Topics Covered Today:

1. Module Layout
2. What is Timeseries Data
3. Problem Statement for TimeSeries
4. Forecasting
5. Properties of time series data
6. Handling Missing Values & Anomalies
7. Linear Interpolation
8. Moving Averages
9. Time series Decompsitions

Lecture Notes are added in the scaler dashboard along with python notebook and Dataset.

You will also receive a slightly more handwritten + text notes at the end of the time series part of the

Timeseries

Start at 9:05 pm

→ 6 classes → Time Series
4 Classes → Rec. System
2 Classes → Interview | Case Study

→ Data against Time

① Stocks ② Sales over time
③ Users over time

Problem statement

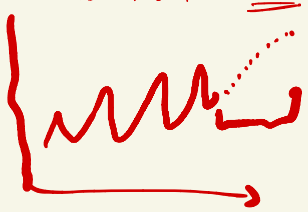
Imagine you are a Data Scientist at MobiPlus, a mobile manufacturing company

You need to forecast their future sales for better planning and revenue.

- **Agenda 1:** We want to understand the patterns in demand to be able to be better plan for factory maintenance / staffing requirements.
- **Agenda 2:** We need a certain level of accuracy. The management requires that the Mean Absolute Percentage Error (MAPE) is not more than 5%.
- **Agenda 3:** Need a range forecast to supplement the point forecast to make educated trade-off wherever needed.

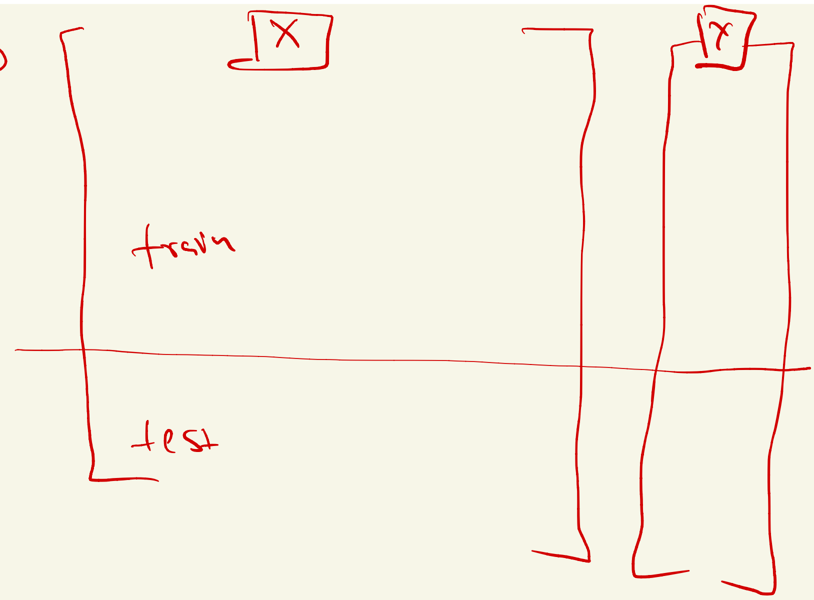
Over the next few lectures, we will be completing these tasks.

① 90 - 110 $M sales
② → 100M $ sales

⇒ 

Agenda 1

⇒ Agenda - 2
↓
measure the accuracy of Prediction.

⇒ 

⇒ How linear Regression & Classification used to look

⇒ Univariate Time series data →

Date	Val1

⇒ only 1 data Column

⇒ Multivariate TS →

Date	Sales	Profit

⇒ Multi Data Column

⇒ **Forecasting** → **given** → $y_1, y_2, \ldots, y_t$

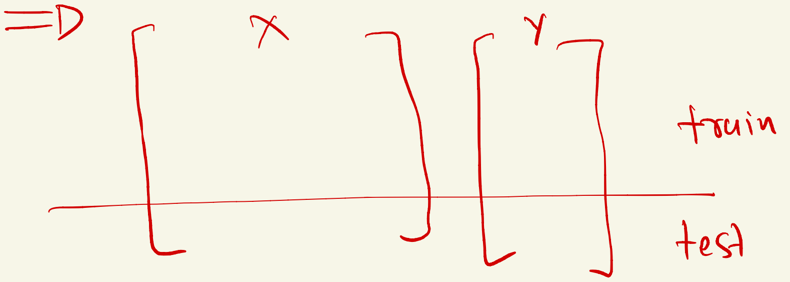
predict → $y_{t+1}, y_{t+2}, \ldots, y_{t+m}$

