Date
Page

Time Series

weather forecasting
finance economics
~~weather~~
medical

→ **Non-time series**
- House Price Prediction
  ↳ size, location, bedrooms, Price
- Regression problem

Time Series
- Sales Data
  ↳ time dependent data

	Sales
Day 1	50 k
Day 2	60 k
Day 3	70 k

- Effect of "previous" timestamp"

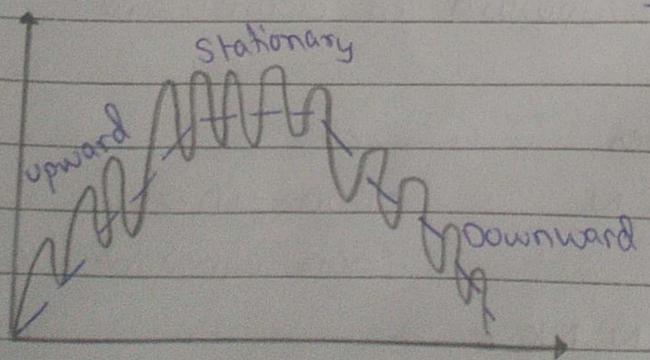
→ Interpolation vs Extrapolation

Interpolation
- Find out value in range itself.
- It assumes there should be a linear relationship

Extrapolation
- Find out value out of the range.
- Based on previous data we try to forecast future data.
- Time series problem statement will be extrapolation

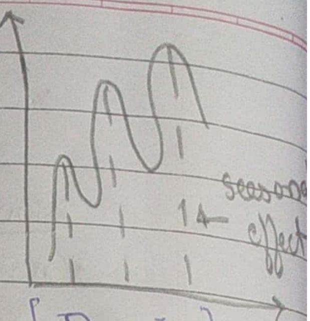
→ Components of Time Series
i. Trend, ii. Season, iii. Cycle, iv. Noise

1) Trend



- overall direction of the series.
  → 3 parts
  i) Upward
  ii) Downward
  iii) flat [horizontal]

- Stock : people's sentiment
- Sales : forecasting.

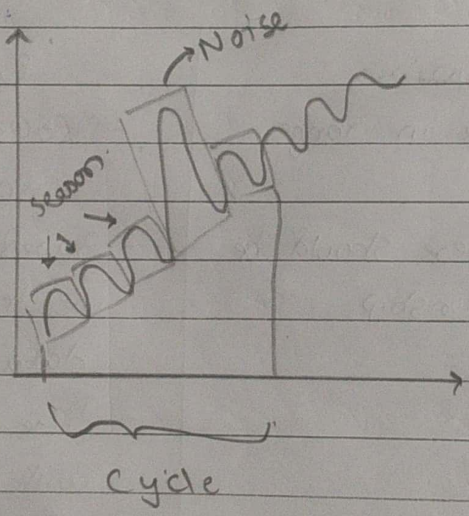


2) Season
- frequent repetition
- Example : i) Sales of ice cream in summer
  ii) Traffic at 5 pM in my area
  iii) Tourist in Goa at end of year [Dec, Jan]

- Repeation on daily, monthly, hourly or yearly basis

3) Cycle
- Time series behaviour over long time / period.
- Cycle = Season + Noise (fluctation)
- Example :

Stock price :



Politics : cycle for 5 years.

4) Noise

Some uncertainity, randomness in my data because of unpredictable reason.
↳ Pandemic, war, report, Current News
- Sudden fluctuation is known as noise.
- fluctuation can be upward or downward.

- Multiplicative Time Series
  $Y_t = T \times S \times N$
  - non-linear
  - non constant variance.

T: Trend, S: Season, N: Noise

• Additive Time Series

$Y_t = T + S + N$

• Linear over the time
• Constant Variance.

* Moving Average → Use for smoothing of Time Series data.

→ Types of Moving average.

• $\boxed{\text{Move}}$: Moving over time axis in specific window time.

