The First Note

09 December 2024 14:03

Hi all welcome to the Series

Time Series Data ?

↳ is a collection of observations collected or recorded over a sequence of time interval

Ex → Price of a stock collected daily

AAPL → 01-01-2024
02-01-2024
03-01-2024
⋮
20-12-2024

Temperature → 01-01-2024 - 00:00 - T_0
00:01 → T_1
00:02 - T_2

Why time series data is treated differently ?

Key characteristics :-

→ chronological order —
↳ ordering of time →

→ Sequential dependence ←

→ Temporal component → - Trend
- Seasonality

→ [illegible]

- Seasonality →
- cyclic patterns →
- Noise →

→ constant frequency → 01-01-2024 ←
02-01-2024
03-01-2024
04-01-2024
04-01-2024 00:01
00:02

→ Dynamic nature → external factors.

Time series Analysis →

understand the past → patterns

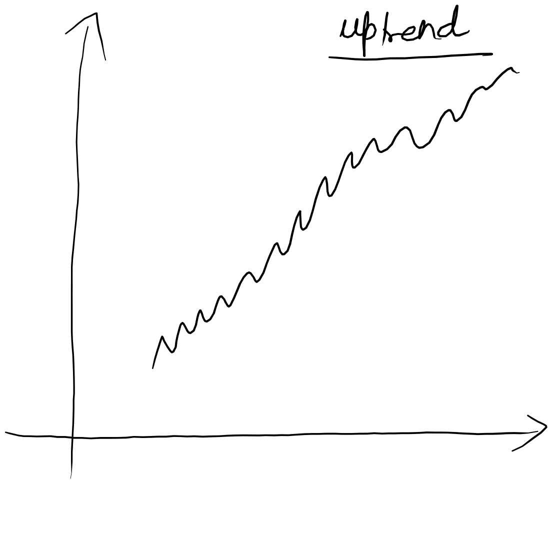
forecast the future

components of Time Series :

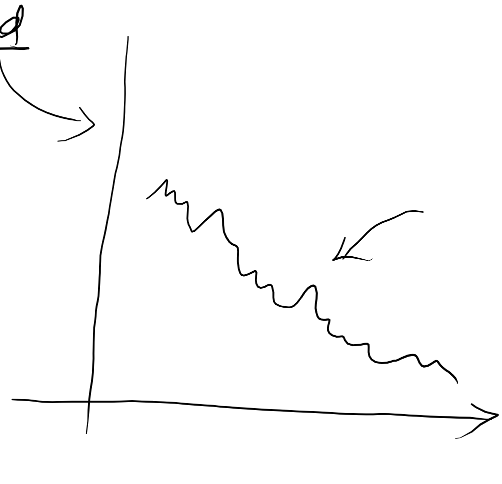
→ Level → the average of your entire data

→ Trend → The long-term direction of data

uptrend

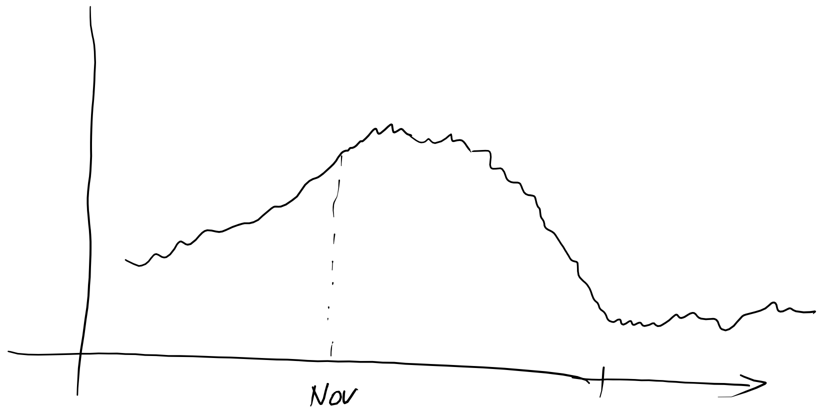


Downtrend



seasonality → september

→ Blankets → Sales of Blanket



in a calender year → daily patterns,

→ weekly patterns →

→ monthly →

→ Quarterly patterns →

→ yearly →

5-yearly pattern is not considered seasonal

→ regular or predictable patterns over a fixed period of interval like daily, monthly, etc -- generally falling in a calender year