⇒ You never predict past **values**

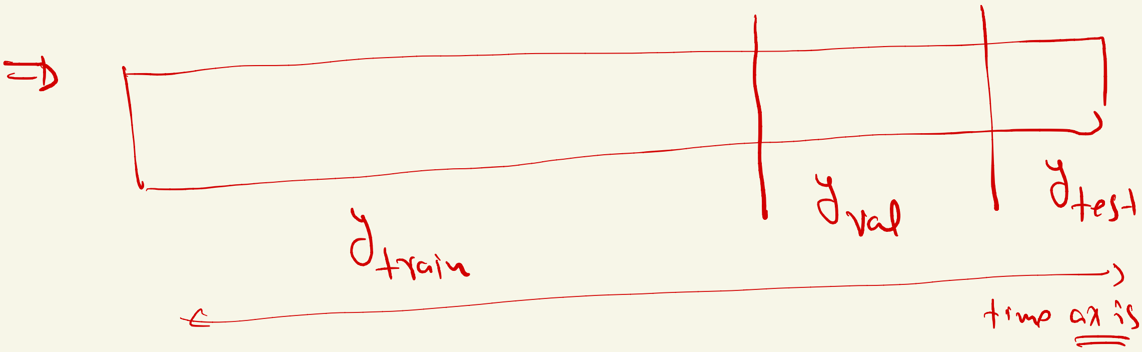
y_{t+1}, y_{t+k}

x x x
x x

predict $y_t \ldots y_{t+r}$



Regression



$\boxed{\text{Timeseries}}$

2001 - 2008 → train
2009 → Val.
2010 → Test data

⇒ EDA

⇒ outliers ⇒ Missing data

① Q Can a Time series data have missing value/ Can you remove some values from TS Data ??

⇒

Date	Sales
1 Jan	100
2 Jan	120
3 Jan	140
7 Jan	

⇒ TS data → order matters

↳ Data should always be in increasing time format.

Date	Sales
2020-01-01	
2020-02-01	
2019-06-01	
2020-03-01	

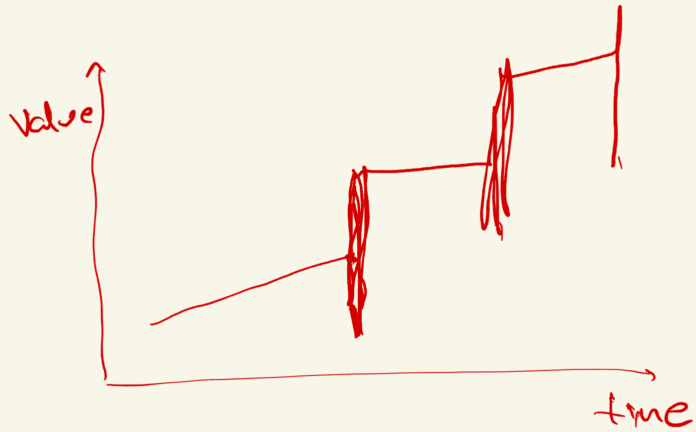
⇒ XX a TS data

↳ TS data should not have duplicate values of time.

⇒ Cannot have this (Duplicate Values)

Date	Sales
2020-01-01	100
2020-01-01	120

⇒ • Row wise operation



⇒ 3 important checks for TS Data

→ Order matters (always be increasing)

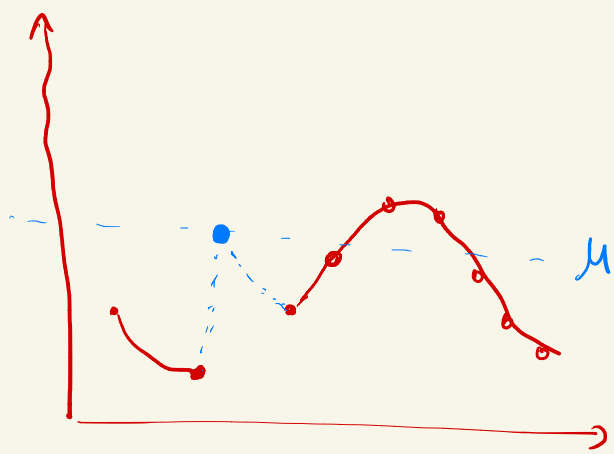
→ No repetition/Duplicates

→ Any missing values.

