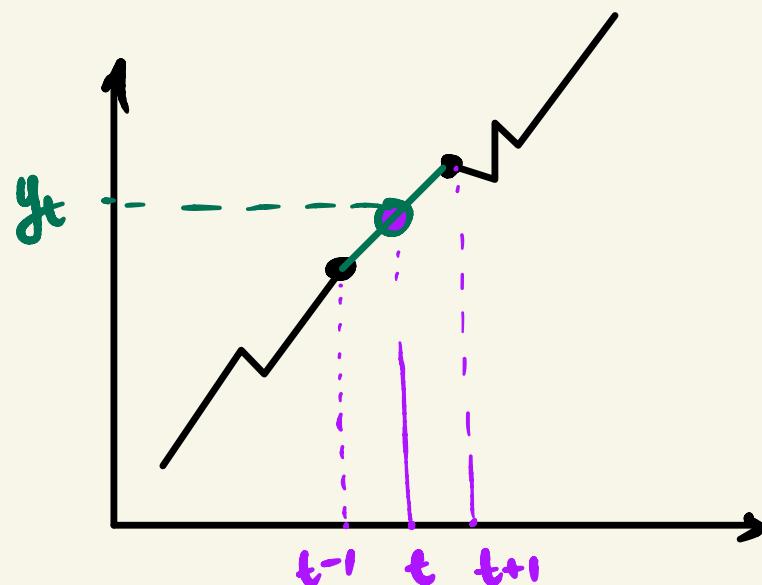


* Not the best strategy in case of TS data

② Backward fill or forward fill

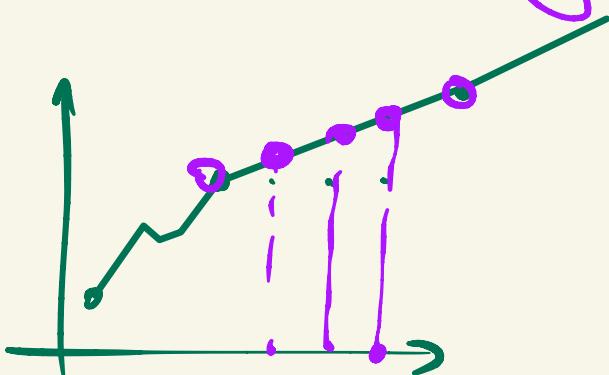


③ Linear interpolation (most widely used strategy)



$$y_t = \frac{y_{t+1} + y_{t-1}}{2}$$

* Another extension of this : Centered MA



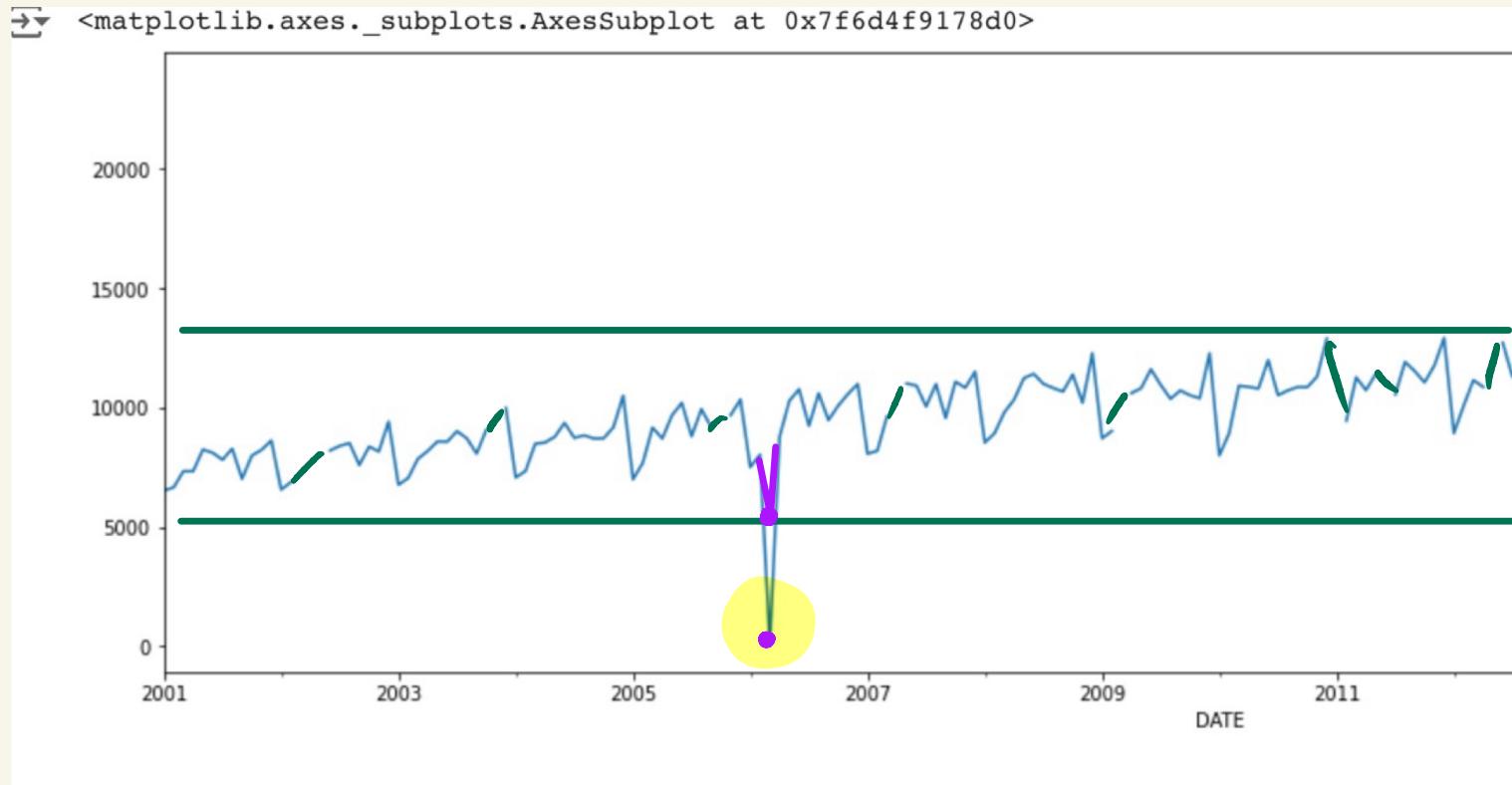
Q) Should we drop the missing value rows from TS data?