cyclic patterns → (i) repeating pattern over a longer duration

(ii) There is no strict criteria of repetition over a fixed duration

Recession

→ 2009-10 → covid

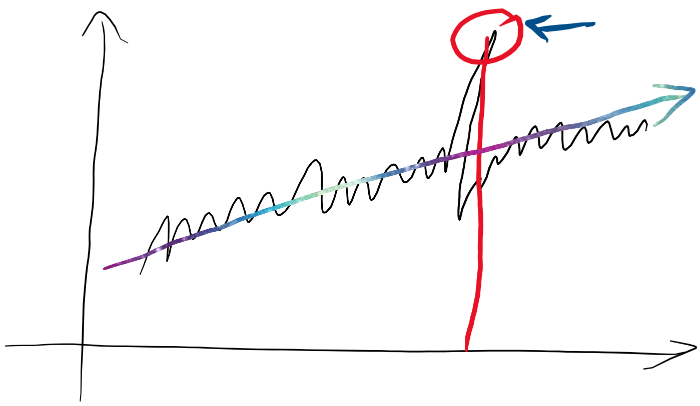
Big Short

→ Two types of cycle :

(i) Economic cycle

(ii) Business cycle →

→ Noise or Residual → Random fluctuations in the data which does not have any explanation



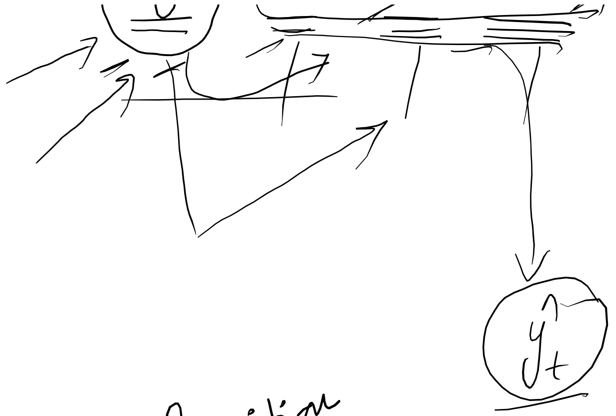
Decomposition Method

→ Additive model

$$y_t = \boxed{T_t + S_t + R_t}$$

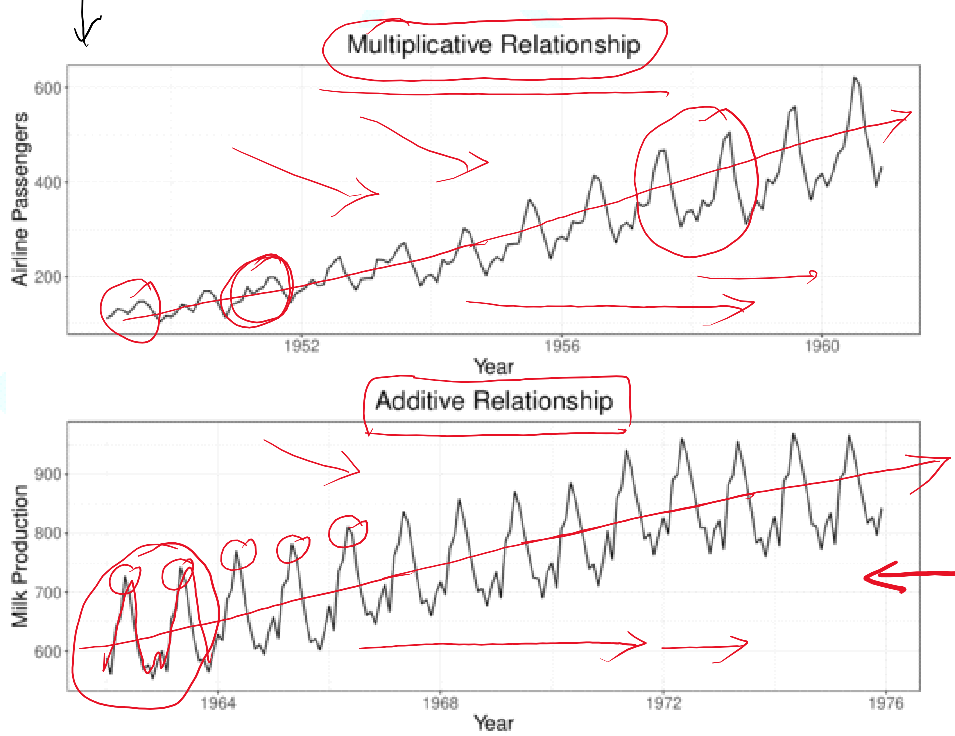
Multiplicative Model

