TIME SERIES

- **Sequence** – Arrangement (Order) of pattern.
- **Pattern** – It is organized structure and it is repeated occurrence of similar things.
- **Series** – Arrangement of sequence and many sequence constitute together along with patterns to form a series.
- A time series is a series of data points indexed or listed or graphed in time order.
- In other words, the arrangement of data in accordance with their time of occurrence is a time series. It is the chronological arrangement of data. Time series is univariate analysis(navie,ARIMA,SARIMA).
- Equal interval and time is called Time Series

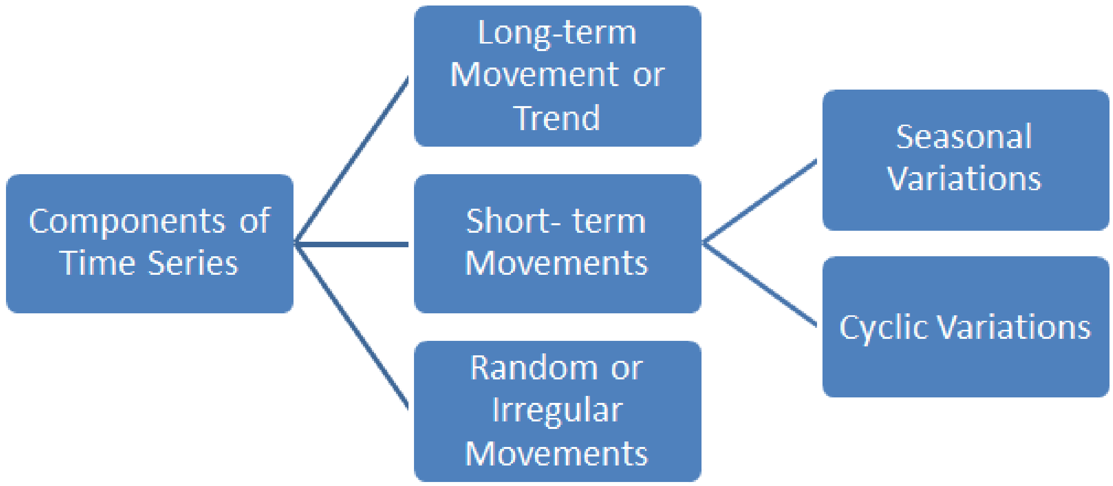
- **Uses of Time Series**
- The most important use of studying time series is that it helps us to predict the future behavior of the variable based on past experience
- It is helpful for business planning as it helps in comparing the actual current performance with the expected one
- From time series, we get to study the past behavior of the phenomenon or the variable under consideration
- We can compare the changes in the values of different variables at different times or places, etc.
- A time series plot make between data (y-axis) and time period (x-axis).

- **Assumptions in Time Series**
- Time series data should be equally spaced over time
- Patterns of past data will propagate into future
- Cannot be used to predict random events (i.e. next tsunami etc.)

- **What is Time Series Forecasting?**
- Time series forecasting is the use of a model to predict future values based on previously observed values.

TIME SERIES

- Components of Time Series



-
- **NOTE :-**
- **alpha = level**
- **Beta = Trend**
- **Delta/Gamma = Seasonality**

- **Trend**
- The trend shows the general tendency of the data to increase or decrease during a long period of time. A trend is a smooth, general, long-term, average tendency. It is not always necessary that the increase or decrease is in the same direction throughout the given period of time.
- It is observable that the tendencies may increase, decrease or are stable in different sections of time. But the overall trend must be upward, downward or stable.
- Taking averages over a certain period is a simple way of detecting trend in seasonal data.
- Change in averages with time is evidence of a trend in the given series, though there are more formal tests for detecting trend in time series

- **Seasonal variation**
- These are the rhythmic forces which operate in a regular and periodic manner over a span of less than a year. They have the same or almost the same

pattern during a period of 12 months. This variation will be present in a time series if the data are recorded hourly, daily, weekly, quarterly, or monthly.

- The seasonal variations the time period should not exceed one year.

- **There are three reasons for the study of seasonal variations:**
- We can establish the pattern of the past changes,
- The projection of past patterns into the future is a useful technique of prediction,
- The effects of seasonality can be eliminated from the time series after their presence is established.