A) No.

TS: ordered, uniform frequency

Mon	800
Tue	900
W	950
Th	1000
Fri	1100
Sat	1200
Sun	1800
Mon	800
Tue	900
W	950
.	.
.	.
.	.

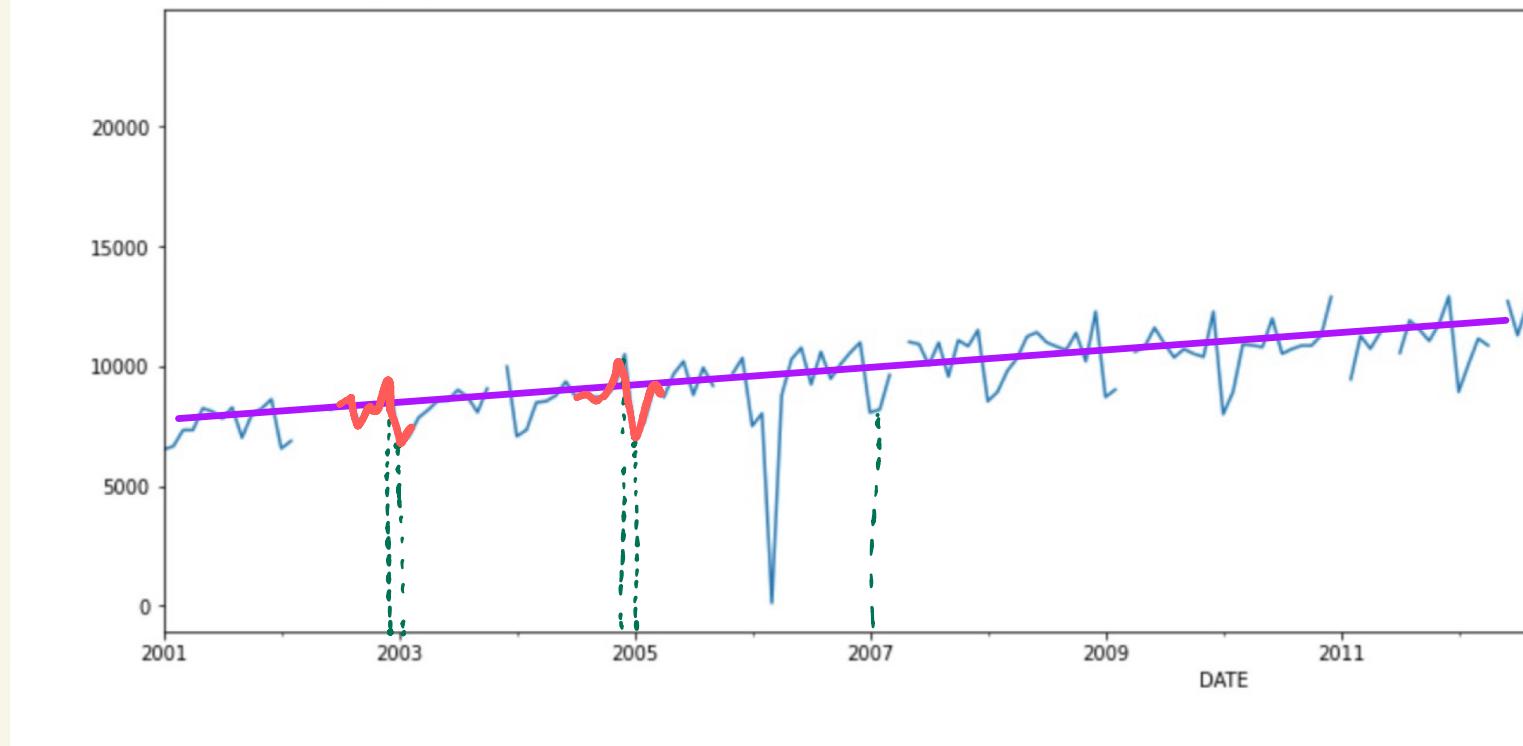
Anomalies



Clipping

Decomposition

→ <matplotlib.axes._subplots.AxesSubplot at 0x7f6d4f9178d0>

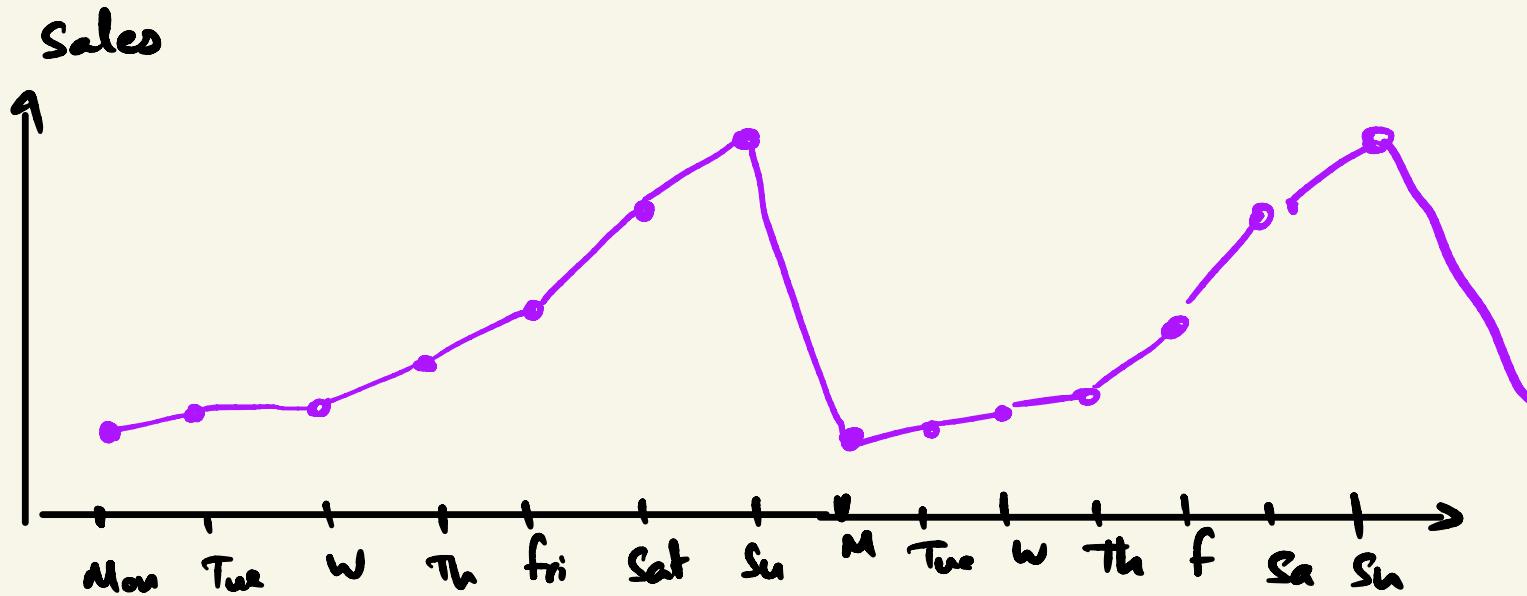


$$TS = \text{Trend} + \text{Seasonality} + \text{Residuals} \\ (\text{Noise})$$

~~~~~

ETS decomposition (Error, trend, seasonality)