$$y_t = T_t * S_t * R_t$$



classical decomposition

and $y_t = \hat{y}_t$



$T_t * S_t$

→ STL decomposition → Seasonal and Trend decomposition

↓

LOESS

locally estimated scatterplot smoothening

→ only supports additive decomposition

→ classical method assumes fixed seasonality

↳ fails to learn the variations in seasonality

↳ fails to learn the variations
in seasonality

on the contrary, STL method can learn the variations
in seasonality.

stationary data :-

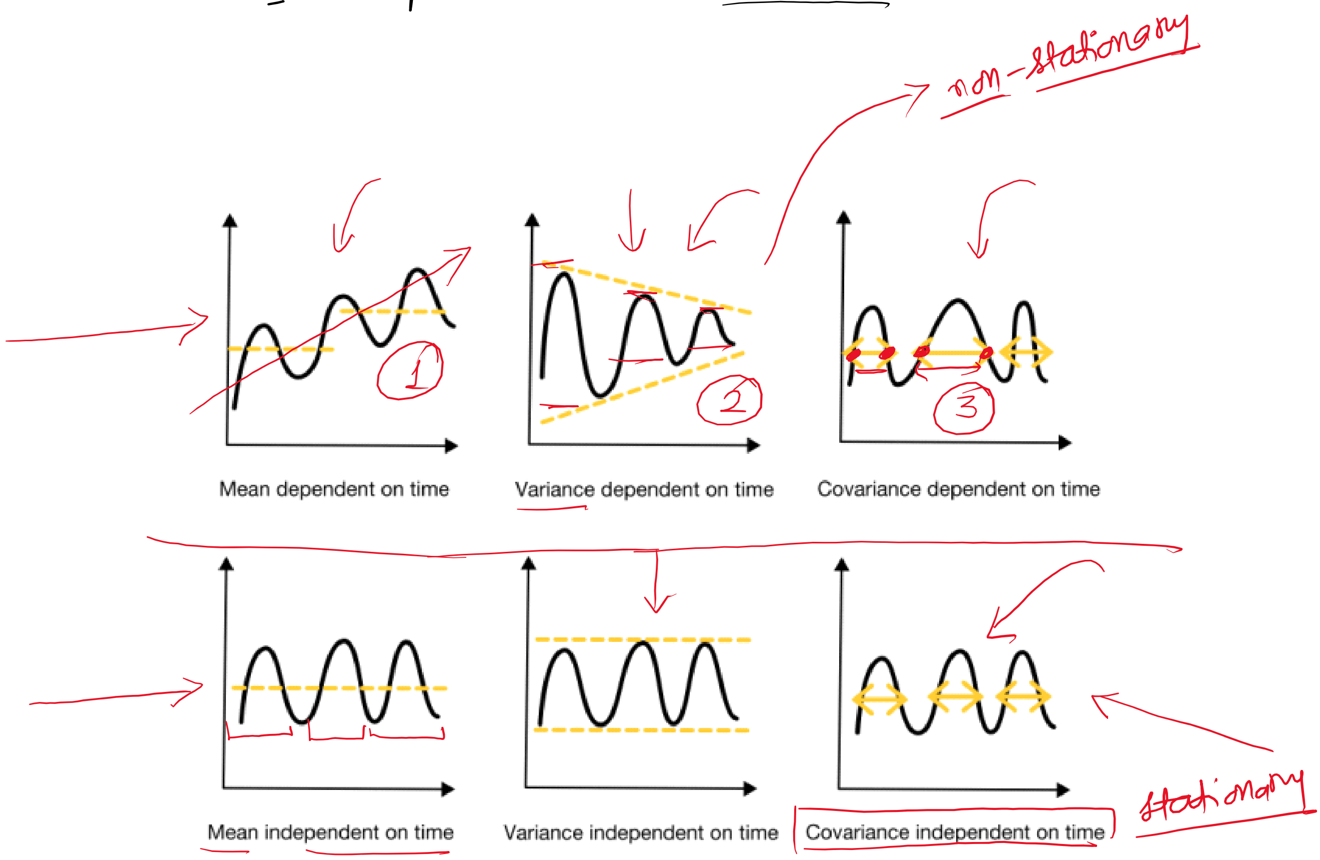
statistical prop → mean
Variance
Auto-correlation

