

* Time Series

time series \rightarrow A series with time component.

x_1 Area of house	x_2 No. of Rooms	y Price of house.
		1

MLR.

(timeskip) Month	Sales
Jan	60k
Feb	40k
March	50k
April	20k
...	...

\rightarrow Time Component is involved

Area of house	Price
1000	60L
950	50L
1100	75L

1100	75L
1000	60
950	50

(Rearrange of row, x -train was selected randomly)

Jan	15k
Feb	20k
Mar	60

\neq

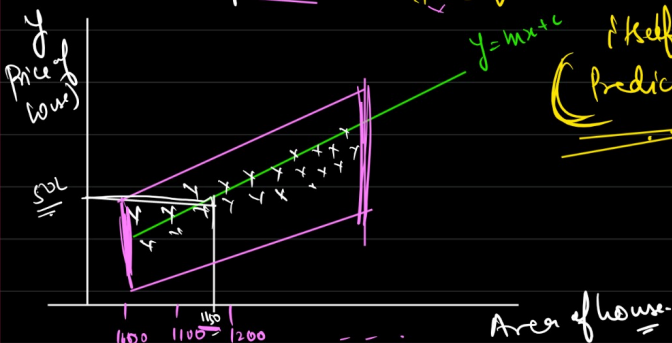
Mar	60
Jan	15k
Feb	20k

\rightarrow Here order matters.

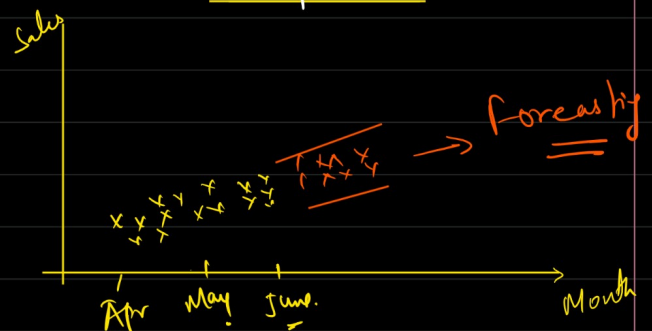
\Rightarrow because the current time period value depends on the previous time period.

* Can we use regression model for time series data?

* interpolation \Rightarrow To find out the value in range (itself) (prediction)



Extrapolation.



* Most of time (99%) the test data will come in training range.

training (Area) = 1000 - 10000

testing (Area) = 11

* if outside training range \Rightarrow wrong prediction

* Time series problem will be extrapolation (forecasting)

* Based on previous history predict next 6 months

* Why not Linear Regression for time series?

\rightarrow time component is involved.

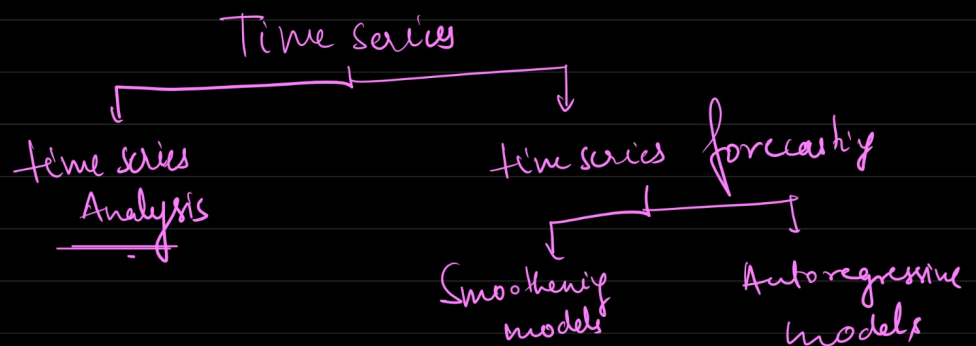
\rightarrow Because of Extrapolation, Prediction may be wrong.

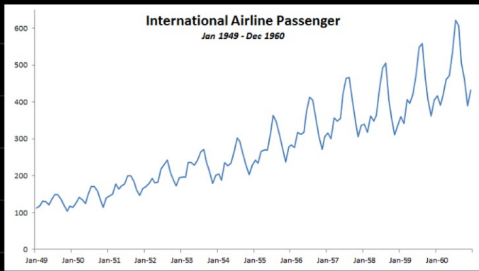
\rightarrow LR assumes linear relationship but in time series the current observation depends on previous obs \rightarrow which is not true for non time series data.