- **Types of seasonal variation**
- **Seasonal variations due to natural forces**: There are variations in time series due to weather conditions and climatic changes. For Example: Sales of umbrellas zoom up during rainy season, Sale of ice cream zoom up in summer.
- **Seasonal variations due to customs**: There are variations due to customs, habits, lifestyle, and conventions of the people in society. For Example: Sale of sweets goes up during festival.

- Seasonal variations that repeat over a specific period such as a day, week, month,season,etc.,

Cyclic Variations

- The variations in a time series which operate themselves over a span of more than one year are the cyclic variations. This oscillatory movement has a period of oscillation of more than a year. One complete period is a cycle. This cyclic movement is sometimes called the ‘Business Cycle’.
- It is a four-phase cycle comprising of the phases of prosperity, recession, depression, and recovery. The cyclic variation may be regular are not periodic. The upswings and the downswings in business depend upon the joint nature of the economic forces and the interaction between them.
- Cyclical variations that correspond with business or economic 'boom-bust' cycles or follow their own peculiar cycles,

- **Irregular** – This component is unpredictable. Every time series has some unpredictable component that makes it a random variable. In prediction, the objective is to "model" all the components to the point that the only component that remains unexplained is the random component.

- **Smoothing Time Series**
- Smoothing is usually done to help us better see patterns, trends for example, in time series. Generally smooth out the irregular roughness to see a clearer signal. For seasonal data, we might smooth out the seasonality so that we can identify the trend. Smoothing doesn't provide us with a model, but it can be a good first step in describing various components of the series.
- Smoothing is a statistical technique for removal of short term irregularities in a time-series data to improve the accuracy of forecasts.
- The process of removing all these jumping around in the data is called smoothing.
- In demand forecasting, we use smoothing to remove random variation (noise) from our historical demand.
- **The most common way to remove noise from demand history is to use different statistical methods to normalize the effect of time series component:**
- **Simple Moving Average**
- **Weighted Moving average.**
- **Simple Exponential Smoothing**
- **Holts Exponential Smoothing/ Double Exponential Smoothing**
- **Winters Exponential Smoothing/Holts Winter Smoothing/ Triple Exponential Smoothing**
- **Simple Moving Average: -** A simple moving average (SMA) is an arithmetic moving average calculated by adding the number of time periods and then dividing this total by the number of time periods.
- **Moving averages** are a simple and common type of smoothing used in **time series** analysis and **time series** forecasting.
- It is average of data but we are decided the time period which take a average of data under the time period and it is used to identify trends and the potential for a change in an established trend.
- The longer the timeframe for the moving average, the smoother the simple moving average.
- A short term moving average is more volatile, but its reading is closer to the source data

TIME SERIES

- Note: - One problem is facing that After a long time period , moving average behave as a constant.
- **Disadvantages of Simple Moving Average**
- Requires saving lots of past data points: at least the N periods used in the moving average computation
- Lags behind a trend
- Ignores complex relationships in data

- **Weighted Moving average: - The weighted moving average (WMA) is a technical indicator that assigns a greater weighting to the most recent data points, and less weighting to data points in the distant past. The WMA is obtained by multiplying each number in the data set by a predetermined weight and summing up the resulting values.**
- the weighted moving average to determine trend direction is more accurate than the simple moving average.

- **<u>Simple Exponential Smoothing (SES)</u>**
- SES assumes that the time series data is made up of two components, a level and some error around that level.
- Level = average(mean) of the data.
- Demand at level t = level + error around the level at time t
- How do we calculate the level?
- $Level_0$ = average of the N period of Actual value(A).
- $Level_1 = Level_0$ + (alpha or smoothing coefficinet)*($A_1 - Level_0$)
- $Level_2 = Level_1$ + (alpha or smoothing coefficinet)($A_2 - Level_1$)

- $Level_t = Level_{t-1}$ + (alpha or smoothing coefficinet)*($A_t - Level_{t-1}$)
- Error = $A_t - Level_{t-1}$
- If you have 36 months of data, the forecast in time period 37 would be $Level_{36}$, and the forecast for time period 40 would be the same, $Level_{36}$.