→ mean → constant
Variance and
auto-correlation → constant over time
then time series data is considered stationary

why is it important?

(i) majority of models for time series forecasting
assumes the data to be stationary

(ii) help us make better prediction



Mean independent on time → Variance independent on time → Covariance independent on time → Stationary

covariance →
↓
when you want to understand how does the value of var1 changes or varies when there is a variation or change in var2

correlation → [-1 to 1]

Variance → understand the average of the spread of the data around mean

→ case-1

→ case-2

savings and expenses → -1

if you inc exp by 3 times, results in dec. savings by 3 times.

salary and YOE → +1

How to check for stationarity

stationarity → weak stationarity
stationarity → strict stationarity

weak stationary → mean, variance and auto-correlation → stationary over time

strict stationarity → exhibit all the properties of weak stationary ... different samples from strict stationarity → exhibit all the properties of weak stationary

- joint distribution of different samples from the same time series should be same.

$$\left(y_{t_1}, y_{t_2}, y_{t_3}, y_{t_4}, y_{t_5} \right) \quad \left(y_{t_{21}}, y_{t_{22}}, y_{t_{23}}, y_{t_{24}}, y_{t_{25}} \right)$$

ADF test → checks for the presence of unit root in the data

Presence of non stationary trend

unit root → non-stationary

Null hypo: (H_0) → has a unit root → (non-stationary)

Alternate hypo (H_A) → does not have a unit root → stationary

p-value: →

p-value < 0.05 → we reject null

p-value :

p-value $< 0.05 \rightarrow$ we reject null

$\geq 0.05 \rightarrow$ we fail to reject the null Hypo

KPSS test :→

H_0 :- the data is stationary

H_A :- non-stationary

both the above tests are for weak stationarity

strict stationarity →

KS test

p-value > 0.05 → no difference in the dist.

↳ stationarity

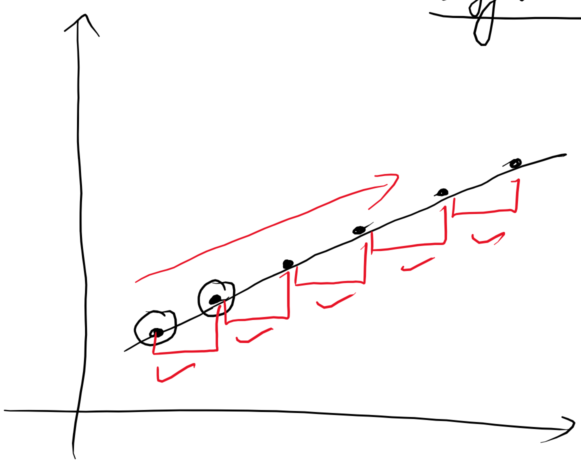
null Hypo → stationary

Alt Hypo → non-stationary

smoothening techniques

— differencing → remove trend and seasonality

de-trending



$$y_t' = y_t - y_{t-1}$$

$$y_{t+1}' = y_{t+1} - y_t$$

$U_{t+1} = y_{t+1} - y_t$

first order differencing

→ second order differencing

$y_1, y_2, y_3, y_4 \; ------ \; y_t$

→ FOD $\quad y_2' \; y_2', y_3' \; ------- \; y_{t-1}'$

SOD $\quad y_1'' \; y_2'', \; ---$

second order differencing

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