Time-Series :-

① Time-series Data :- Data points collection, often regular, time intervals.
Cross-sectional Data :- Comparing different groups at a specific moment.

- key components of time series :- Trend, Seasonality, Cyclical variation, Irregular variation
- A series is stationary :- mean, variance, autocorrelation - remain constant over time
- How do you test for stationary :- statistical test :- ADF (Augmented Dickey-Fuller)
  KPSS (kwiatkowski-phillips-schmidt-shin)

⇒ ADF :-

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p} \delta_i \Delta y_{t-i} + \epsilon_t$$

$\Delta y_t = y_t - y_{t-1}$
t = time trend
p = lag order
$\gamma = 0$ (test static)

- p-value < 0.05 reject H_0 → series is stationary
- p-value > 0.05 fail to reject H_0 non-stationary

⇒ KPSS :-

$$y_t = \gamma_t + \beta t + \epsilon_t$$

γ_t = random walk
β_t = deterministic trend
ϵ_t = stationary error term

- p-value < 0.05 non-stationary
- p-value > 0.05 stationary

- When the ADF test fail to reject the unit root null, but the KPSS test rejects its stationarity null. In this case, the series is often trend-stationary.

- A time series is stationary if it's mean, variance, autocovariance are constant over time.
  if not → transform it.

a) Differencing :- Remove trends & unit roots.

$$y'_t = y_t - y_{t-1}$$
$$y''_t = y'_t - y'_{t-1}$$

b) log transformation :- $y'_t = \log(y_t)$
→ stablizes variance

c) Square Root/Cube Root :-

d) Box-Cox transformation :-

$$y'_t \begin{cases} \dfrac{y_t^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log(y_t) & \lambda = 0 \end{cases}$$

e) Seasonal Differencing :-

$$y'_t = y_t - y_{t-s}$$

f) standardization/Normalization :-

g) Detrending by model fitting :-

- stationarity :- How a time series is changing will remain the same in the future.
- A unit root is a stochastic trend called a "random walk with drift"
  Randomness can't be predicted

Important
- unit root present = not stationary (unpredictable)
- unit root absent = stationary

→ To test for stationarity with a unit root test.
assume : Null hypothesis (H_0) : no unit root
Alternative hypothesis (H_1) : unit root

② White Noise :- A sequence of random values with, $(\mu = 0) + (\sigma^2 = \text{constant}) +$ [No autocorrelation]
↳ In time series modeling, pure randomness.

• Residuals = actual data − model's fitted values.
↳ After fitting ARIMA/ETS/Regressions
if residuals ≈ white noise, then model has captured all systematic patterns. (trend, seasonality, autocorrelation)

⇒ Trend stationary & Differenced stationary series :-

Series that becomes stationary once you remove a deterministic trend. (straight line)

$$y_t = \alpha + \beta t + \epsilon_t$$

Series that becomes stationary only after differencing.

$$y_t = y_{t-1} + \epsilon_t$$

$$\Delta y_t = y_t - y_{t-1} = \epsilon_t$$

⇒ AR vs MA vs ARMA :-
AR → Autoregressive model
MA → moving avg.
ARMA → autoregressive moving avg.

• Current value of the series depends on the past [value + noise]. (like momentum)

• $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$

y_t : current value
y_{t-1}, y_{t-2} : past value
ϵ_t : white noise

[• yesterday's return influences today's return]

MA :-
• current value = past shocks (errors)

• $y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$

$\epsilon_{t-1, t-2}$: past shocks

[• Today's return may depend on yesterday's partly.]

ARMA :-
• current value = (AR + MA)

• $y_t = c + \phi_1 y_{t-1} + \cdots + \epsilon_t + \theta_1 \epsilon_{t+1} \cdots$

[• Most important] → if series is stationary → ARMA
if non-stationary → ARIMA

ARIMA :- (Autoregressive Integrated moving avg.)