⇒ Missing Values →

↳ mean of Column ↳ meadian ↳ Zero

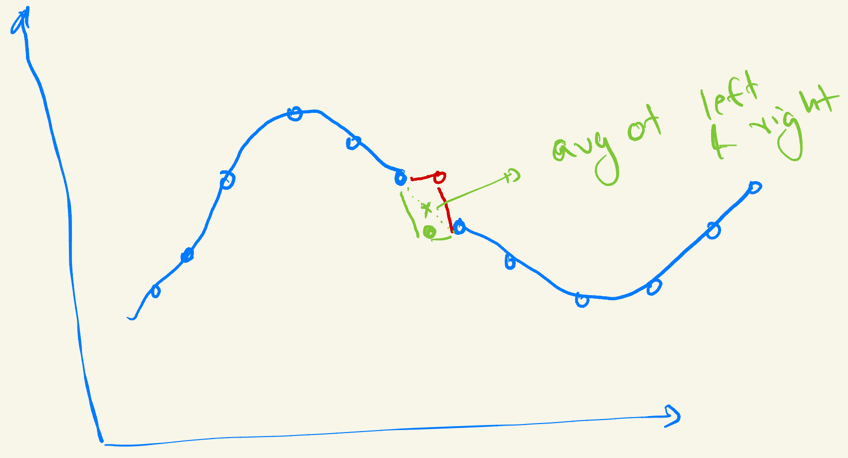
Q ↳ Mean be used in TS ??



