* Time Series

In Time series $\xrightarrow[\text{have}]{\text{we}}$ Timestamp

① Data → univarient time series
→ Multivarient time series

component
↳ Seasonality
↳ trade
↳ cycle)
↳ Error

② Analysis of Time series $\xrightarrow[\text{series are}]{\text{Time}}$ Stationary or non-stationary
(mean, var)

③ Processing (If data is Non-Stationary $\xrightarrow[\text{can make}]{\text{we}}$ Stationary)

④ Model Building

⑤ Evaluation → Evaluation of Time Series can be done using (AIC, BIC). we can find Loss also (mse)

- w.r.t Time series we are able to see
  - ↳ Time Series Analysis
  - ↳ Time series forcasting (Just a prediction) (model building)

- **Model**
  - ↳ ARIMA

It is called classical technique inspired by Linear Regression

variation (ARIMA)
AR ARMA SARIMA SARIMAX

ARIMA model is based on Auto Regression

w.r.t ARIMA we are able to see → ARIMA
p d q

Corelation within a feature

- facebook $\xrightarrow{\text{from facebook}}$ prophet
  (we are using this "Prophet" Library for Time series)

- Google $\xrightarrow{\text{from google}}$ Temperal fusion Trasformer (TFT)

Time stamp

→ Data →

	Time	Temp
1	5:00 am	54° f
2	6:00 am	54° f
3	7:00 am	58° f
4	8:00 am	60° f
5	9:00 am	62° f
6	10:00 am	64° f
⋮	⋮	⋮

feature

univarient Time Series (Single feature)

we are taking timestamp
↳ 0 to 5 $\xrightarrow[\text{predicting}]{\text{and we are}}$ 6th
then ↳ 6 to 10 $\xrightarrow[\text{predicting}]{}$ 11th

within a variable we are going to create a Timestamp (or) Predict

* we can solve using VAR (Multi-variant)

* eg
* Time series → (1) Stock price
  (2) Bitcoin / Crypto
  (3) weather forcasting

Date → Time (Timestamp)	Temp	cloud cover	dev point	humidity	wind
5: am	59° f	97 %	51° f	74 %	8 mph SSE
6: am	59° f	89 %	51° f	75 %	8 mph SSE
7. am	58° f	79 %	51° f	76 %	7 mph S
8: am	60° f	74 %	51° f	77 %	7 mph S
9: am	62° f	74 %	52° f	74 %	8 mph S

Multivariant timeseries

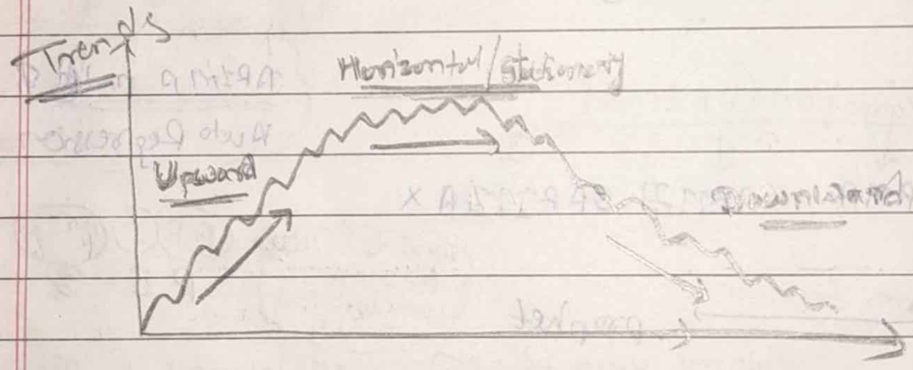
Can be solved using **VAR** (Vector Auto Regressive) ← Inspired from the ARIMA model.

Imp. • Analysis
* → Trend → overall direction of the series.
* → seasonality → Monthly (or) Quartely pattern. eg:- (1) Selection of PM (2) world cup (olompic)
* → Cycle → long-term business cycle.
* → Irregular remainder → random noise left after Extraction of all the components (eg:- flood, corona virus, war)

* We can do Analysis on Time Series based on diff-diff Components

Repeating trend

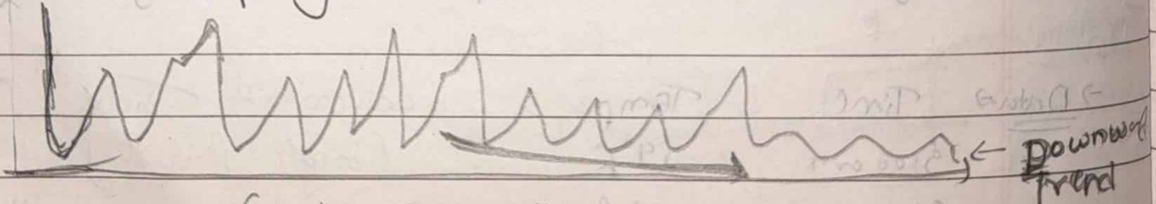
→ Trends



eg:- Sales (Increase / Decrease)

