19 March 2024 19:56

Agenda Time Series Analysis

https://chat.whatsapp.com/JvnoQ9yFj7NBHNtNzGKUzx

https://chat.whatsapp.com/CNg5iaeoSxhIraeEUK4UQY

* (Rohit Jindal)

7982691190

→ # Doubts **

Logistics → (TTS → Evening 9 pm)

After every class → 3 things

1. Class Notes ✓
2. 5 minute Summary →
3. Interview Qs and Answers

Agenda →
- A1 Introduction to Time Series
- A2 Moving Averages
- A3 Decomposition

Rohit Jindal → Sr Data Scientist @ Target

→ Mastercard AI Garage

→ SSGA (financial ML)

→ AM at Ircon

Originally I am Civil Engineering

→ AM at Ircon

Originally I am Civil Engineering

Educ^n → Civil → DCE (DTU)

DS → Master in D.S (IISc Bangalore)

University of Cambridge → Master in IP & Substainability

(1 year)

General ⇒ Sharing my numbers → WA groups

Updates

Admin Only

https://chat.whatsapp.com/JvnoQ9yFj7NBHNtNzGKUzx

Discussions

https://chat.whatsapp.com/JvnoQ9yFj7NBHNtNzGKUzx

Mobiplus ← DS

forecast their future sales

March → April → May

100 units

(Better Forecast)

** 200

** *(Under) < 200

150

Model says

Incur loss

(over) > 200

Model → 300 mobile phones

Real → 180 - 200 phones

Stock those 120 phones

(120 extra phones

Model says

$

Incur ~~Loss~~

50 mobile phones

Stock this ...

⇒ Inventory (120 extra pho...

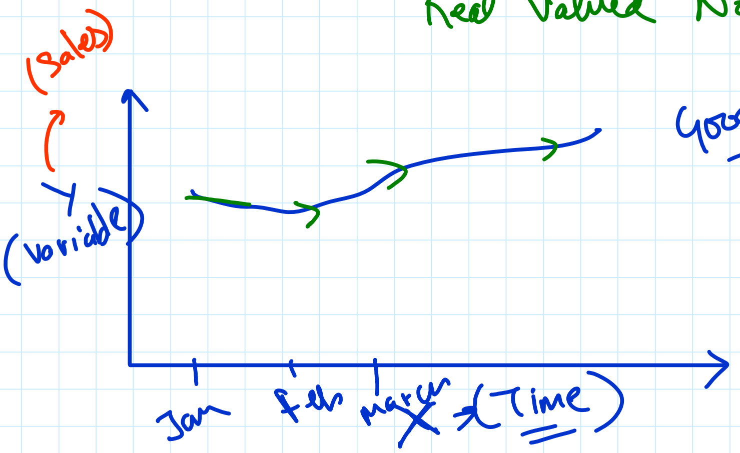
Time Series

Date	Sales
1st Jan	60 $
1st Feb	70 $
1st March	80 $
⋮	
1st Dec	
1st Jan	

Time Series Data → "Signal" + "Ordered Time Stamp"

x random

Real Valued No

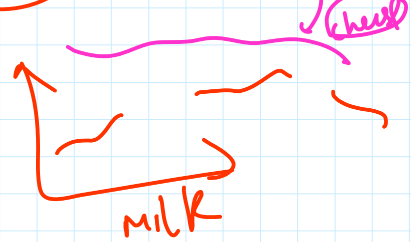


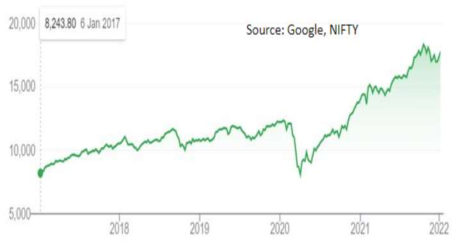
Google

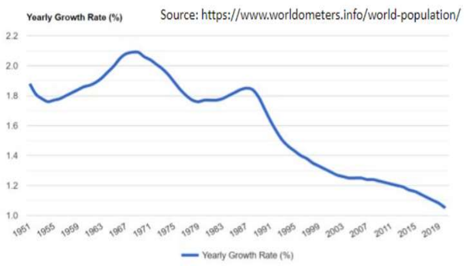
Industry Terminology

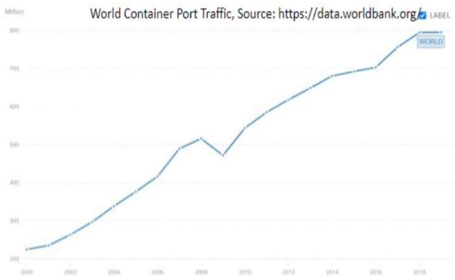
Sales → Data Coverage is low "

✓ check

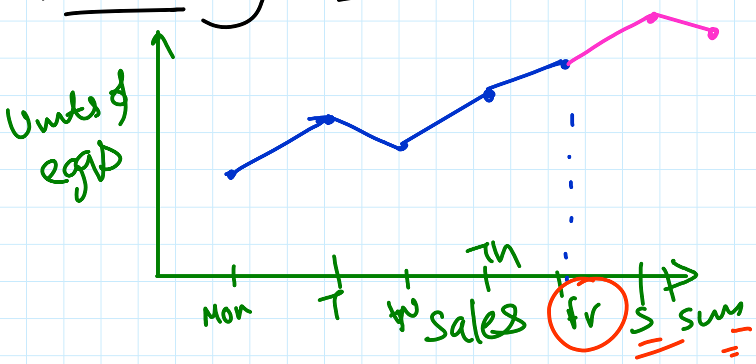








Forecasting → Tell us what the future holds, so we may know that you are gods.

