\* When we want to predict something related to time, we use time series forecasting.

\* In time series we cannot use machine learning.

\* Regression means linear regression.

\* We study about many projects on time series, that project you show as it is in interviews and explain to interviewers.

\* SEBI hired data scientists, if you get the opportunity to work in SEBI, accept it.

\* Time series analysis isn't needed everywhere; it's mainly used by companies that deal with data changing over time, like brokerage firms or businesses that make products.

\* In next class we discuss 2 jupyter notebook, in one notebook we predict stock price and, in another notebook, we predict how many beer bottles / cans to produce.

\* We also study about 2 concepts ACF (Auto correlation) and PACF (Partial Auto Correlation).

\* In ACF & PACF we check the correlation between previous values which we take for analysis.

There are 2 types of Moving Average: -

(I) Simple Moving Average (SMA): Simple moving average give equal weightage to all previous values.

(ii) Exponential Moving Average (EMA): Exponential moving average gives more weightage and importance to the most recent data. It is also known as exponentially weighted moving average.

Example: When any company declares his result, then next day its stock price will fluctuate. In this case we have to use exponential moving average because in this case model gives more importance and weightage to the most recent data.

\* Parameter for moving average is ‘ma’. In this parameter we have define that how many previous values to take and compute their average.

\* ARIMA: Auto Regressive Integrated Moving Average

\* ARMA: Auto Regressive Moving Average

\* SARIMA: Seasonality Auto Regressive Integrated Moving Average

\* While using these we have to define a parameter named as ‘p’ which means how many previous values to use in autoregression.

\* Lag: How many previous values to use in autoregression.

\* If data is more than we cannot use full data because it will increase the noise / residual in data.

\* In time series we have 3 different components: -

(I) Seasonality: Specific type of variations which repeat over time.

\* Seasonality may occur day wise, month wise or year wise.

Example: - In malls from Monday to Friday, footfall is low but on weekend it is high.

\* Traffic is lighter during working hours and at night, but it is heavier in the morning and evening when people are commuting to and from work.

\* During summers sales of A.C & Refrigerators are high and in winters sales is low.

(II) Trend: Overall pattern of data

(III) Random Variations / Residual: Variations which do not comes in seasonality & trends are consider as random variations.

\* Decomposition: It means to separate all components from data. The parameter for decomposition is ‘seasonal\_decompose’.

\* Stationary Data: When mean and variance / std. deviation of data is constant or almost same, not more than 10%. It means mean and variance of all data points are constant.



\* In real time data is non-stationary.

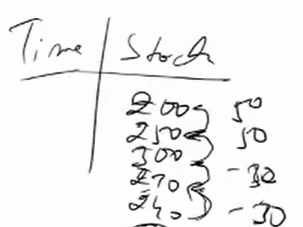
\* Non-Stationary Data: When mean and variance / std. deviation of data is not constant.



\* This is a sales example of any random company. Here, the mean of starting point and any random point at middle is not same.

\* We cannot do analysis on non-stationary data. We first need to make it stationary.

\* Do convert non-stationary data into stationary we have ‘i’ parameter. This parameter means differencing.



\* We do differencing in this way. 200-250=50, 250-300=50, 300-270=-30, 270-240=-30.

\* After every level differencing we check that data becomes stationary or not. If data not becomes stationary then again do differencing, we do that until data becomes stationary.

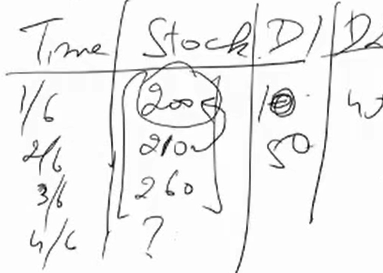


\* When data becomes stationary then after we further provide data to model.

\* When data is non-stationary we do multiple level differencing. If after doing multiple level differencing data doesn’t convert into stationary, then we do transformations such as logarithmic, square root, cube root, exponential.

\* After applying transformation we again do differencing, and there are high chances that after transformation and differencing data become stationary.

\* There is a problem in differencing, when we apply differencing, every time one records getting less in this way.



There are two ways by which we can check data is stationary or not: -

(I) Basic way: Compute mean & Variance / std. deviation of min. and max. value of data and compare their mean & variance if the difference of both is not more than 10%, it means data is stationary.

(II) Ad Fuller Test (Augmented Dicky Fuller Test): If P value of this test is less than 0.05, it means data is stationary.

\* Internally this test defines H0 (H not) = Data is not stationary & Ha = Data is stationary.

\* If p-value ≤ chosen significance level (e.g., 0.05), reject the null hypothesis.

\* If p-value > chosen significance level, fail to reject the null hypothesis.

Q) How to apply ARIMA?

1) ‘p’: means how many previous values to use in autoregression.

2) ‘d’: To convert non-stationary data into stationary we have ‘i’ parameter but in ARIMA we use ‘d’. This parameter means differencing / Integration.