- **Holt's Trend Corrected Exponential Smoothing**
- This method extends the simple exponential smoothing technique to create forecast from data that has a linear trend.
- Demand at level t = level + t*trend + error around the level at time t

TIME SERIES

- If you have 36 months of data, the forecast in time period 37 would be $Level_{36}$ + 1 month of the trend, and the forecast for time period 40 would be, $Level_{36}$. + 4 months of the trend.
- We need an initial value of the level and the trend:
- One common way is to plot half the data that is given. The slope of the line is $trend_0$, and the intercept is $level_0$.
- $Level_1 = Level_0 + Trend_0 + alpha * (Demand_1 – (Level_0 + Trend_0))$
- So,
- $Level_{current\ period} = Level_{previous\ period} + Trend_{previous\ period} + alpha * (Demand_{current\ period} – (Level_{previous\ period} + Trend_{previous\ period}))$

- $Trend_1 = Trend_0 + gamma * alpha * (Demand_1 – (Level_0 + Trend_0))$
- So,
- $Trend_{current\ period} = Trend_{previous\ period} + gamma * alpha * (Demand_{current\ period} – (Level_{previous\ period} + Trend_{previous\ period}))$

- **Holt Winter's Exponential Smoothing**
- It is the logical extension of Holt's Trend Corrected Smoothing. It accounts for a level, a trend, and the need to adjust the demand up or down on a regular basis due to seasonal fluctuations.

- Demand at time t = (level + t*trend) * seasonal adjustment for time t * irregularity adjustments we can't account for
- So, as above, if we have 36 months of data,
- Forecast for month 39 = ($level_{36}$ + 3 * $trend_{36}$) * $seasonality_{27}$
- $Level_1 = Level_0 + Trend_0 + alpha * (Demand_1 - (Level_0 + Trend_0) * Seasonality_{-11})/Seasonality_{-11}$
- $Trend_1 = Trend_0 + gamma * alpha * (Demand_1 - (Level_0 + Trend_0) * Seasonality_{-11})/Seasonality_{-11}$
- $Seasonality_1 = Seasonality_0 + delta * (1- alpha) * (Demand_1 - (Level_0 + Trend_0) * Seasonality_{-11})/(Level_0 + Trend_{0})$

- **Autocorrelation** : - the linear relationship between *lagged values* of a time series and we can also said that relationship between current value and past value.

TIME SERIES

- **Random Walk or Drunken Walk or White Noise** : - Time series that show no autocorrelation means there is no correlation between present and past value
- Note:- Firstly we check the random walk if they present then stop the model and leave the model.
- RMSE Model will be less than RMSE Simple Average but due to present of Random Walk it will show opposite(RMSE Model > RMSE Simple Average).
- For a white noise series, we expect 95% of the spikes in the ACF to lie within $\pm 2/\sqrt{\boxed{T}}\pm 2/T$ where TT is the length of the time series.

- **White Noise** : - Mean = 0 and Standard Deviation = Constant (Important)

- **Stationarity** :- A stationary process has the property that the mean, variance and autocorrelation structure do not change over time.
- Stationarity : - Mean = Constant and STD.Dev = Constant
- We can say that every stationarity is white noise but every white noise is not a stationarity.

- **Unit root tests:-**
- Unit root is a characteristic of a time series that makes it non-stationary.
- **The Dickey-Fuller Test:-** The ADF test belongs to a category of tests called ‘Unit Root Test’, which is the proper method for testing the stationarity of a time series.
- The p-value is very less than the significance level of 0.05 and hence we can reject the null hypothesis and take that the series is stationary.

- **The KPSS Test(**Kwiatkowski-Phillips-Schmidt-Shin):- KPSS test is a statistical test to check for stationarity of a series around a deterministic trend. Like ADF test, the KPSS test is also commonly used to analyse the stationarity of a series. However, it has couple of key differences compared to the ADF test in function and in practical usage. Therefore, is not safe to just use them interchangeably.
- if p-value is < signif level (say 0.05), then the series is non-stationary. Whereas in ADF test, it would mean the tested series is stationary.

- **what is a ‘deterministic trend’?**

TIME SERIES

- Ans:- The word ‘deterministic’ implies the slope of the trend in the series does not change permanently. That is, even if the series goes through a shock, it tends to regain its original path.
- **The Zivot and Andrews Test**

- **ACF(Autocorrelation function):-** the correlation between the observation at the current time spots and the observation at the previous time spots.
- **PACF(Partial Autocorrelation function):**- the correlation between the observation at two time spots given that we consider both observation are correlated to observation at other time spots .EX- today price are correlated with day before yesterday .

- **White Noise:-** A series purely random in nature.
- Mean = 0,variance=constant and uncorrelated
- Average is best forecast of this series.

- **Auto Regressive model(AR):-** Yt depends only the past value Yt-1,Yt-2,etc
- Yt = f(Yt-1,Yt-2,..)
- Yt = B0 + B1Yt-1 + …
- Auto Regressive as the name implies we regress to itself with the time lag.