1) window size
2) Average Calculation

1] SMA [Simple Moving Average]

Example:
window = 3

Time	Sales	Avg
D_1	10	NAN
D_2	12	NAN.
D_3	15	12.33 ← 1st avg
D_4	13	13.33 ← 2nd avg
D_5	14	14
D_6	16	14.33
D_7	17	15.66

rows

$\boxed{\text{* SMOOTHING}}$

- To remove all effect from the data.
- We perform smoothing with the help of MA

→ Benefits.
1) Pattern Recognition from data
2) Analyzing trend of data
3) Reduce effect of outlier
4) Enhancing the visualization

2] CMA [Cumulative Moving Average]

- findout average of all datapoint upto given time stamp.

D_1	10	$D_1 = 10$
D_2	12	$D_1 + D_2/2 = 10+12/2 = 11$
D_3	15	$D_1 + D_2 + D_3/3 = 10+12+15/3$
D_4	14	$10+12+15+14/4 =$
D_5	16	$10+12+15+14+16/5 =$
D_6	17	$10+12+15+14+16+17/6 =$

• We use CMA for long time period.

3] EMA [Exponential Moving Average] or
EWMA [Exponential weighted ....]

- In EMA we give more weightage to the recent datapoint or give more weightage to recent time stamp.

$$V_t = \beta V_{t-1} + (1-\beta)\theta_t$$

$\because$ V_t = EMA at time t.
β = $0 < \beta < 1$
V_{t-1} = EMA at previous time stamp.
θ_t = Data at t time stamp.

Ex: $\beta = 0.9$

D_1	25
D_2	13
D_3	17
D_4	31
D_5	43

$V_0 = 0$
$V_1 = 1.3$
$V_2 = 2.87$

$V_1 = \beta \times V_0 + (1-\beta)\theta_1$
$V_1 = 0.9 \times 0 + (1-0.9)\, 13$
$V_1 = 0 + 0.1 \times 13 = 1.3$

$V_2 = \beta \times V_1 + (1-\beta)\theta_1$
$= 0.9 \times 1.3 + (0.1)\, 17$
$= 1.17 + 1.7$
$V_2 = 2.87$

* Stationary & Non-Stationary Time Series.

Stationary $\rightarrow$ Mean & variance will be constant

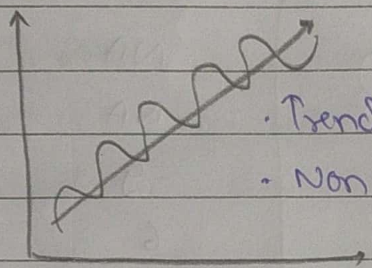
Non-Stationary $\rightarrow$ Mean & variance will not be constant.

Time Series Data → { Time Series is ST or NonST } $\xrightarrow{ST}$ Model
- ARIMA
- ANN, RNN.

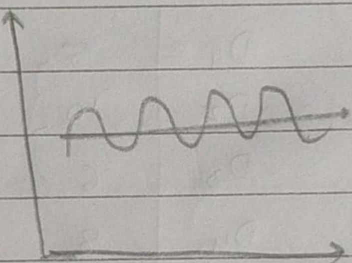
Check → Non-ST → Make → ST

→ For checking whether Time Series is ST or Non-ST ??

1) Visualization.
   - check trend, moving average.



- Trend: Upward
- Non Stationary



Trend: Flat.

- Stationary

- Moving average is increasing
- Moving average is constant.

2) Stats Based test.

ADF : Augumented dickey fullar test.
1) Static test
2) P- value
3) Critical value

H_0 = Null hypothesis
- My data is non stationary

H_a : Alternative hypothesis
- data is stationary

$\therefore P \le 0.05$
- reject null hypothesis
- Stationary

$\therefore P > 0.05$
- accept null hypothesis
- non-Stationary.

- White Noise
- random process is white noise process.
- Errors are serially uncorrelated if they are independent & identically distributed (iid).

→ From Non-Stationary To Stationary
↳ transformation.

