1) Autoregressive Integrated Moving Average (ARIMA) is one of the most popular technique for time series modeling. This is also called Box-Jenkins method, named after the statisticians who pioneered some of the latest

2) What is an ARIMA modeling-?

An ARIMA model has following main components. However, not all models need to have all of the below mentioned components.

- Autoregressive (AR)

Value of a time series at time period t (yt) is a function of values of time series at previous time periods 'p'

yt = Linear function of yt-1, yt-2,....., yt-p + error

- Integrated (I)

To make a time series stationary (discussed below), sometimes we need to difference successive observation and model that. This process is known as integration and differencing order is represented as'd' in an ARIMA model.

- Moving Average (MA)

Value of a time series at time period t (yt) is a function of errors at previous time periods 'q'

yt = Linear function of Et-1, Et-2,....., Et-q + error

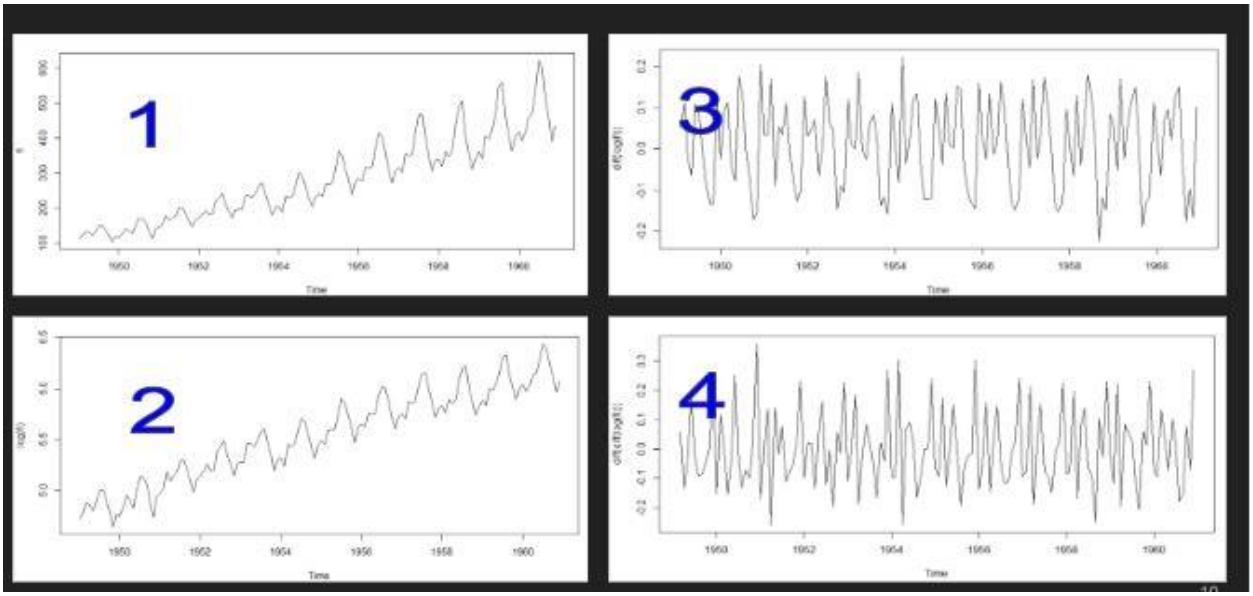
Based on the combinations of the above factors, we can have following and other models-

- AR- Only autoregressive terms
- MA- Only moving averages terms
- ARMA- Both autoregressive and moving average terms
- ARIMA- Autoregressive, moving average terms and integration terms. After the differencing step, the model becomes ARMA

A general ARIMA model is represented as ARIMA (p, d, q) where p, d and q represent AR, Integrated and moving averages respectively. Whereas each of p, d and q are integers higher than or equal to zero.

3) Stationarity of a time series-

A time series is called stationary where it has a constant mean and variance across the time period, i.e. mean and variance don't depend on time. It other words, it should not have any trend and dispersion in variance of the data over a period of time. This is also called white noise.



The main assumption in a time series model is that it is stationary meaning it has constant mean and constant variance across the time period.

From the plot of our air passengers time series, we can tell that the time series is not stationary. Moreover, a time series needs to be stationary or made stationary before being fed into ARIMA modeling.

We can see pic 4 over here where mean is almost constant and variance across the time period is almost constant.to arrive at pic4 from pic1 we may need to do differencing i.e. let us say value of one time period is 80 and value of other time period is 72 and if we do differencing of the time period it is called differencing and we build our model on that.one time it is called differencing for one time, similarly we may do differencing for the second time and so on. Normally differencing for the second time is sufficient enough to arrive at a stationary time series.

Say picture 1st is FL i.e. original time series and we can clearly see in picture one that time series is not stationary i.e. it doesn't have constant mean and constant variance across the time period.

Then we take log (FL) and arrive at picture2.

Then we do differencing of log (FL) i.e. diff (log (FL)) which is d=1 to arrive at picture3.

Then we further do differencing for second time i.e. diff(diff(log(FL))) which is d=2 to arrive at picture4 and here in picture4 we can see that it has constant mean and constant variance across the time period and we can conclude that we made the time series as stationary.

4) ADF TEST FOR TESTING SATIONARITY OF TIME SERIES:

A) Statistically, Augmented Dickey–Fuller test is used for testing the stationarity of a time series. Generally speaking the null hypothesis (H0) is that the series is "Non-Stationary" and the alternative hypothesis (Ha) is that series is "Stationary".

B) If the p statistics generated from the test is less than <0.05 we can reject the null hypothesis. Otherwise, we need to accept the null hypothesis.

C) From the ADF test we can see that the p values is close to 0.78 and which is more than 0.05 and hence we need to accept the null hypothesis that is the series is "Non Stationary"

5) How do we make a time series stationary? Well, we can do it two ways-

- Manual- Transformation and Differencing etc. Let's look at an example.
- Automated- The Integrated term (d) in the ARIMA will make it stationary. This we will do in the model fitting phase. Generally speaking we don't require d>1 to make a time series stationary
- Auto.arima ( ) will take care of this automatically and fit the best model

6) Fit a model, evaluate model's accuracy and forecast

We will use auto.arima ( ) to fit the best model and evaluate model fitment and performance using following main parameters.

7) A good time series model should have following characteristics-

- Residuals shouldn’t show any trends over time.
- Auto correlation Factors (ACF) and Partial Auto correlation Factor (PACF) shouldn’t have large values (beyond significance level) for any lags. ACF indicates correlation between the current values to all the previous values in a range. PACF is an extension of ACF, where it removes the correlation of the intermediate lags.
- Errors shouldn’t show any seasonality
- Errors should be normally distributed
- Error (MAE, MAPE, MSE etc.) should be low
- AIC, BIC should be relatively lower compared to other alternative models.