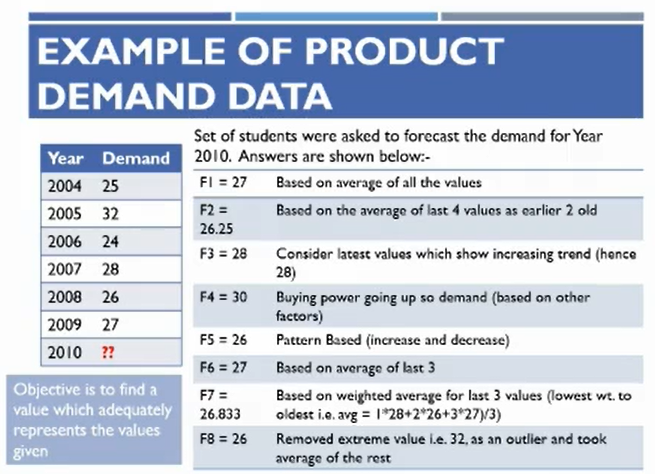
3) ‘q’: In this parameter we have define that how many previous values to take and compute their moving average.

\* In this if we don’t take ‘d’ value, it becomes ARMA.

\* In this if we take seasonality component, it becomes SARMIA.

\* In SARMIA if we add external factors or variations like covid 19 recession or any unexcepted thing, it will become SARIMAX.

Base Line Models

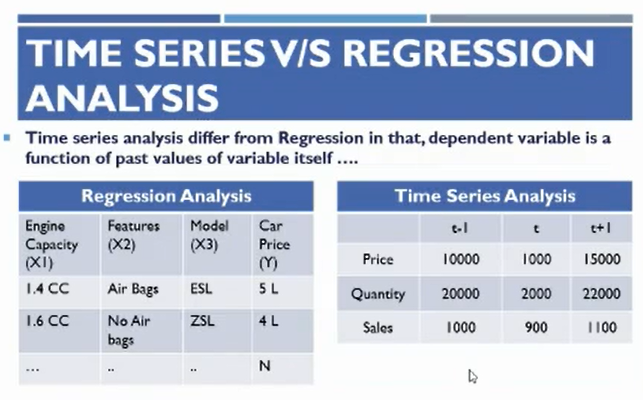


\* If some doesn’t study time series, then these are ways by which someone can do predictions. These techniques are called ‘base line models’

\* We can use these techniques as base line models.

\* If model perform good then it will be approved.

\* After model building through time series we can compare our model’s performance with these techniques, which one can perform better we can select that one. Management denies to approve time series model because base line model performs better.



\* In this image t is present value, (t-1) previous values and (t+1) next value which we want to forecast.

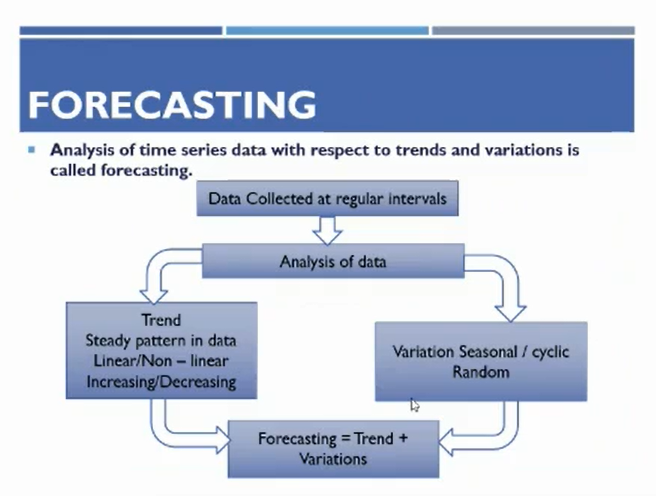
\* (t-n) presents lags, means how many previous values to use for forecasting and (t+n) represents future values or forecasted values.

Q) Difference between seasonality & cyclic variations?

Duration of seasonality is less like 1 day, 1 week, 1 month, 1 year and it repeats over time like everyday traffic is more in morning & evening, sales of cooling appliances are more during summers and low during winters, etc these type of variations & patterns repeats over time. Variations of seasonality are same like summer came every year in same month, every day traffic goes high at same time.

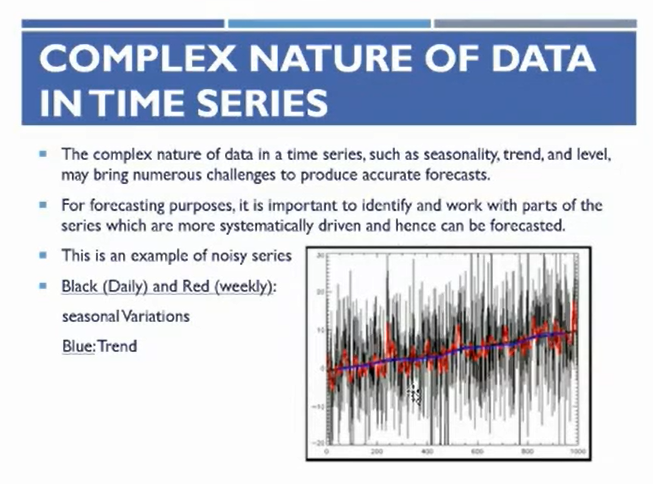
Duration of Cyclic variations are long term like 5 years, 10 years. Example: Recession, last recession came in 2008 after then in 2020, it means their time is not fixed. Usually, cyclic variations are long term variations and their time is not fixed. In cyclic variation, there is no seasonality usually of longer period.