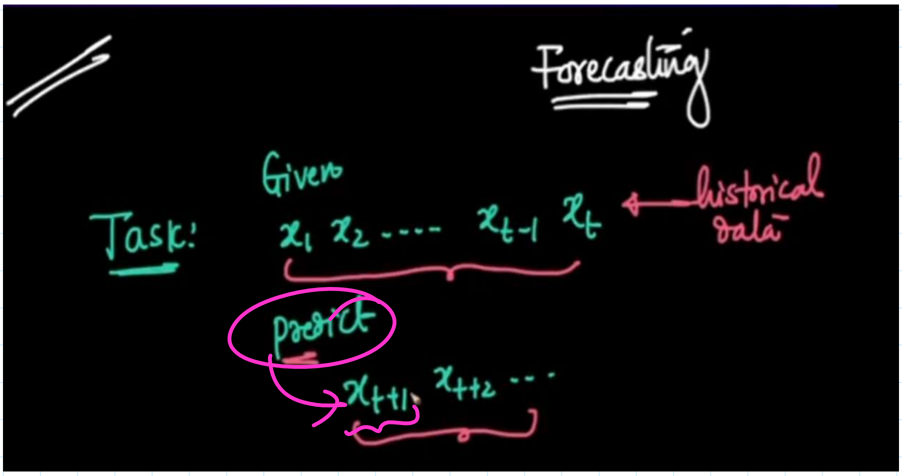


Close to Regression → ∈ R

→ Sales
→ Salary
→ House Price

→ weather
→ sales
→

Temp → High
→ Med
→ Low



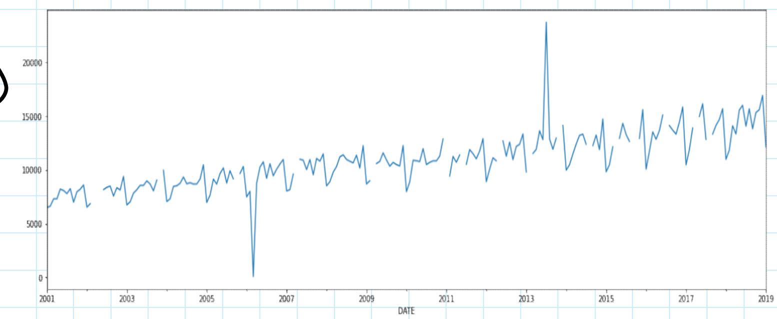
		sales	
2024	Jan	10 unit	x_1
	Feb	20 units	x_2
Here	March	22 units	x_3
Predict	April	30 unit	$\hat{x}_4$
	May	45 unit	

31st may

T/False

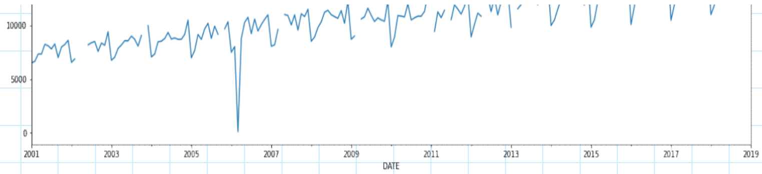
True/False


```
# df['sales'].plot()
```