If you use Column mean the pattern will be lost

XX

⇒

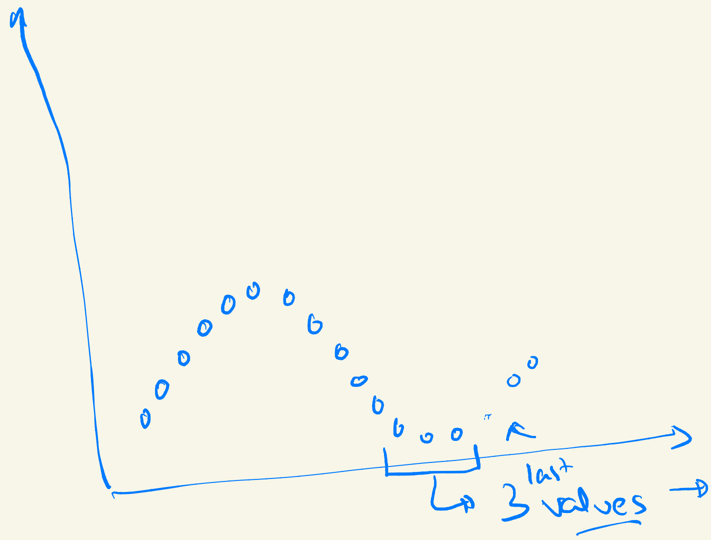


⇒ Moving Average →

⇒ Mean → avg of entire data

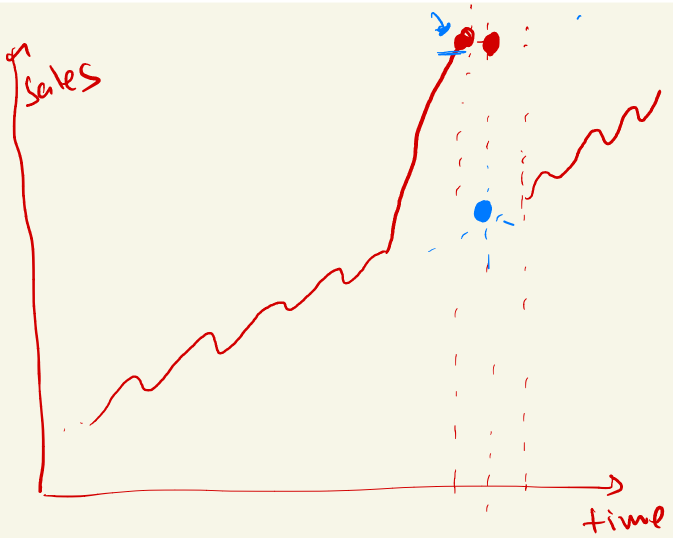
⇒ mean of last 3 values

⇒



mean of all the data ⇒ XXX

$$\hat{y}_t = \frac{y_{t-1} + y_{t-2} + \dots y_{t-m}}{m}$$



→ Linear Inter

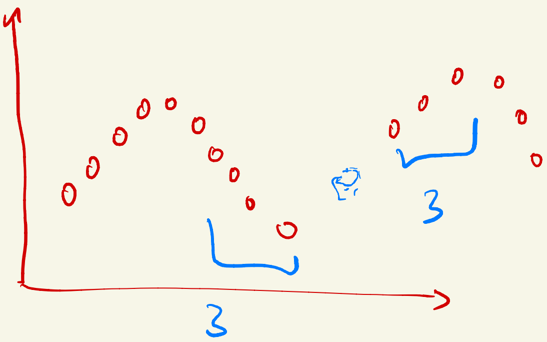
→ last value probably will not

→ avg of last 7 values

⇒

⇒ Central MA →

windows on both side of missing data



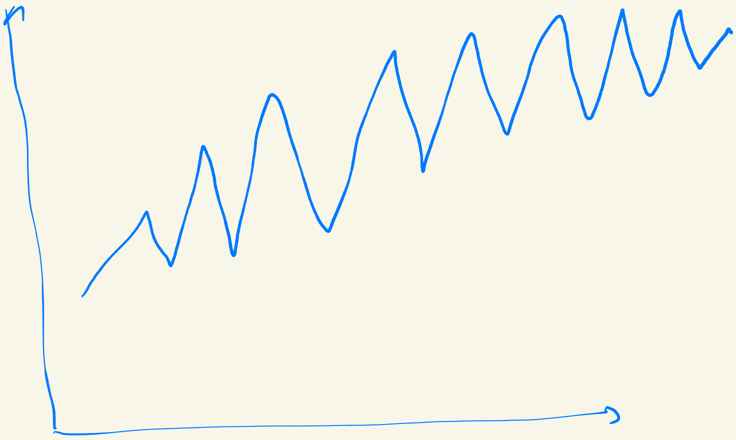
⇒ Avg of (last 3 + next 3)

⇒

$$\text{CMA at any point} = \frac{\overbrace{y_{t-m} \quad y_{t-m+1} \cdots y_t}\ + \cdots \overbrace{y_{t+1} + \cdots y_{t+m}}}{2m+1}$$

⇒ CMA and linear Interpolation → $\frac{y_{pre} + y_{next}}{2}$

linear Interp → a type of CMA (m=1)