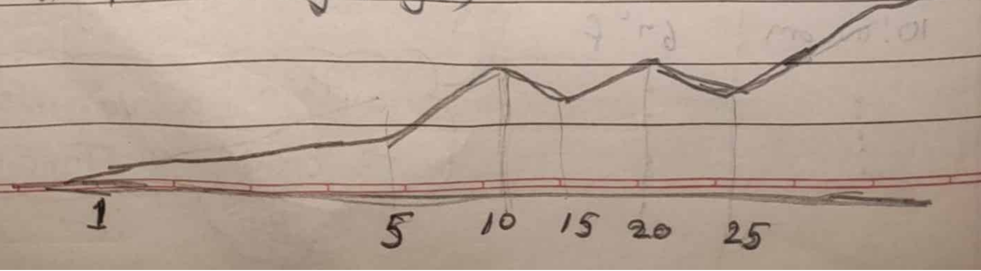
→ Trent + Seasonality

→ Seasonality :- Repeating trends



Google Trends - "Snowboarding"

→ Cyclical :- Trends with no set repetition
(Time stamp is v.very Huge)



- Stationary Vs Non-Stationary Data

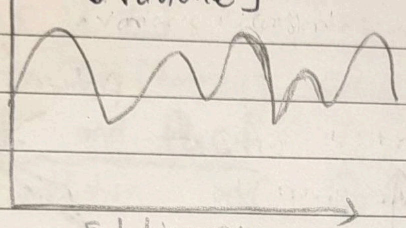
has constant mean and variance over time

- we are not having constant mean ~~and variance~~

Stationary → Mean needs to be constant

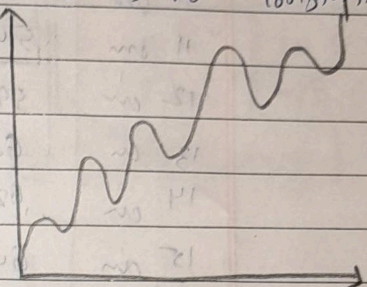
- Mean
- Variance

} are constant



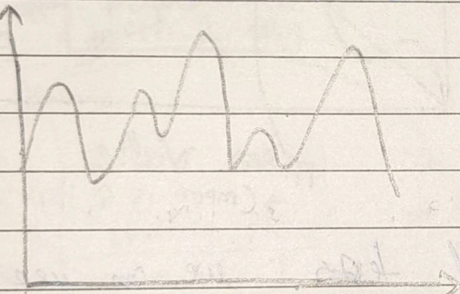
Stationary

- Mean is not constant

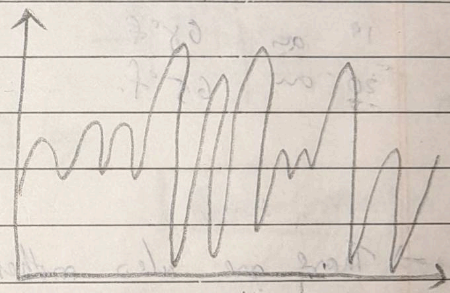


Non-Stationary

→ Variance should not be a function of time



Stationary



Non-Stationary

- Stationary dataset will allow our Model to predict that the mean and variance will be the same in future periods

- Mean and Variance are Constant → Stationary Time Series

whenever we are talking about mean/variance one thing comes into picture that is

→ Rolling Statics $\xrightarrow{\text{inside Rolling Statics one more terminology is}}$ Windows