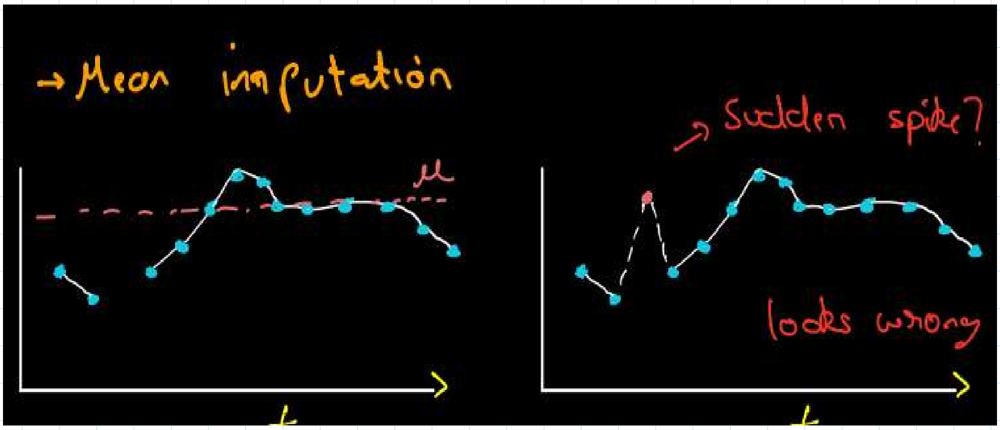


Problem
- → Missing Values
- → Anomaly

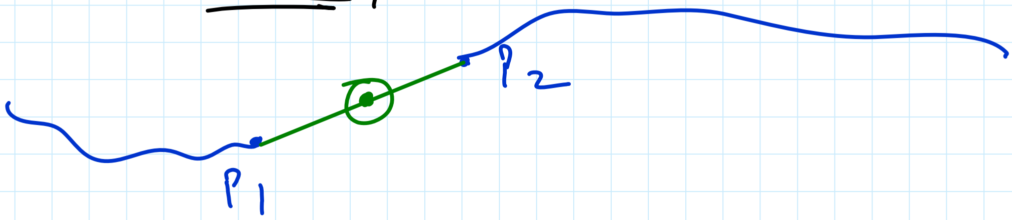
Problems → Missing Values
→ Anomaly

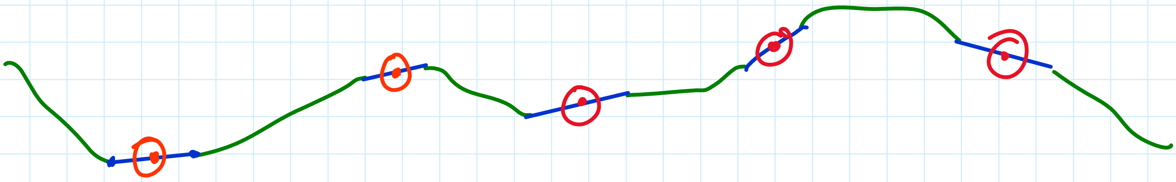


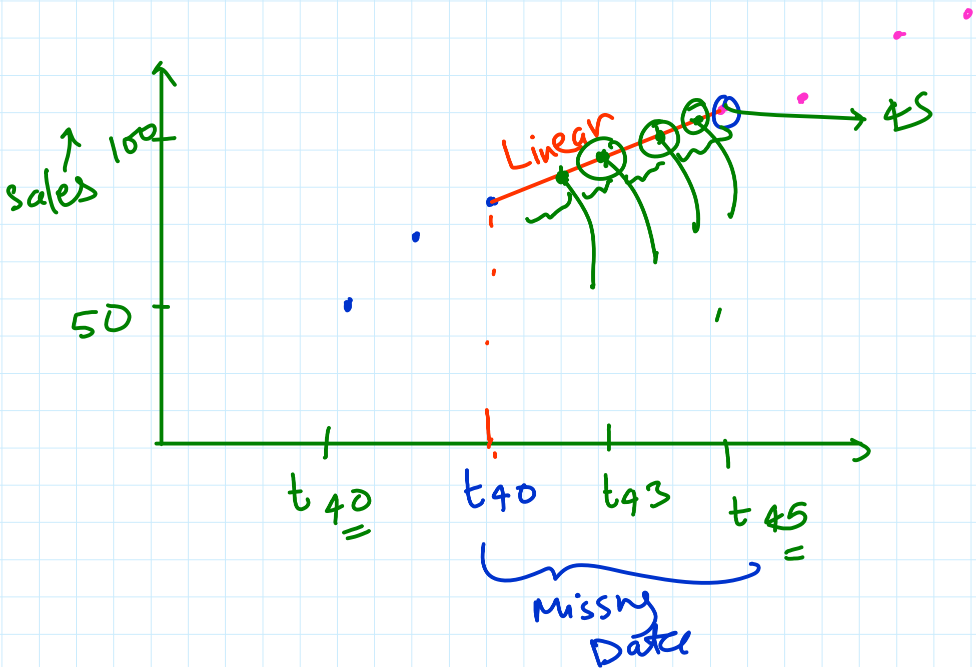
Mean Imputations



III Linear Interpolation

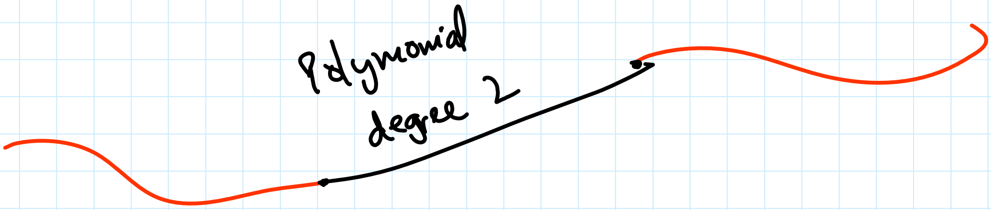




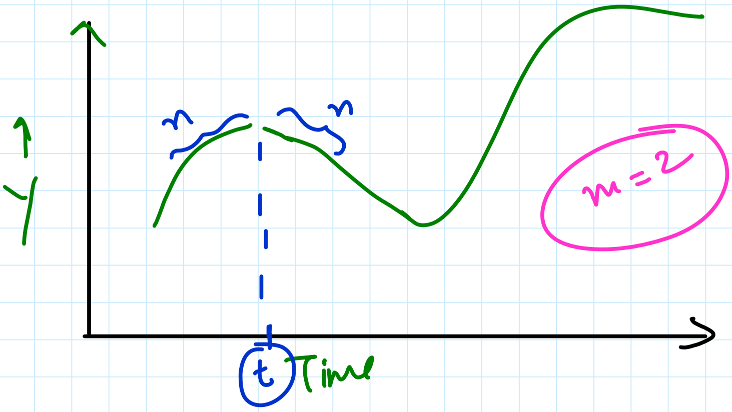


40
41
42 } missing
43
44
45

Missing Data



Extend this Idea

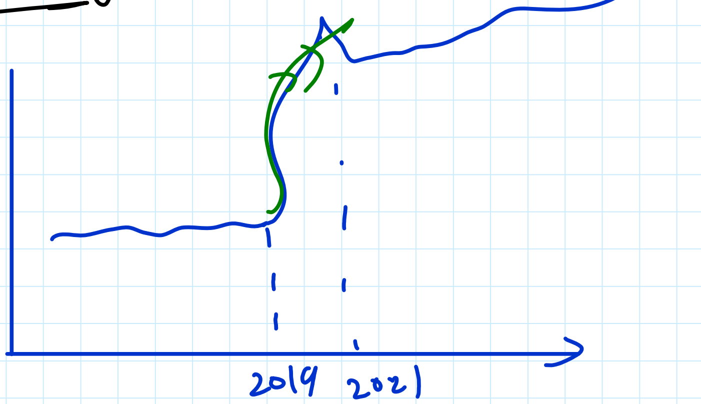


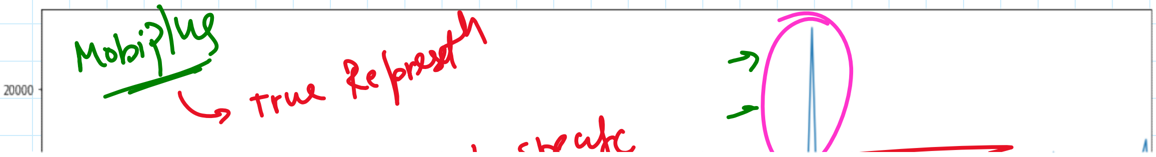
Sales
20
25
28
missing
30
31
33