⇒ 

→ whenever rolling avg
↳ smoothen the chart

⇒ 1, 2, 3, ... 100

20^{th} position → $\frac{16 + \dots 20}{5}$

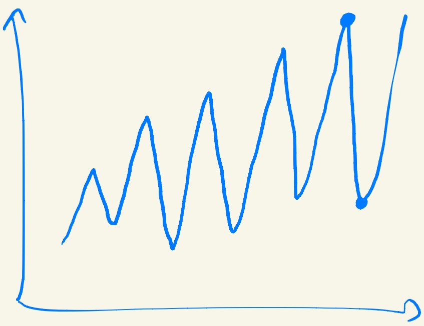
$\frac{11 + \dots \dots +20}{10}$

$\frac{1 + \dots \dots +20}{20}$

21^{th} $\frac{17 + \dots 21}{5}$

$\frac{12 + \dots 21}{10}$

$\frac{2 + \dots +21}{20}$

⇒ 

⇒ 5, 15, 30, 20, 18, 17, 30

⇒ 5^{th} position → $\frac{18+20}{2}$ 2 day

$\frac{30+20+18}{3}$ 3 day

$\frac{5+15+30+20+18}{5}$ 5 day

numerator →

Denominator

2

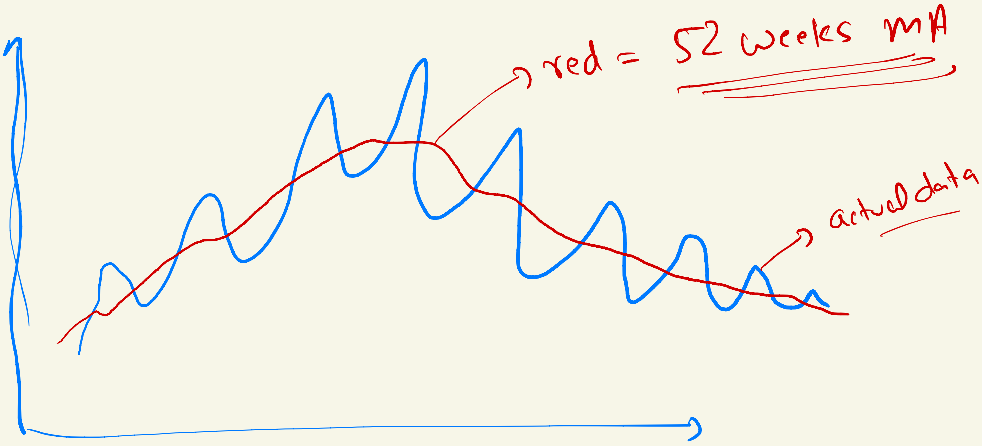
3

5

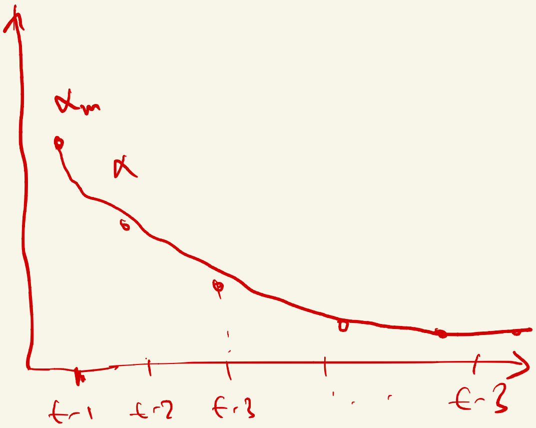
fraction ↓

6^{th} position → $\frac{17+18}{2}$

⇒ How is the sales on a given day impacted by
- ↳ 1 day before ←
- ↳ 7 day before ←
- ↳ 1 year before ←

⇒ 

$$\Rightarrow \frac{y_{t-1} + y_{t-2} + \dots \; y_{t-m}}{m}$$



⇒ you can also give weightage based on the difference in date.

$$= \frac{\alpha_{t-1} y_{t-1} + \alpha_{t-2} \cdot y_{t-2} + \dots \; \alpha_{t-m} \, y_{t-m}}{m}$$

Weighted MA ⇒

⇒ Components of TS →

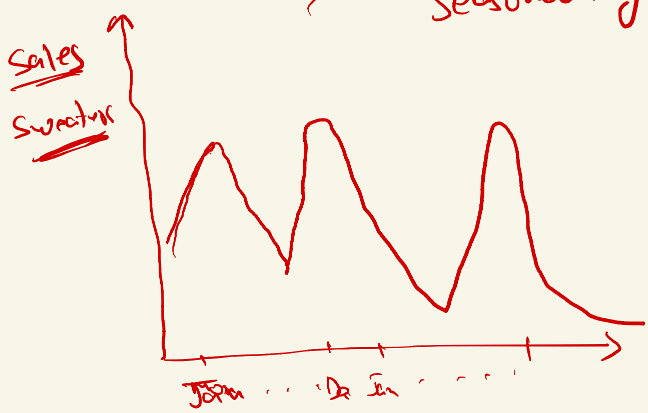
→ Pattern (increasing, decreasing, changing) → $\boxed{\text{Trend}}$