Data :-

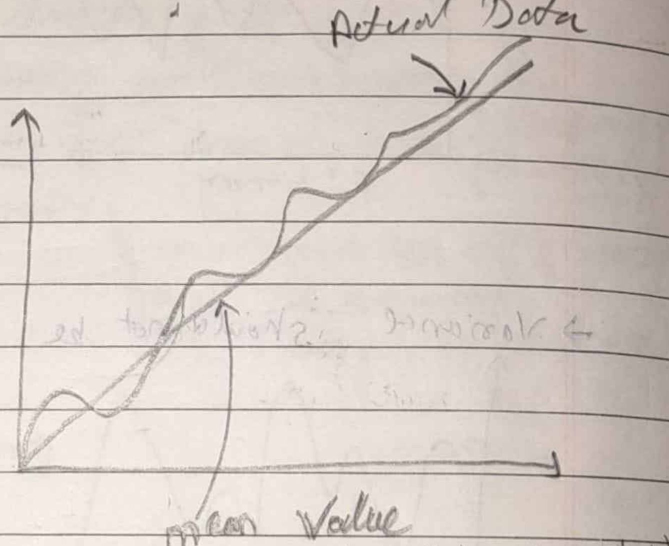
Time	Temp
6 am	59°f
7 am	54°f
8 am	60°f
9 am	62°f
10 am	60°f
11 am	54°f
12 am	59°f
13 am	60°f
14 am	62°f
15 am	64°f
18 am	58°f
17 am	57°f
18 am	61°f
19 am	68°f
20 am	65°f

5

5

5

1st i'll take this 5 window
↓ then
calculate mean
2nd take 5 window
↓ then
calculate mean
⋮



→ (mean is ↑, it is not constant)

→ There are also mathematical tests we can use to test for stationary in your data.

→ A common one is the Augmented Dickey - fuller test (ADF)

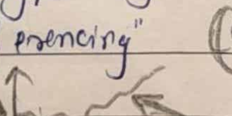
→ If we've determined your data is not Stationary (either visually or mathematically), we will then need to trasform it to be Stationary in order to evaluate it. (This is a preprocessing step)

→ Eg

What type of ARIMA terms you will use

→ "Differencing" (Using Differencing we can make Data into Stationary)

Constant



→ In my Time Seore mean is increasing. So at that time we can do a differencing, then my mean will be a constant.

→ Non Stationary $\xrightarrow{\text{to}}$ Stationary {that time will use Differencing}

Date:
Page:

→ Differencing concept ⇓

1 lag — first diff (1 NA)

2 lag — second differ (2 NA)

Original Data	
Time 1	10
Time 2	12
Tim 3	8
Tim 4	14
Tim 5	7

(for 2nd diff leave these 2 points)

First	Difference
Time 1	NA
Time 2	2 (12-10)
Time 3	-4 (8-12)
Time 4	6 (14-8)
Time 5	-7 (7-14)

Second	Difference
Time 1	NA
Time 2	NA
Time 3	6 (-4-2)
Time 4	10 (6-(-4))
Time 5	-13 (-7-6)

Note ⟹ (using Differencing we can make Data into Stationary)

- **Acf and Pacf**

(Autocorrelation and Partial Autocorrelation)