35

$$x_t = \frac{x_{t-1} + x_{t+1}}{2}$$

Centered Moving Average

II Anomaly → abnormal / unusual

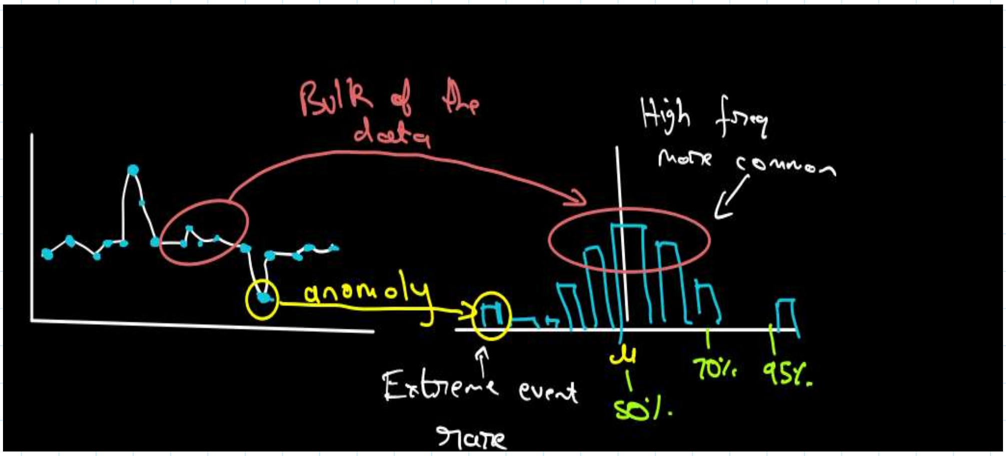




2023

1. NaN

2. Clip



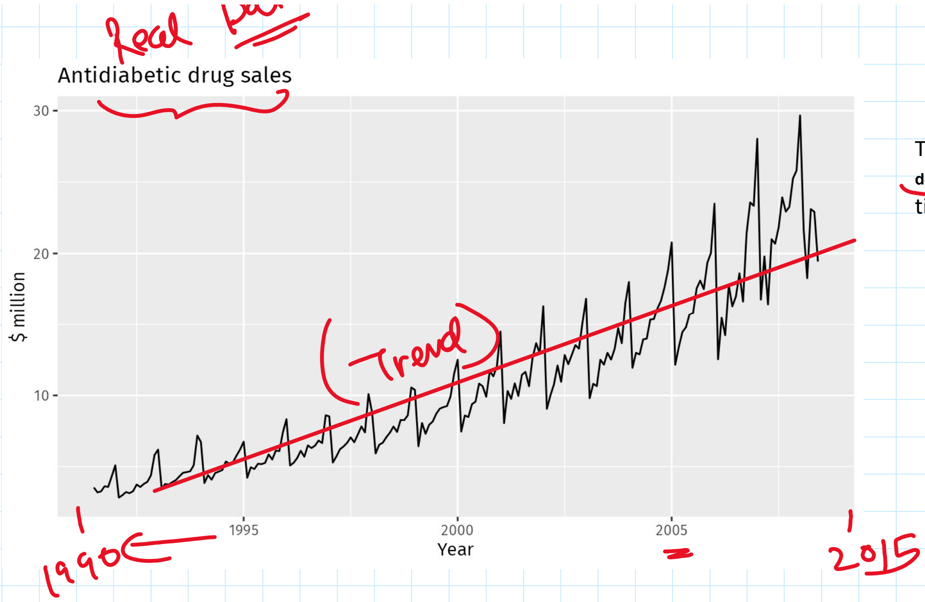
Clip → 10 – 30 → lower
upper

Sales	After clip
3	→ 10
11	→ 11
20	→ 20
25	→ 25
33	→ 30
39	→ 30

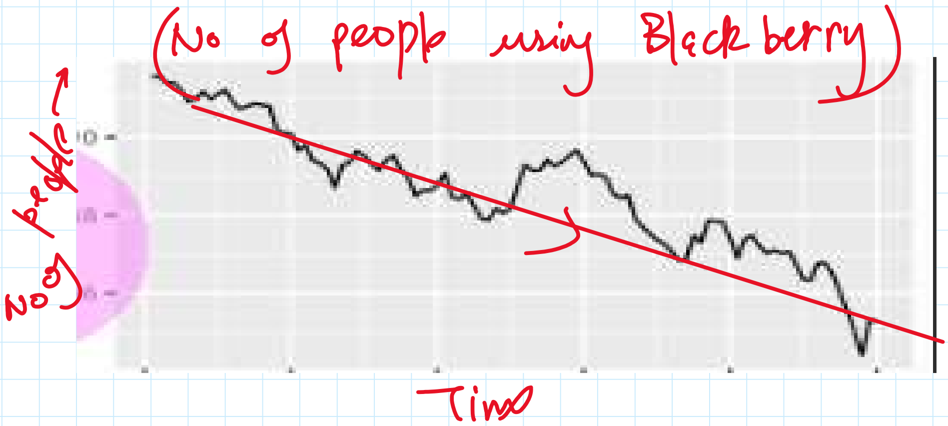
22:25 →

Real Data

Antidiabetic drug sales

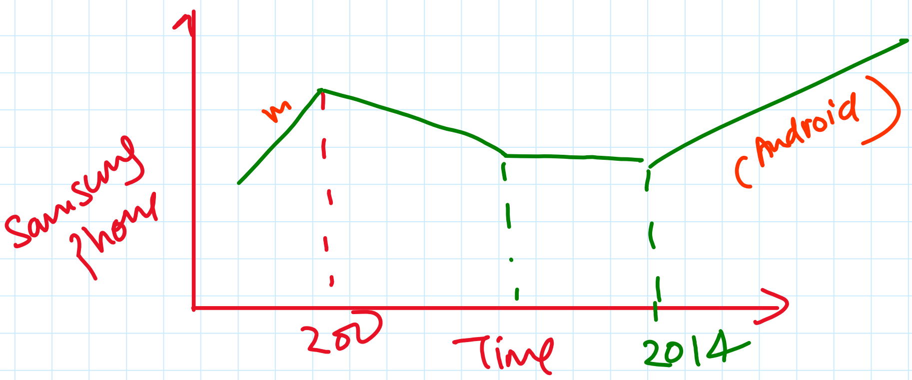


Trend can be thought of as the linear **increasing or decreasing behavior** of the series over a long period of time.