* Motivation

- ① Weather forecasting \Rightarrow weather pattern day wise, month wise, year wise
- ② Medica
- ③ Stock Market
- ④ Ecommerce / Finance — Sales, bond price.

* Whenever time component is there in data, you can use time series.





Components of time series

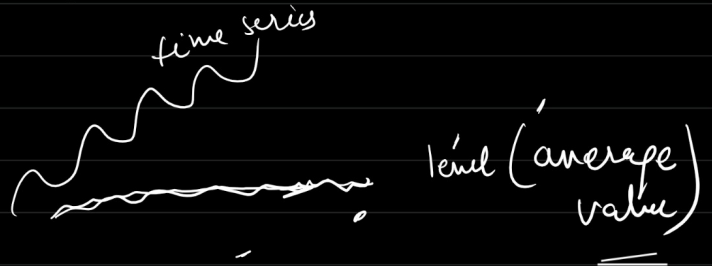
- ① Level
- ② trend
- ③ Season / seasonality
- ④ Cycle / cyclicity
- ⑤ Noise

① level → The base value of a time series on which other components are added.

② trend

↳ long term movement

or direction in data over a long period of time



① upward

② downward

③ Horizontal



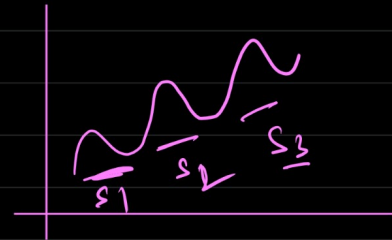
③ Season → frequent repetition in a data over regular interval. daily, monthly, annually

ex. traffic in peak hours

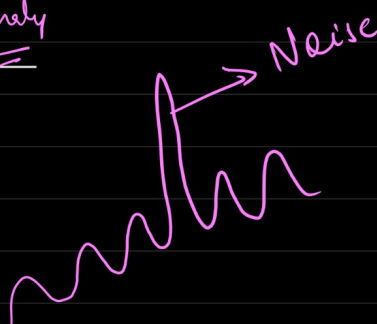
ex. Sales of TV in diwali

ex. No. of tourists in peak season

ex. No. of Ice-cream sales.

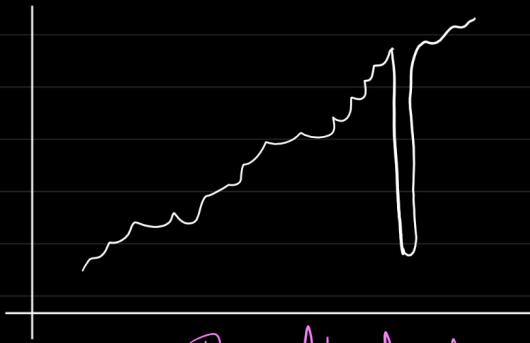


④ Noise / Anomaly



Some uncertainty / randomness in time series data because of unexpected reason.

- ① News
- ② Reparat



③ Pandemic

④ war

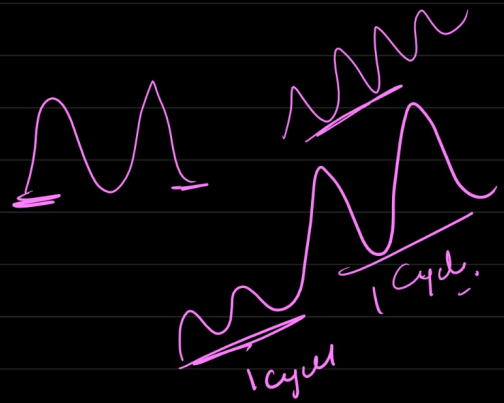
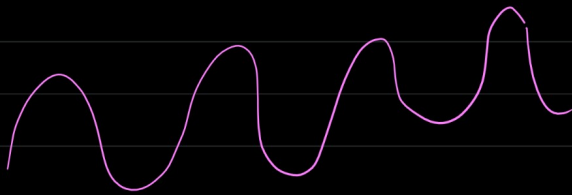
⑤ election

⑥ influenza

⑤ Cycle \Rightarrow The fluctuation in the data over a longer period of time. These periods are not fixed and can vary.

$$\text{Cyclicality} = \text{Season} + \text{Noise}$$

Pharmaceutical industry Stock Price



tractor making coupon \Rightarrow Harvesting months