→ Repeatition ( ) → $\boxed{\text{Seasonality}}$

⇒ Seasonality → A Repetetive behaviours

→ weekly seasonality

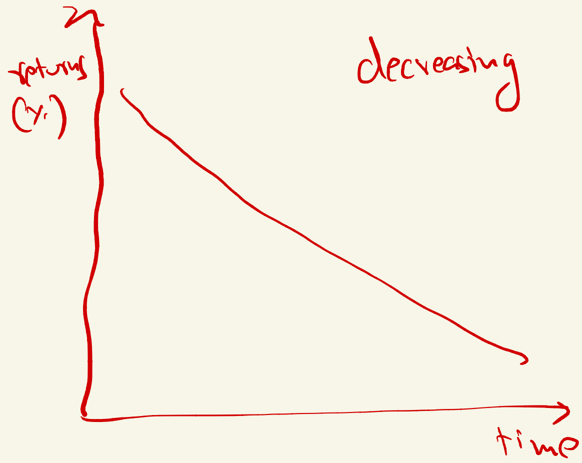
→ yearly seasonality



⇒ Trend

Increasing trends

decreasing



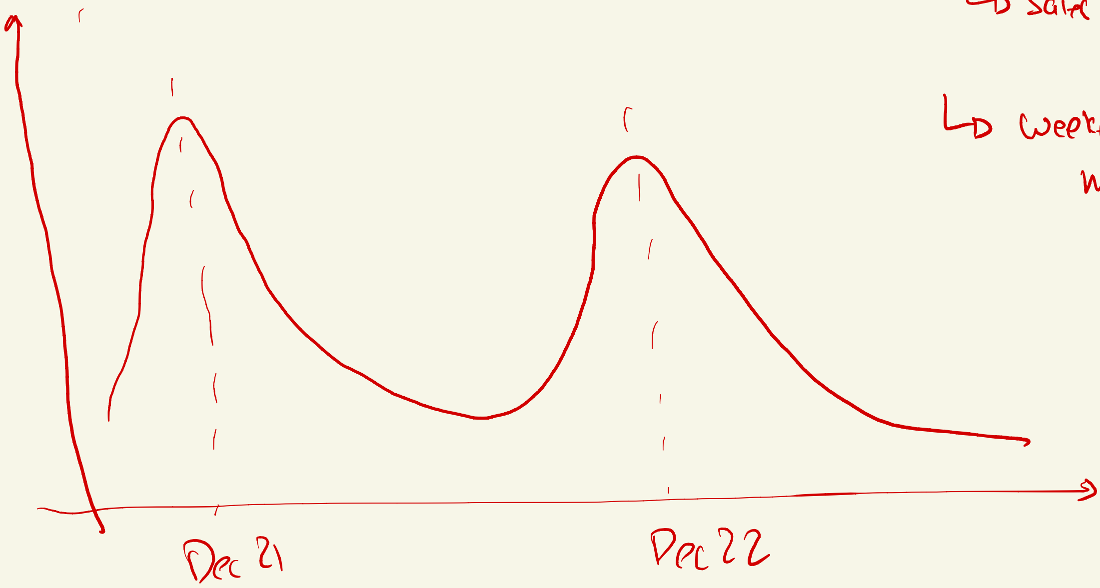
Changing trends

⇒ multiple Seasonality →

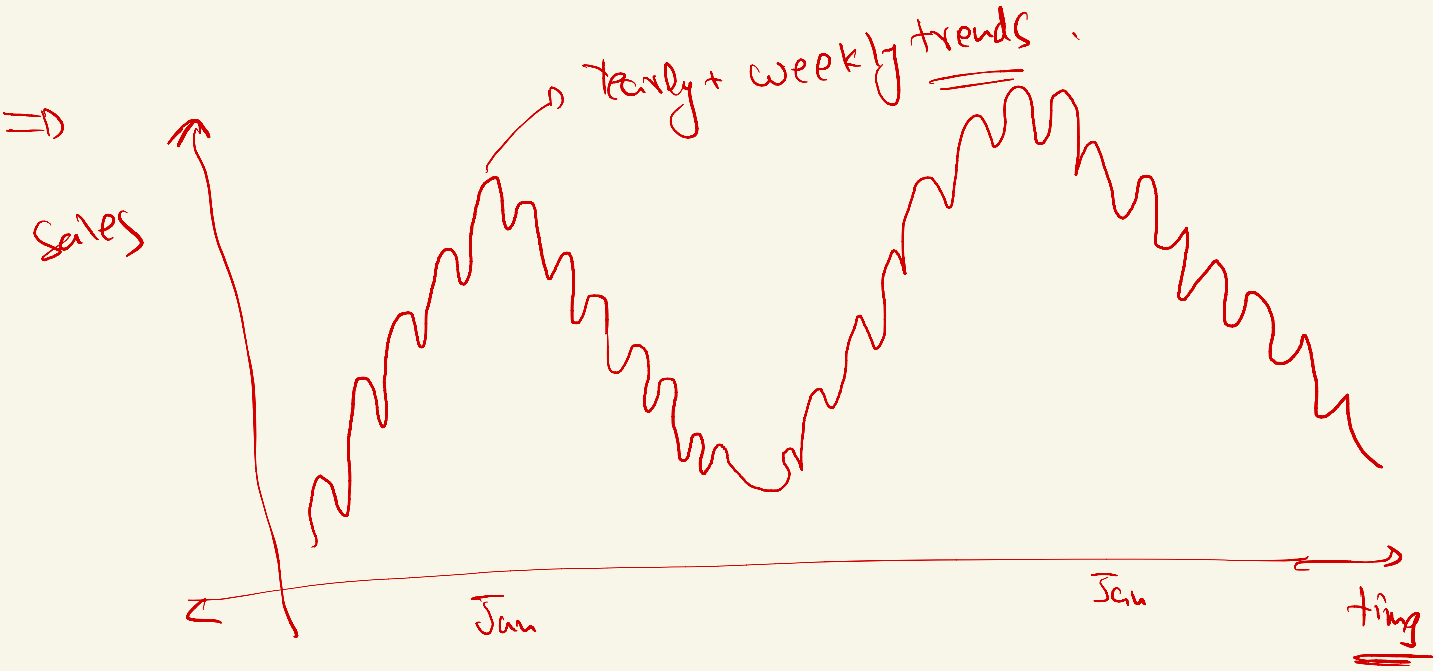
Sweater Sales

↳ Sales will increase in Dec–Jan