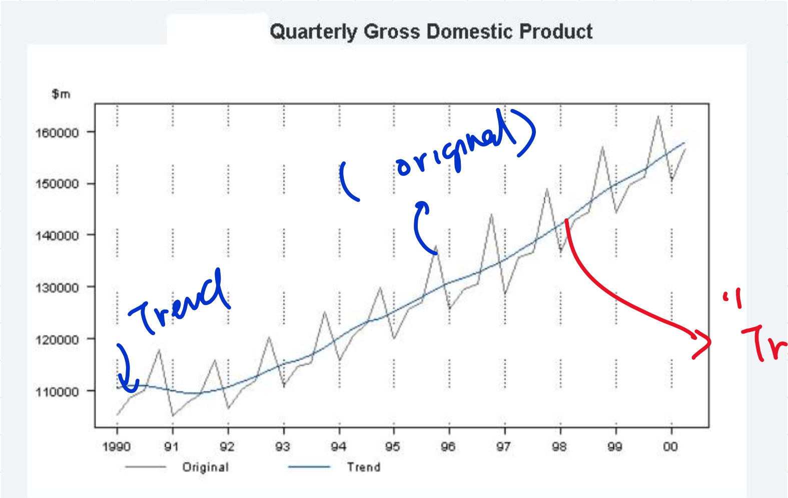
Trend →

Trend can be thought of as the linear **increasing or decreasing behavior** of the series over a long period of time.



(Time < 2000 → ↑ Trend)

Time → < 2014 → change



qualitatively

→ mostly linear

→ smooth Predictable fn

"Trend Line" ⇒ ($y = mx + c$) (type)

→ this (mostly)

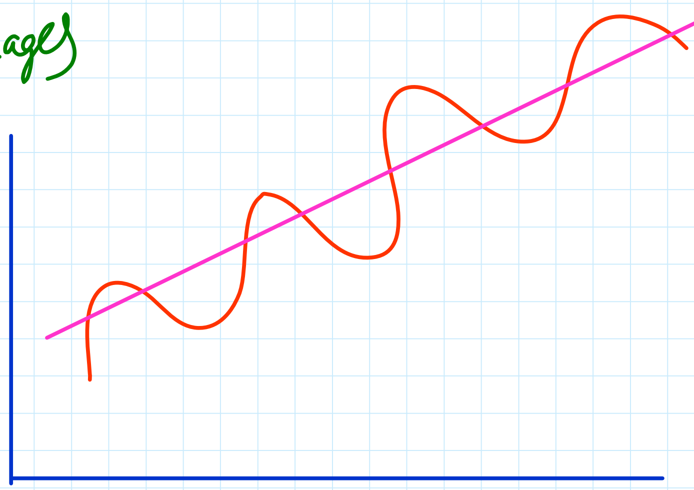
* Polynomial ($y = mx^2 + cx + d$)

How To Calculate Trend?

Rolling Averages

Sales

pd.rolling



↓ trend line

(Trend)

Sales | Rolling Average (window size = 2)

20 →

25 →

30 → (25+30)/2 =

32 → (32+30)/2

34 → (34+32)/2

38

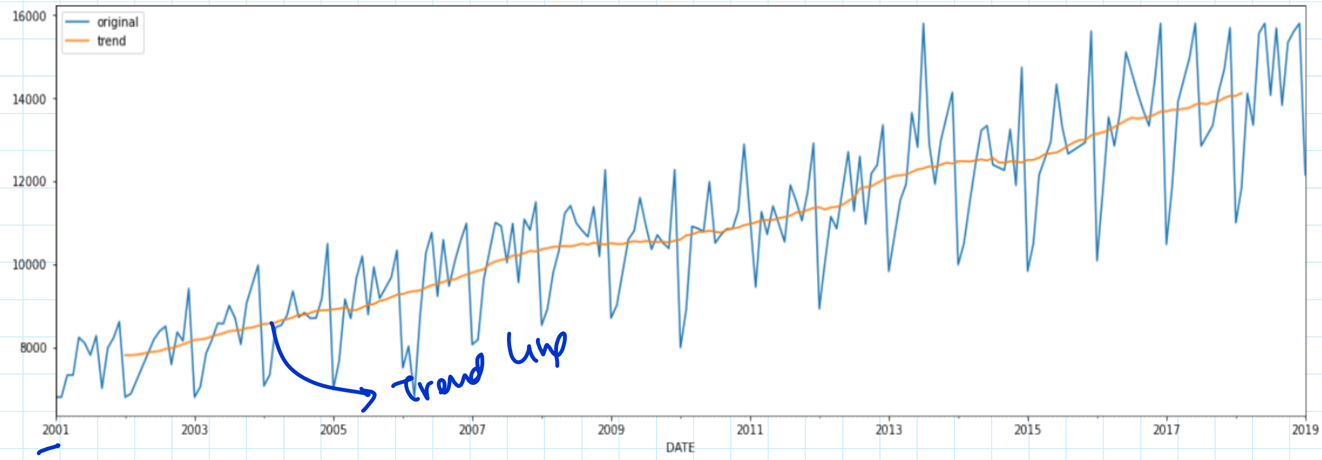
40

45

40
45

window size = 24 = (2year)