1) Differencing
2) log
3) root
4) Seasonal adjustment.

- Differencing.

		First difference.	2nd differency
D1	5	NA	NA
D2	10	5	NA
D3	6	-4	-9
D4	8	2	6
D5	15	7	5

↳ Again check until we get stationary data.

If data is stationary stop, if not do 2nd differencing

* ACF [Auto Correlation function]

AUTO + Corelation.

corelation of itself in the feature

↳ It's a relationship b/w two features.

- Pearson
- Spearman rank
- kendall

- ACF is a funᶜ of time displacement of time series itself.
- ACF Measures the correlation between time series and its lag value.

	Timeseries		1st lag	2nd lag	3rd lag
D_1	10		NA	NA	NA
D_2	25	⟷	$D_1 \rightarrow 10$	NA	NA
D_3	14	⟷	$D_2 \rightarrow 25$	$D_1 \rightarrow$ ~~NA~~ 10	NA
D_4	16	⟷	$D_3 \rightarrow 14$	$D_2 \rightarrow 25$	$D_1 \rightarrow 10$
D_5	20	⟷	$D_4 \rightarrow 16$	$D_3 \rightarrow 14$	$D_2 \rightarrow 25$
D_6	32	⟷	$D_5 \rightarrow 20$	$D_4 \rightarrow 16$	$D_3 \rightarrow 14$
	$\boxed{y_t}$		$\boxed{y_{t-1}}$	$\boxed{y_{t-2}}$	$\boxed{y_{t-3}}$

$Corr(y_t, y_{t-1})$, $Corr(y_t, y_{t-2})$, $Corr(y_t, y_{t-3})$

* PACF [Partial Auto Correlation function]

- PACF is the conditional correlation between two variables under the assumptions that the effects of all previous lags on time series.

$t-2 \longleftrightarrow t$

$t-1$

x x

Here, t is dependent on t-1

t-1 is dependent on t-2

∴ Cancelling t-1.

* Auto Regression [AR(p)]

Auto — Regression itself in the variable

Regression

x	y

x → Independent

y → Dependent

$\boxed{y = mx + c}$

$y = m_1 x_1 + m_2 x_2 + \dots m_n x_n + c$

$$\boxed{y_t = \varphi y_{t-1} + c}$$

y_t → value at current timestamp

φ → coefficient term

c → constant

ε → error

$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_n y_{t-n} + c$

→ Machine learning

Data → Model
[Linear Reg, lasso, log, SVM, DT]

Time Series Data → Model
[ARIMA, SARIMA, SARIMAX, DL, RNN, Attention, transformer]

* ARIMA

AR	+	I	+	MA
[Auto Regression]		[Integration]		[Moving Average]
⇓		⇓		⇓
p		d		q
[0,1,2...., n]		[0,1,.... .n]		[0,1.... n]

we decide AR value using PACF

Differencing [Non stationary/ Stationary]

ACF [Corrlogram]

- Removing intermediate value & using partial value.

- Integration ⇒ Difference

$D_1 = (Y_t - Y_{t-1})$

$D_2 = (Y_{t-1}) - Y_{t-2}$

$D_2 = (Y_t - Y_{t-1} - Y_{t-2})$

- Moving Average

$$y_t = \varepsilon_{t-1}\varphi + c$$

comes with error

$$y_t = \varepsilon_{t-1}\varphi_{t-1} + \varepsilon_{t-2}\varphi_{t-2} + \dots \varepsilon_{t-n}\varphi_{t-n} + c$$

ARIMA

AR(1) $p = 1$

I(2) $d = 2$

MA(1) $q = 1$

$$y_t = \underbrace{(y_{t-1}\varphi_{t-1} + c)}_{AR} + \underbrace{(y_t - y_{t-1} - y_{t-2})}_{I} + \underbrace{(\varepsilon_{t-1}\varphi_{t-1} + c)}_{MA}$$

→ Variants of ARIMA

SARIMA
↳ seasonal

SARIMAX
↳ exogeous [outlier]