\* Random variations / noise: These are the variations which no one can predict. For example: - Covid, no one can predict that covid came and stock price were fell. This is called noise. Anyhow earthquake come, no one predict that. If CBI raid in random company and found any illegal documents then suddenly stock price of that particular company fell down, this is the case that no one can predict. These types of unpredictable variation due to any cause is known as random variations / noise. These variations are cannot comes under seasonality, trend & cyclic variations.



\* Always remember that time series data is in equal time interval.

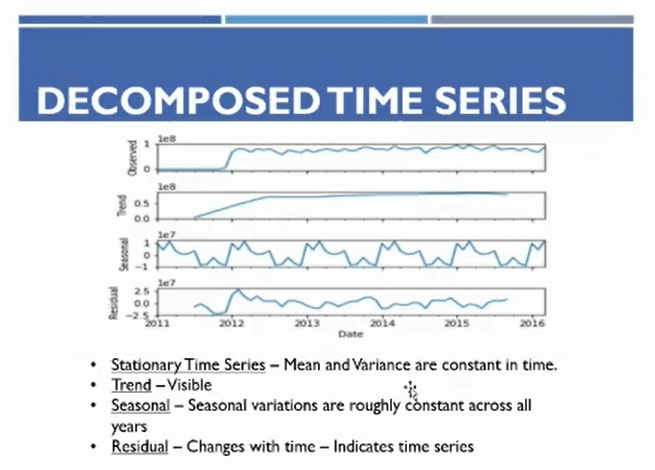
\* Time duration between all pairs of data points should be same then after model perform good. If there are variations in data then model unable to capture pattern correctly which results in model doesn’t perform better.



\* We saw on trading view platform that how stock price fluctuates and we also saw 2 candles green and red. Green candles represent that today stock price is increase by yesterday and red candle represents that today stock price closed lower price than yesterday.

\* During trading hours stock price fluctuates every minute.

Decomposition

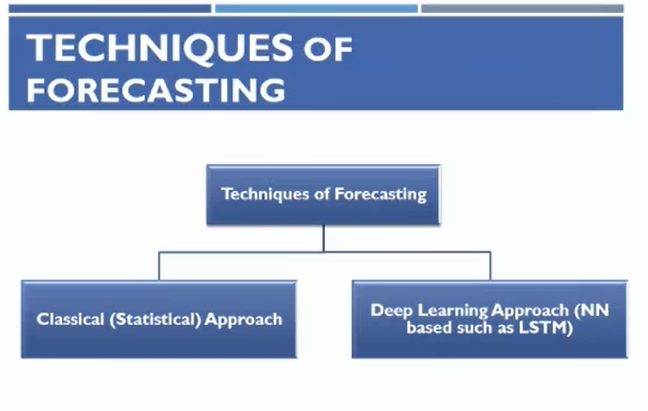


\* Separate all components from data.

\* Trend in 2012 is upward and after 2012 trend is constant, it means company doesn’t show any growth and it stables their performance.

\* Clearly, seasonality is there in data.

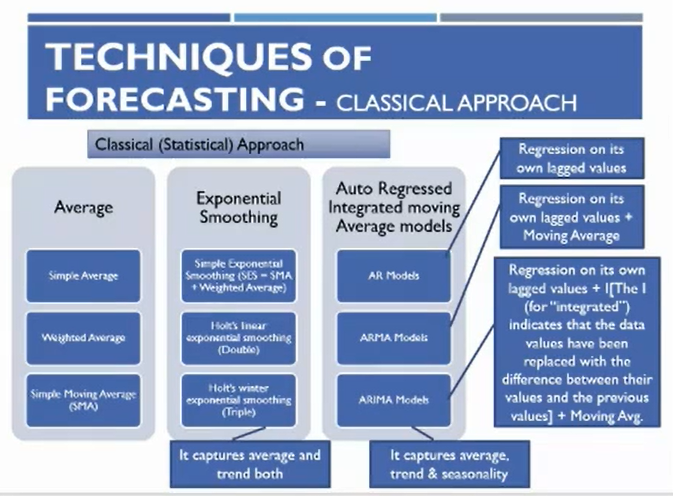
Forecasting Techniques



\* Classical Approach: ARIMA, AR, ARMA, SARIMA, SARIMAX, SMA, EMA

\* In Deep Learning approach we use neural network such as LSTM (Long Short-Term Memory), which is an advance version of RNN.

\* Simple RNN is not used because it forgets previous data and only member recent data.



\* Average means simple moving average.

\* Moving Average is every important thing in trading as well as in time series.



\* These 3 lines are moving averages. These moving average lines are used in trading for buying and selling opportunities. Black line represents 200 days moving average. Red line represents 50 days moving average and green light represents 20 days moving average.

\* It is generally said that when any share price is nearby his 200 days moving average or below it, purchase that share. On the basis of moving average, we can make the trading calls.

In this way, we calculate Simple Moving Average