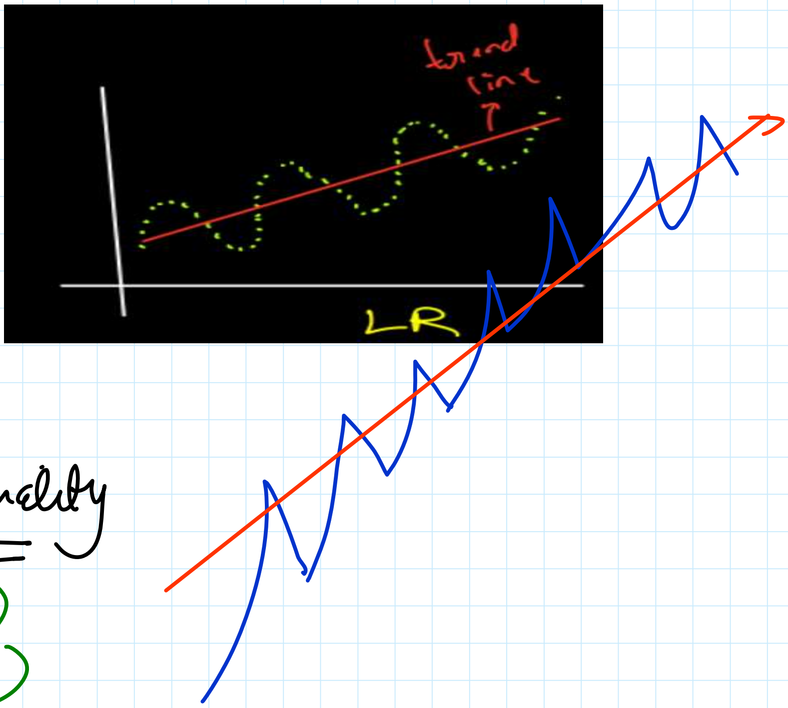
```
mobile_sales.Sales.plot(label='original')
mobile_sales.Sales.rolling(24, center=True).mean().plot(label='trend')
plt.legend()
```



2002

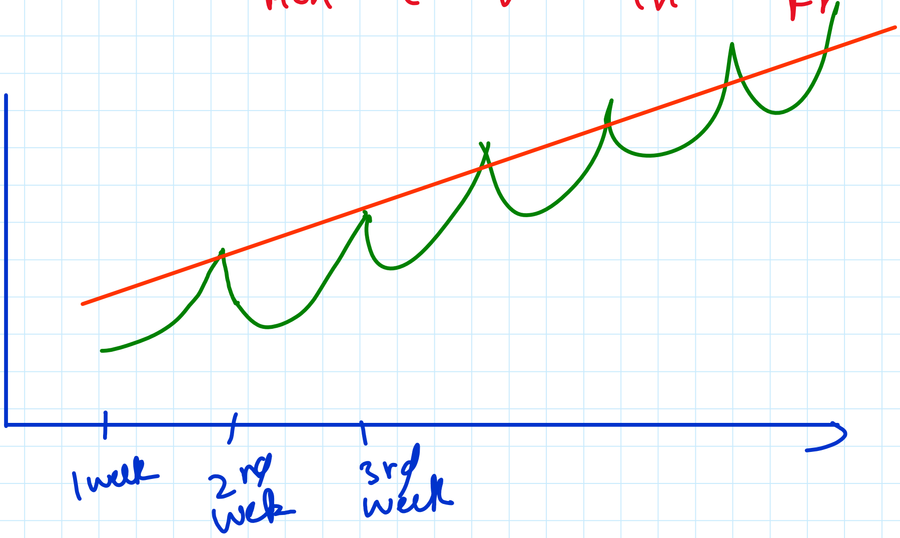
M1 # → MA of long window size (24 months)

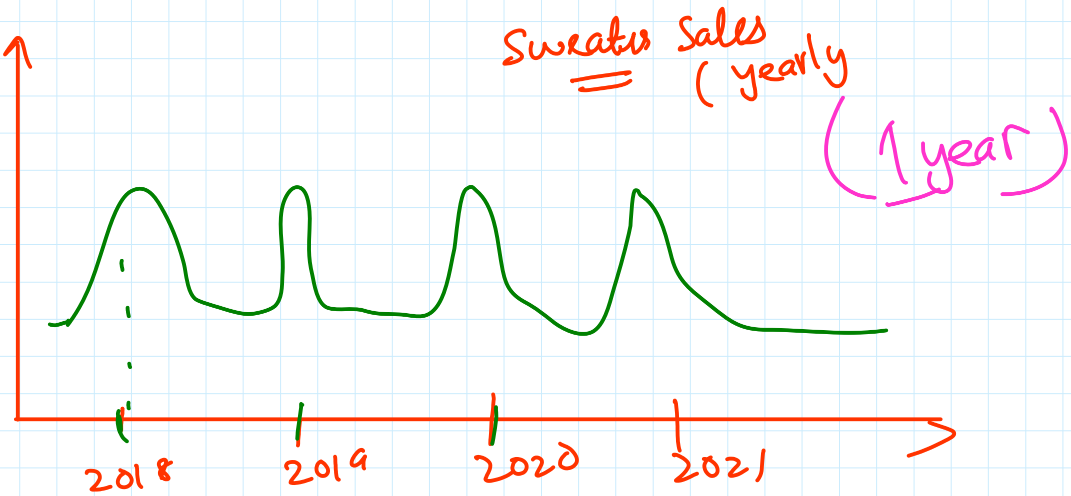
Second method → (Hint : Regression)