• AR(p) + MA(q) + ARMA(p,q) + ARIMA(p,d,q) → p = no. of autoregressive terms
q = no. of moving avg. terms
d = no. of differences
↳ handle non-stationary

• ARIMA(1,0,0) → AR(1)
" (0,0,1) → MA(1)
" (1,0,1) → ARMA(1,1)
" (0,1,0) → Random walk
" (0,1,1) → Exponential smoothing
" (1,1,1) → $Y_t = \phi_1 Y_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$

SARIMA :- (seasonality)
• SARIMA(p,d,q)(P,D,Q,s)
↳ seasonal period

⑧ ACF & PACF :- (Partial Autocorrelation Function) :-

- Correlation of y_t with y_{t-1}, y_{t-2} ..... , it's tell you how past values in general affect the current value.
- Correlation of y_t with y_{t-2}, after removing y_{t-1}

[• Before plotting ACF/PACF, you must make the series stationary.
• PACF helps guess p, ACF helps guess q.]

⇒ Confidence band :- (statistical threshold for significance)

- The spikes in ACF & PACF plots tells you whether a correlation at a given lag is statistically significant or just noise.

• Bands :- A threshold dashed blue lines.
• How to calculated :- For a large sample size N: std. error of autocorrelation $\approx \frac{1}{\sqrt{N}}$
• All spikes within the band, series behave like white noise.

④ MAE (mean absolute error) → $mae = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n}$

RMSE (root mean squared error) → $rmse = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2}$

MAPE (mean absolute percentage error) → $mape = \frac{100}{n}\sum_{t=1}^{n} \left|\frac{y_t - \hat{y}_t}{y_t}\right|$

AIC (akaike information criterion) → $AIC = -2\ln(L) + 2K$

BIC (bayesian information criterion) → $BIC = -2\ln(L) + K \cdot \ln(n)$

L : likelihood of the model
K : no. of parameter
Lower AIC → better model.

• These are used to measure How well the forecasted values match actuals value.
• How well the model structure fit, penalizing complexity.

⑤ How to Handle the missing values in time series data? → Random missing / systematic missing.

i) Forward Fill :- $y_t = y_{t-1}$
ii) Backward Fill :- $y_t = y_{t+1}$
iii) linear interpolation :- Missing points b/w known values. → $y_t = y_{t-1} + \frac{(y_{t+1} - y_{t-1})}{2}$
iv) Time-based Interpolation :- Spline/polynomial (use curve line)
v) Model-based :-
vi) static (mean, median, mode) :-

• How to Backtest a Forecasting model.
→ Train-test Split
→

⑥ GARCH (generalized autoregressive condition Heteroskedasticity) :-

- Designed for variance (volatility). where ARIMA designed for mean (Returns)
- today volatility = past volatility + past forecast errors.
- GARCH (P, Q)

$$h_t = \alpha_0 + \sum_{i=1}^{P} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{Q} \beta_j h_{t-j}$$

$\because h_t = \sigma_t^2$

- h_t → conditional variance (volatility at time t)
- ε_{t-i}^2 → yesterday's & before squared error
- h_{t-j} → yesterday's & before volatility

- Use Forecast risk (VaR, option pricing)
- position size based on expected volatility
- Identify High-risk & low-risk regimes

⑦ Exponential Smoothing Models :-

- Recent data gets more weight, older data get exponentially less weight.
- Simple ES :- $\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$ → data has no trend → mean
- Holt's linear trend :- mean + trend

Level : $L_t = \alpha y_t + (1-\alpha)(L_{t-1} + T_{t-1})$

Trend : $T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

Seasonal : $S_t = \gamma(Y_t - L_t) + (1-\gamma)S_{t-m}$

⑧ Machine Learning Approaches :-

- Random Forests :- & logistic Regression :-
- Support vector Machine :- (SVM)
- Recurrent Neural Network :- (RNN) + Reinforcement learning
- Long short-term memory :- (LSTM)
- Gated Recurrent units :- (GRU)
- XGBoost classifier :-

⑨ How to deal with large datasets or high-dimensional feature spaces.

- Optimization Algorithms :- {Stochastic gradient Desent (SGD) (Time consumming + power computation)
  Bayesian Optimization
- Parallel Computing :- Use Amdahl's law $S = \frac{1}{(1-P) + \frac{P}{N}}$ (S → speedup, P → program, N → Number of processors)
- Dimensionality Reduction :-
  - PCA (Principal component analysis)