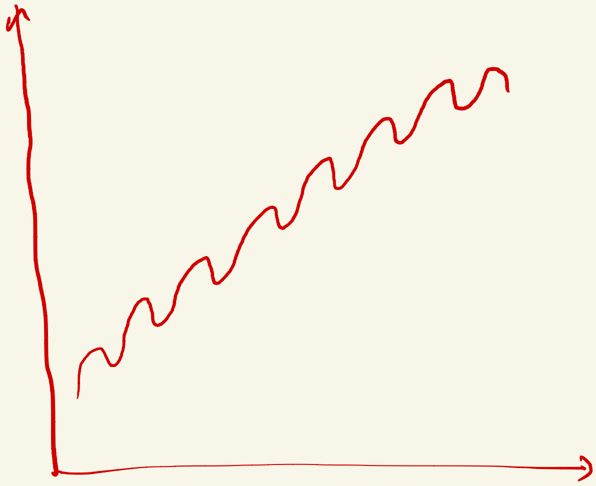
↳ weeken will have more Sales



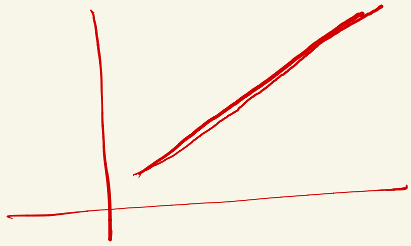
⇒



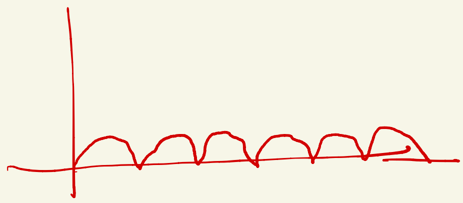
=> Seasonality + Trend



Trend



Seasonality



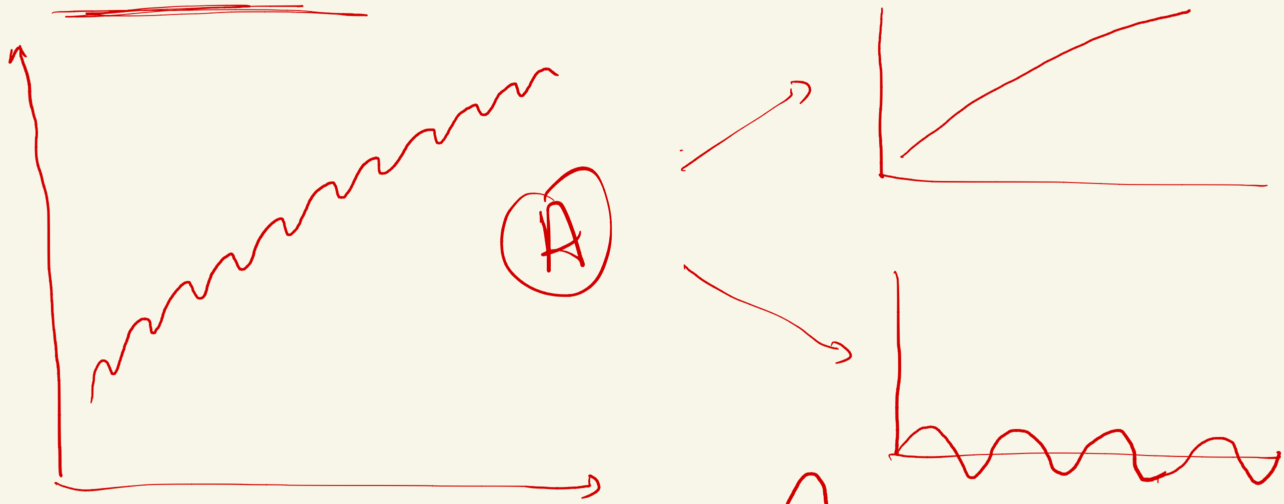
$$\boxed{y(t) = b(t) + s(t) + e(t)}$$

Trend Season residue / noise

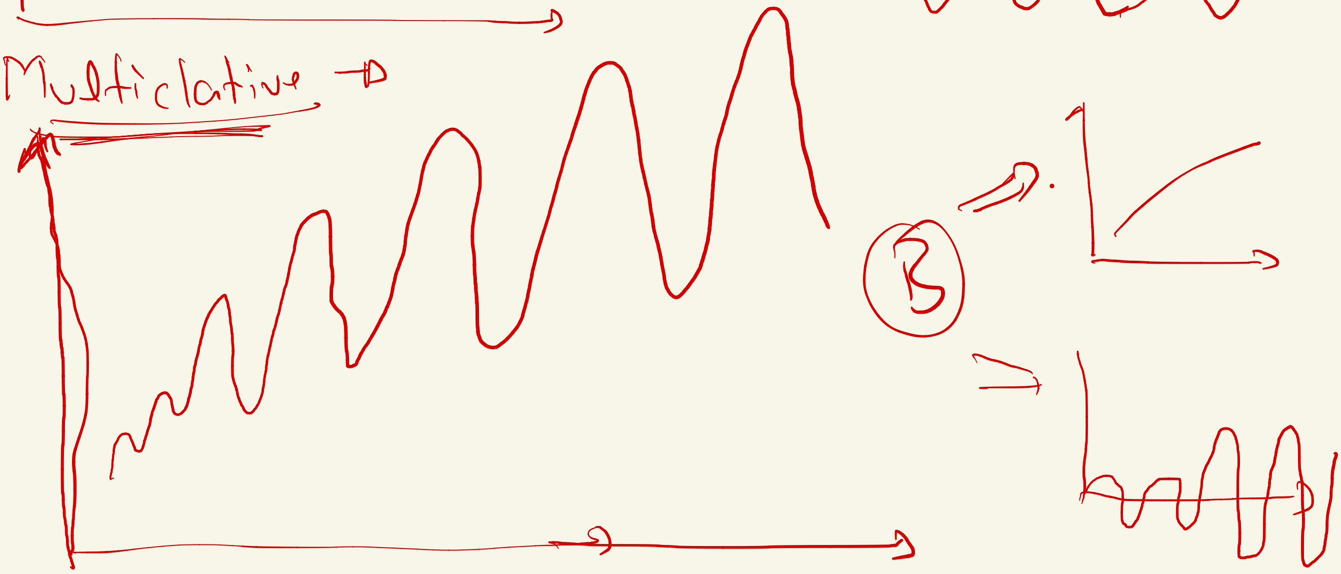
$$e(t) = y(t) - \underbrace{\left[b(t) + s(t) \right]}_{\hat{y}(t)}$$

$\hat{y}(t) \rightarrow$ Prediction

⇒ Additive Trend



⇒ Multiclative ⇒



⇒ A

→ overall increasing

→ Seasonality is there

$$y(t) = b(t) + s(t) + e(t)$$

B

→ over all increasing

→ Seasonality is also there

$$y(t) = b(t) \cdot s(t) \cdot e(t)$$

=> Jan 21

Monday 100
wed 150
Sund 200

50
50

Dec 21

Monday 150
wed 200
Sun 250

50
50

Jan 21

50
50

Dec21

Monday 150
wed 220
Sun 300

70
80