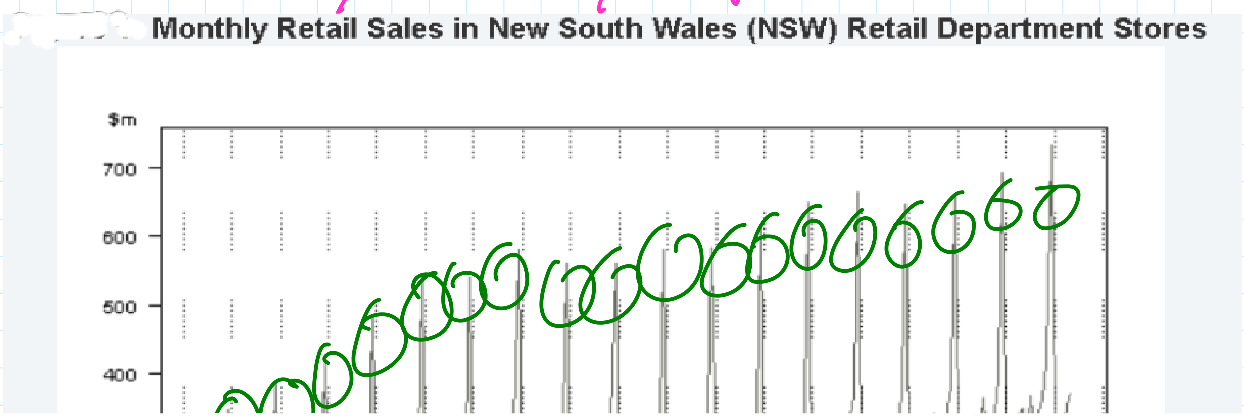
Seasonality

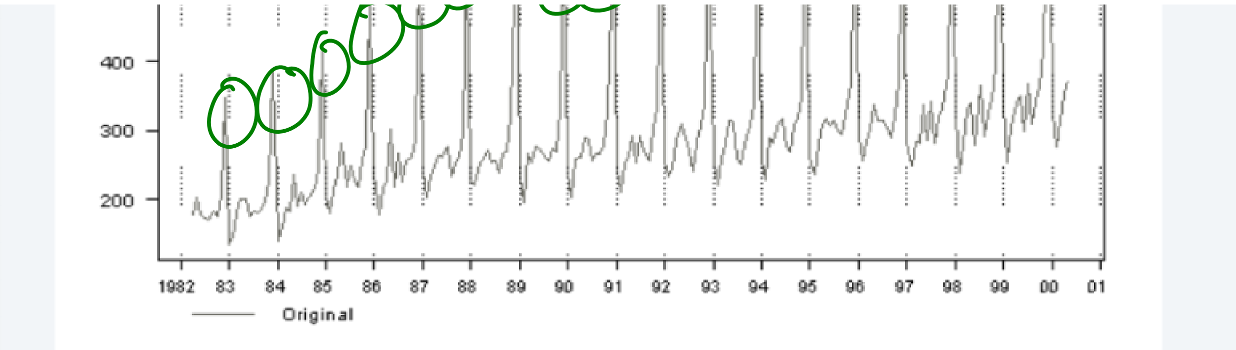
(Qtr)
(1year)



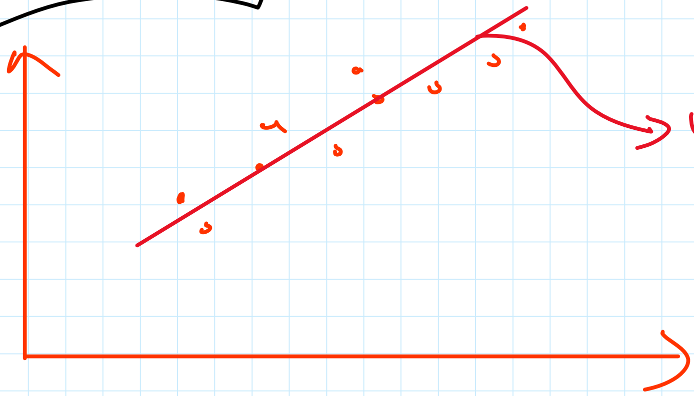


Reliance Fresh / Walmart / Target



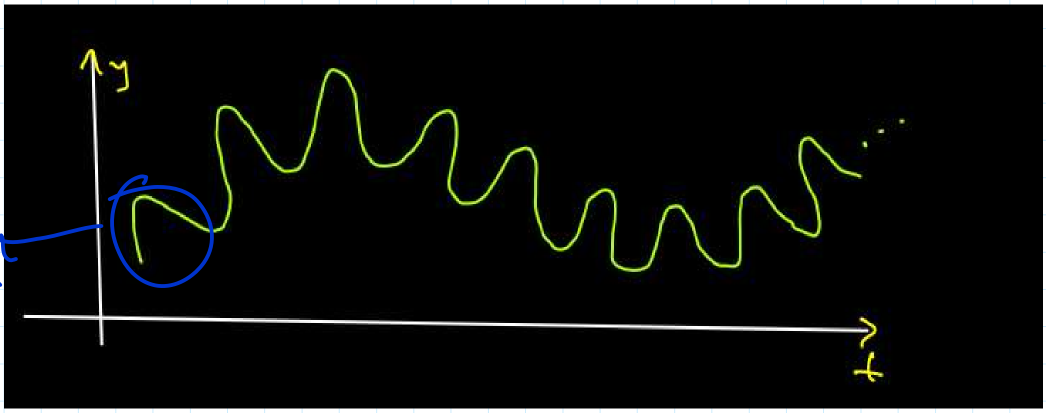


$$y(t) = \underbrace{\text{trend} + \text{Seasonality}} + \epsilon$$



$$y = (a_1x_1 + a_2x_2 + \dots a_nx_n)$$

Short Term



Retail → Short Term

Long Term (12 months)

$$Y(t) = \text{Trend} + \text{Seasonality} + \epsilon$$

||

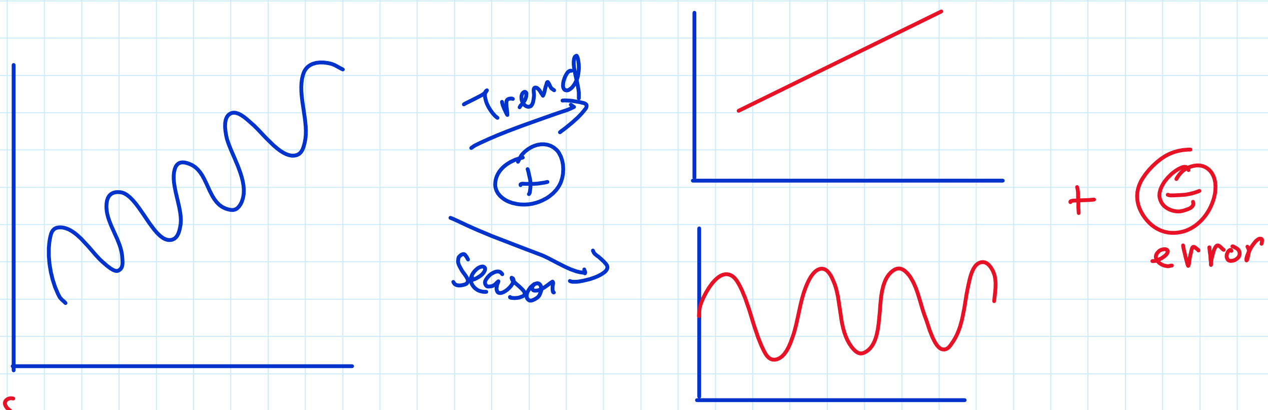
{ Long Window MA or Linea Reg }

Uniform Distrib^n → No of children bor each month

Time Series Decomposition (ETS decomp)