- **Moving Average model(MR):-** Yt depends only the random error terms
- Yt = f(Et,Et-1,…) or Yt = B+ Et + O1Et-1+ ..

- **Autocorrelation** : - Autocorrelation is the similarity between two observationa as a function of time lag between them.

- **<u>Model Selection Criteria</u>**
- **AIC (Akaike Information Critera)** :- The AIC is defined by a simple equation from the sum of square and number of degree of freedom of the two method.AIC = -2 ln(L) + 2k
- **Schwartz Bayesian Criterion (SBC)** = -2 ln(L) + k ln(n)
- where L = likelihood function
- k = number of parameters to be estimated,
- n = number of observations. Ideally,
- **the AIC and SBC should be as small as possible**

TIME SERIES

- **Diagnostic Checking**
- The model that is finally chosen is the one considered best based on a set of diagnostic checking criteria. These criteria include
    - t-tests for coefficient significance
    - residual analysis
    - model selection criteria

NOTES: -

- All stationary AR processes exhibit autocorrelation patterns that "die down" to zero as k increases.
- The autocorrelation coefficient of a **non-stationary** AR process is always 1 for all values of k.
- MA processes are always **stationary** with autocorrelation functions that cut off after certain lags.
- **Random walk**: autocorrelation remains at one for all values of k.

- **There are two methods for the time series analysis:**
- **Frequency Domain Method**
- Frequency domain models are based on the idea that time series can be represented as a function of time using sines and cosines. These representations are known as Fourier representations. Frequency domain models utilize regressions on sines and cosines, rather than past and present values, to model the behavior of the data.
- It includes wavelet analysis and spectral analysis.

- **Time Domain Method**
- The time-domain approach models future values as a function of past values and present values.E.g. ARMA, ARIMA, VAR
- It includes cross-correlation and autocorrelation.

- **The time series analysis technique can be further divided into the following:**
- **Parametric: -**The process assumes that the underlying stationary process follows a structure that can be explained in small parameters.
- **Non- parametric: -**It estimates the covariance instead of assuming any structure.The time series analysis can also be classified into linear, non-linear, univariate, and multivariate.

TIME SERIES

- **Univariate versus Multivariate Time Series Models**
- Univariate time series models are models used when the dependent variable is a single time series. E.g. SRIMA
- Multivariate time series models are used when there are multiple dependent variables. In addition to depending on their own past values, each series may depend on past and present values of the other series. E.g. VAR.
-

  Time Series Modeling
- There are different models of time series analysis to bring out the desired results:
- **ARIMA Model**
- ARIMA stands for Autoregressive Integrated Moving Average model, which is a type of regression analysis that measures the influence of one dependent variable corresponding to changing variables. The model is used to forecast moves in the financial market, analyzing the differences in values in a series rather than the actual values.

- **ARIMA can be classified into three components:**
- **AR -** stands for Autoregression, where the dependent relationship is used between observation and many lagged observations.
- **I -** stands for Integrated, where raw observation is differenced and is used to make the time series stationary.
- **MA -** or the Moving Average, use the dependency between observation and residual error.

- Each component is defined as a parameter which is substituted as integers to indicate the usage of the ARIMA model.
  Following are the parameters:
- **d**: - denotes the degree of differencing or the number of times the raw observations are differenced.
- **p**: - denotes lag order or the number of lag observations.
- **q**: -denotes the order of moving average or the size of the moving average window.

- **ARIMA and Stationarity**

TIME SERIES

- A stationary model is when there is consistency in data over a period. ARIMA model makes the data stationary by differencing. For example, most of the economic data reflects a trend. Differencing of data removes the trends to make it stationary.
- Below is an example of monthly index values that are analyzed monthly. The plot suggests that the data is non-stationary, showing an upward trend. Therefore, the ARIMA model can analyze and forecast and make the data stationary:
- **Autoregressive Model (AR)**
- The Autoregressive (AR) model forecasts the future, deriving the behavioral pattern from the past data. It is useful when there is a correlation between the data in a time series. The model is based on the linear regression of the data in the current time series against the previous data on the same series.
- **Moving Average Model (MA)**
- Moving the average process is used to model the univariate time series. The model defines that the output variable is linearly contingent on present and the past data of a time series. It uses past errors in the forecast in a regression instead of the past value of the forecast variable.