ACF • Autocorrelation → correlation itself / correlation within a feature

• Lag # Autocorrelation with 1 lag

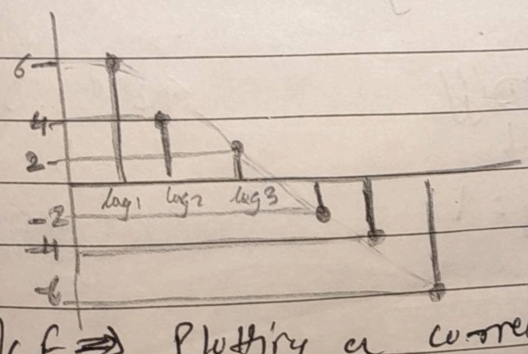
→ 1st Lag → 1st Difference

Data → original time series	1 Lag
10	
15	15
20	20
25	25
30	30

writing these value in 1 Lag →

Auto-Correlation means
↓
Correlation Within feature w.r.t Timestamp, (or) Lag

• Corr lag graph



Correlation means
↓
between 2 feature we are cheking similarities

Acf ⟹ Plotting a correlation with lags

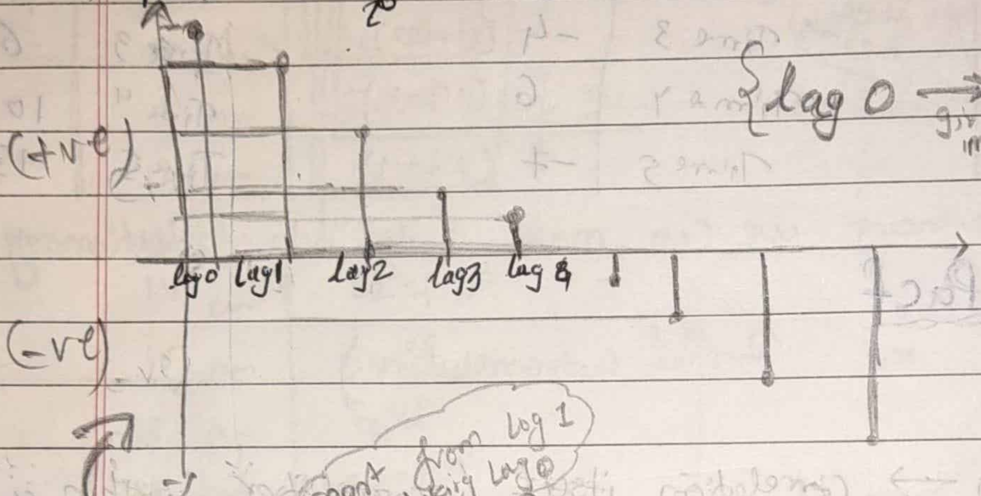
Correlation ⇒ Between 2 feature we are cheking similarity,

Auto correlation ⇒ within feature we are cheking relation w.r.t Lag

- AR → P, I → D, MA → Q — This 3 are Different Model. PDQ is the Representation of AR I MA

AR → we can represent

→ ACF $\xrightarrow[\text{to}]{\text{is related}}$ Correlation



{lag 0 $\xrightarrow[\text{giving me}]{}$ Correlation 1

apart from lag 1 not taking lag 0 then

⇒ for creating Auto Regressor model which lag i will choose?

→ Lag 1 (because correlation is % very Good over here)

*(from the above Graph)

eg:- Data ⇒

1	1
2	2
3	3
4	4

{Lag = 0, Correlation = 1}

* {for Better understanding ~~Refer~~ of LAG Refer DIFFERENCE eg.

original Data → {lag = 0, corr = 1}

1st Diff → lag = 1

2nd Diff → lag = 2

Auto Regresive
Based on previous Time Stamp we are going to predict the Next Time Stamp with a feature

Moving Average
The Moving Average model Assumes that the observed time series can be represented by a Linear Combination of white noise error terms.
~~The time series will always~~ Stationary

AR + I + MA

Auto Regresive — Integration (or) difference — Moving Average

→ In AR$\boxed{I}$MA we are going to conclude my Integrated model

$y = AR + I + MA$

version of ARIMA is → ARMA

I ⇒ (Integration (or) Difference)

eg:- (Day term) ⇒ 40 42 45 48 50 — 2 3 3 2 — pattern — Difference is too low

Suppose taking I = 1 (lag = 1)

eg:- (month term) → 25 30 40 35 ⇒ Difference is too High

In ARMA :- we are not going to conclude this "I", here we are not going to conclude this difference
Because we are not able to find out any pattern between a time stamp, not able to find out any pattern between Data (or) Difference

ARIMA Equation

↳ $y(t) = a + w^* y(t-1) + e$ (Generalize form Equ)

- **Model Evaluation**

(1) Akaike Information Criterion (AIC) ↳ Lowest AIC Value that will be my best model

→ $AIC = -2^* \ln(\text{Likelihood}) + 2^* K$,

→ $AIC = \frac{-2}{N^* LL} + \frac{2}{K/N}$

⇒ $AIC = -2/N^* LL + 2^* K/N$

ACF | PACF

(2) Bayesian Information Criterion. (BIC)

$\rightarrow BIC = -2 * LL + \log(N) * K$ that ↳ Lowest BIC will be my best ARIMA model

---

Blog ⇒ Towards Data Science → arima simplified ← (search)

- Auto-correlation (ACF) :- which consider both direct and indirect effects
  - ACF → is also called corrlg graph
- Partial Auto-correlation (PACF) :- which consider only the direct effects.