E → error, T → Trend, S → seasonality

ETS



ETS

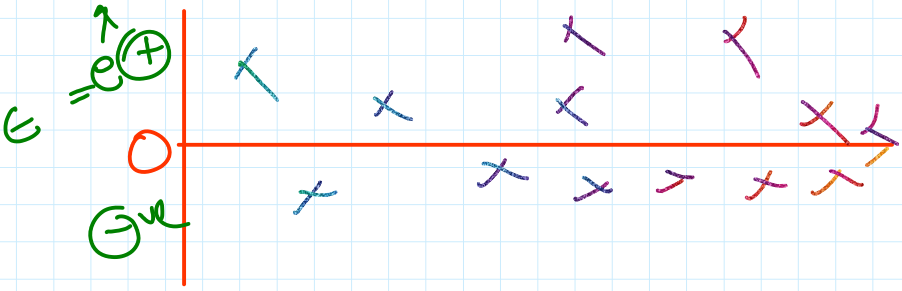
$$\{ y(t) = b(t) + s(t) + e(t) \}$$

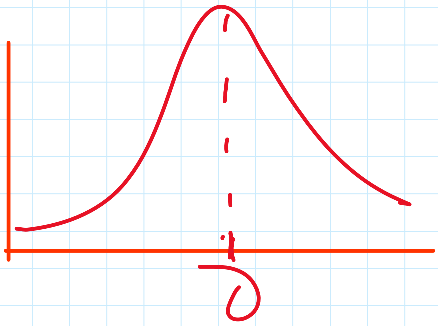
$b(t)$ → Trend, $s(t)$ → seasonalit, $e(t)$ → error

Regression $\Rightarrow y(t) = w_0 + w_1 x_1 + w_2 x_2 + \cdots w_d x_d + \epsilon_i$

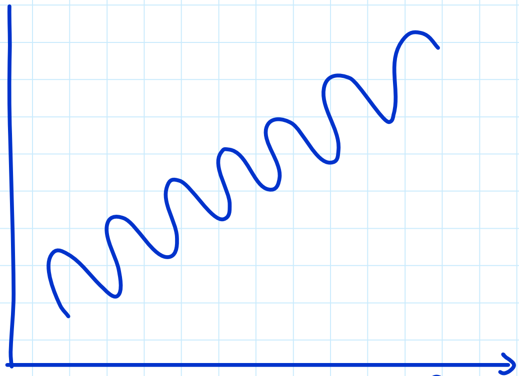
TS

$$\boxed{y(t) - b(t) - s(t) = \epsilon \text{ (Residual)}}$$



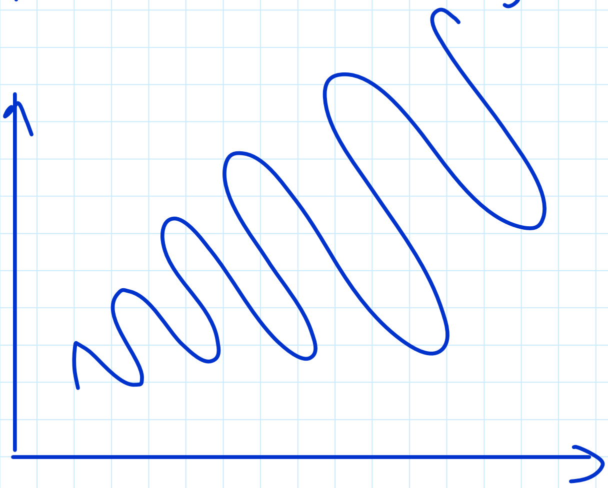


*(Additive Seasonality)



$$y(t) = b(t) + s(t) + \epsilon$$

*(Multiplicative Seasonality)



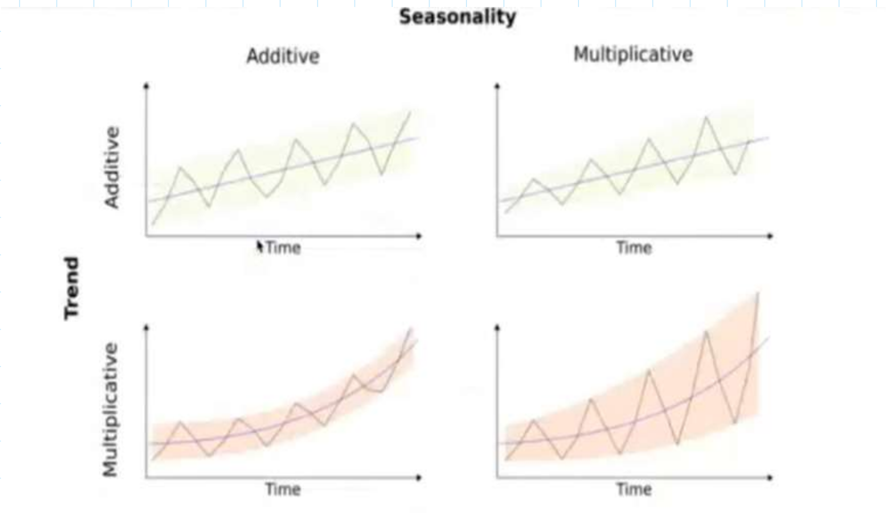
$$\{ y(t) = b(t) * s(t) * e(t)$$

$$e(t) = \frac{y(t)}{b(t) \cdot s(t)}$$

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

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

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



After → I will share a Gdrive link

Code (Colab)

Notes scribble

5 minute summary (1 page)

Interview Qs & Answer

Every 2 class

→