### 5. Sneaker store in a mall



Seasonality = weekly seasonality

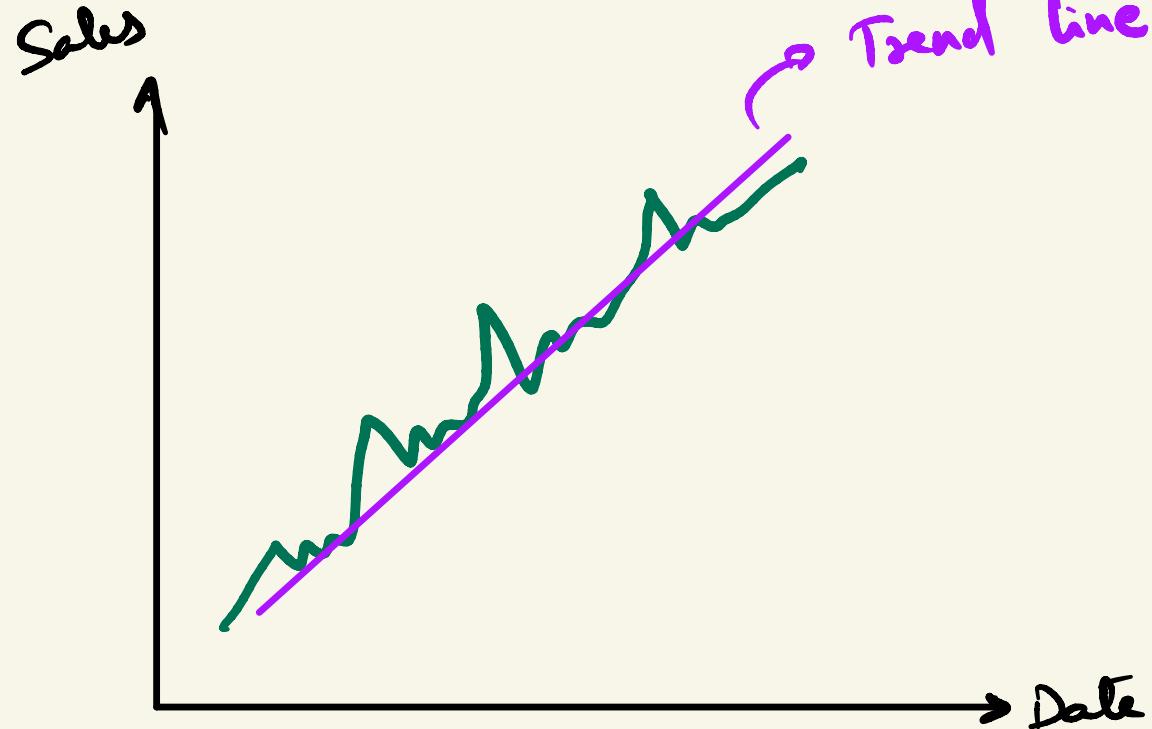
Seasonality is domain specific

Q) Can we have multiple seasonalities?

A) Yes. We can have multiple seasonalities.

In our <sup>case</sup>, we only deal with single seasonality

## Trend



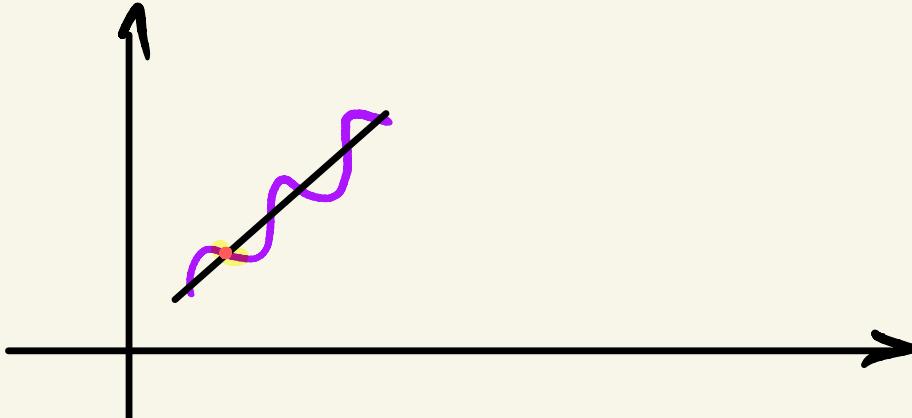
Trend line → multiple ways to obtain

- ① linear seg.
- ② line joining first and last point
- ③ Moving averages

Moving Averages → rolling average

window = 3

| index | values |
|-------|--------|
| 0     | 10     |
| 1     | 15     |
| 2     | 12     |
| 3     | 15     |
| 4     | 30     |
| 5     | 12     |
| 6     | 13     |
| 7     | 17     |
| 8     | 20     |



0 1 2 3 4 5 6  
10 15 12 15 30 12 13  
NA NA 12.33 14 19

$$10 + 15 + 12 = \frac{37}{3} =$$

Moving average → smoothing effect on the data

## Doubts