$\{ARIMA \xrightarrow[\text{3 model}]{\text{Integrating}} Y_t = AR + MA + I\}$

⇒ Rolling | window Concept

Data :- ↳

350
400
450
480
470
420
370
320

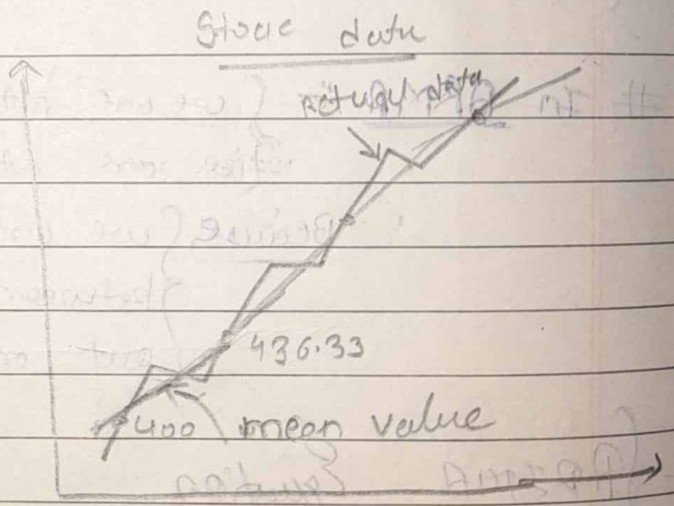
window = 3

mean = $\frac{350 + 400 + 450}{3}$

= 400

mean = $\frac{400 + 450 + 480}{3}$

= 436.33



I'll chek Mean ↑

This concept is Totally related to Rolling st

what is Rolling Stats → Rolling stats means → we are going to role ↳ role over the WINDOW

- window ⇒ eg:- suppose i am taking window = 3

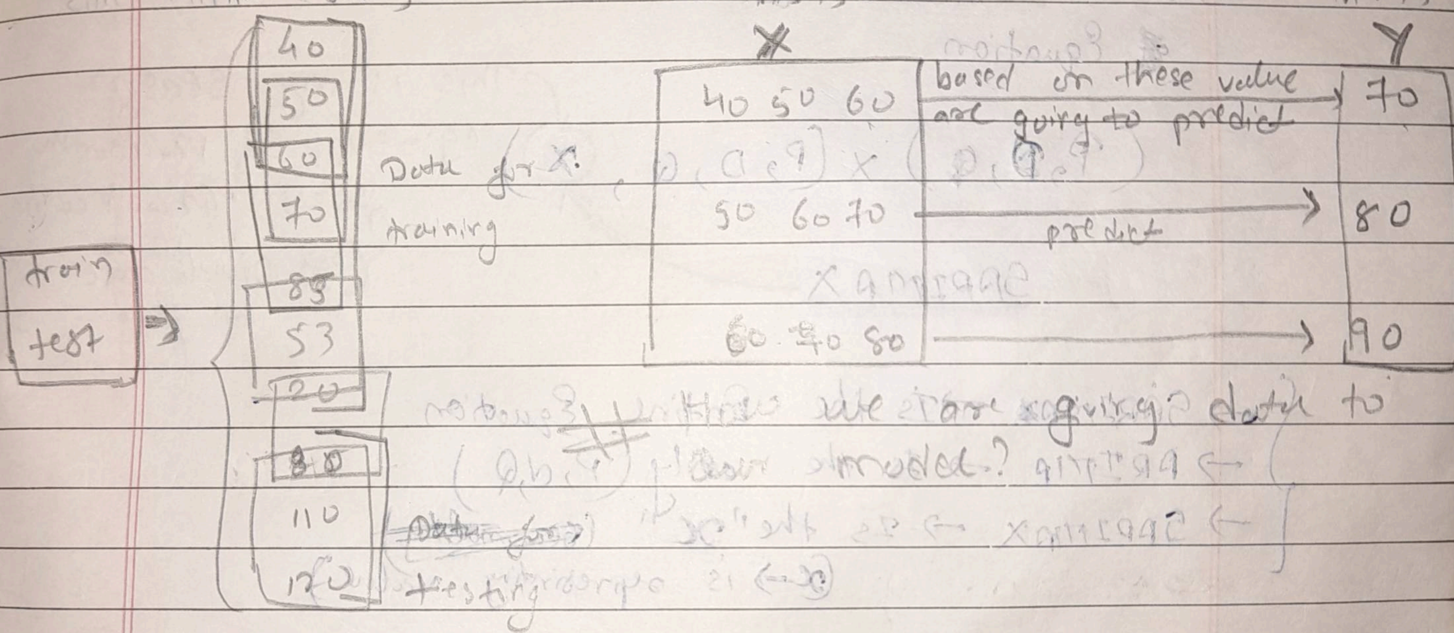
Decompose the time Series :

→ we will try to check the trend in the ~~trea~~ data and will breakdown the series inside, season, cycle and error

---

while handling with main Data set in Jupyter Notebook

↳ How we are going to do train-test split in Time series ?



80 110 120 ⟶ ? { one step ahead w.r.t these data i.e (our dataset)

→ **One step ahead prediction**

⇒ The get_prediction and conf_int methods calculate predictions for future points in time for the previously fitted model and the confidence interval associated with a prediction, respectively.

⇒ The dynamic = false argument causes the method to produce a one-step ahead prediction of the time series.

eg

↳ 10 data point $\xrightarrow[\text{have to predict}]{\text{is given we}}$ 11th