**Q1.Explain the time series methods used for time series Forecasting.**

* Time series forecasting is a technique for the prediction of events through a sequence of time. It predicts future events by analyzing the trends of the past, on the assumption that future trends will hold similar to historical trends
* Common types include: Autoregression (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving-Average (SARIMA)
* AR, MA, ARMA, and ARIMA models are used to forecast the observation at (t+1) based on the historical data of previous time spots recorded for the same observation. However, it is necessary to make sure that the time series is stationary over the historical data of observation overtime period. If the time series is not stationary then we could apply the differencing factor on the records and see if the graph of the time series is a stationary overtime period.**ACF (Auto Correlation Function)**
* Auto Correlation function takes into consideration of all the past observations irrespective of its effect on the future or present time period. It calculates the correlation between the t and (t-k) time period. It includes all the lags or intervals between t and (t-k) time periods. Correlation is always calculated using the Pearson Correlation formula.

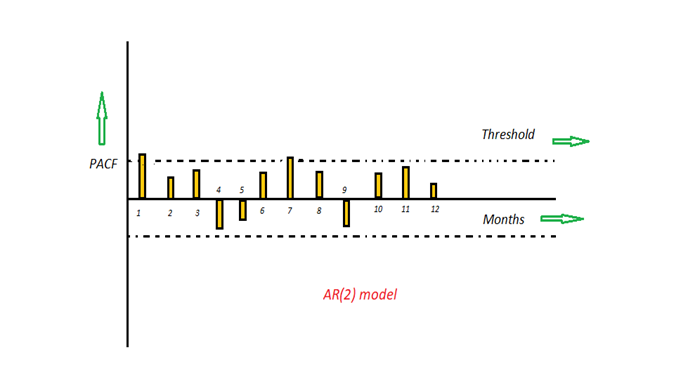
**PACF(Partial Correlation Function)**

* The PACF determines the partial correlation between time period t and t-k. It doesn’t take into consideration all the time lags between t and t-k. For e.g. let's assume that today's stock price may be dependent on 3 days prior stock price but it might not take into consideration yesterday's stock price closure. Hence we consider only the time lags having a direct impact on future time period by neglecting the insignificant time lags in between the two-time slots t and t-k.

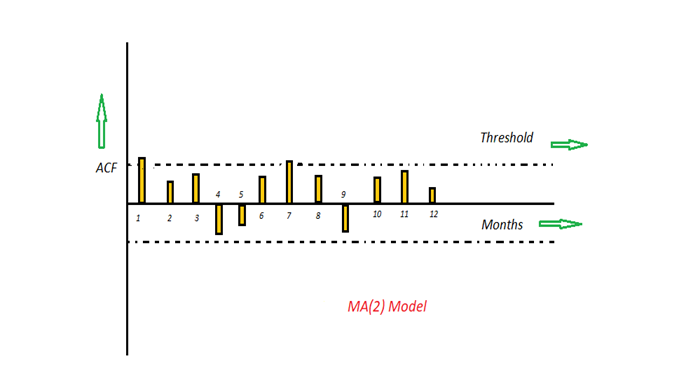
**How to differentiate when to use ACF and PACF?**

* Let's take an example of sweets sale and income generated in a village over a year. Under the assumption that every 2 months there is a festival in the village, we take out the historical data of sweets sale and income generated for 12 months. If we plot the time as month then we can observe that when it comes to calculating the sweets sale we are interested in only alternate months as the sale of sweets increases every two months. But if we are to consider the income generated next month then we have to take into consideration all the 12 months of last year.
* So in the above situation, we will use ACF to find out the income generated in the future but we will be using PACF to find out the sweets sold in the next month.

**AR (Auto-Regressive) Model**

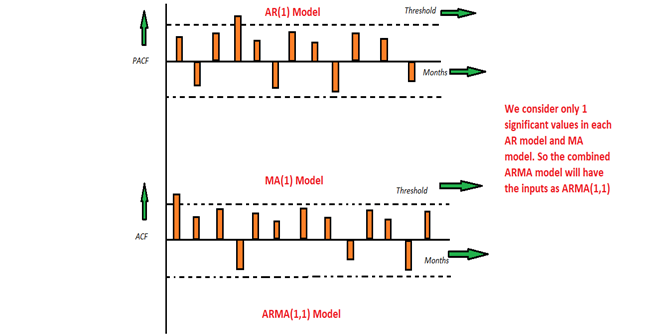
* 
* The time period at t is impacted by the observation at various slots t-1, t-2, t-3, ….., t-k. The impact of previous time spots is decided by the coefficient factor at that particular period of time. The price of a share of any particular company X may depend on all the previous share prices in the time series. This kind of model calculates the regression of past time series and calculates the present or future values in the series in know as Auto Regression (AR) model.
* **Yt = β₁\* y-₁ + β₂\* yₜ-₂ + β₃ \* yₜ-₃ + ………… + βₖ \* yₜ-ₖ**
* Consider an example of a milk distribution company that produces milk every month in the country. We want to calculate the amount of milk to be produced current month considering the milk generated in the last year. We begin by calculating the PACF values of all the 12 lags with respect to the current month. If the value of the PACF of any particular month is more than a significant value only those values will be considered for the model analysis.
* For e.g in the above figure the values 1,2, 3 up to 12 displays the direct effect(PACF) of the milk production in the current month w.r.t the given the lag t. If we consider two significant values above the threshold then the model will be termed as AR(2).

**MA (Moving Average) Model**



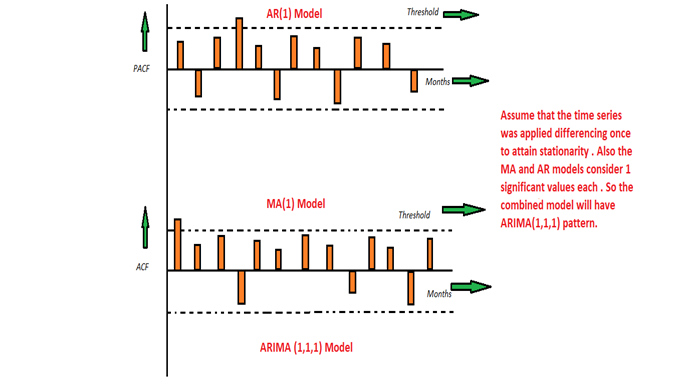
* The time period at t is impacted by the unexpected external factors at various slots t-1, t-2, t-3, ….., t-k. These unexpected impacts are known as Errors or Residuals. The impact of previous time spots is decided by the coefficient factor α at that particular period of time. The price of a share of any particular company X may depend on some company merger that happened overnight or maybe the company resulted in shutdown due to bankruptcy. This kind of model calculates the residuals or errors of past time series and calculates the present or future values in the series in know as Moving Average (MA) model.
* **Yt = α₁\* Ɛₜ-₁ + α₂ \* Ɛₜ-₂ + α₃ \* Ɛₜ-₃ + ………… + αₖ \* Ɛₜ-ₖ**
* Consider an example of Cake distribution during my birthday. Let's assume that your mom asks you to bring pastries to the party. Every year you miss judging the no of invites to the party and end upbringing more or less no of cakes as per requirement. The difference in the actual and expected results in the error. So you want to avoid the error for this year hence we apply the moving average model on the time series and calculate the no of pastries needed this year based on past collective errors. Next, calculate the ACF values of all the lags in the time series. If the value of the ACF of any particular month is more than a significant value only those values will be considered for the model analysis.
* For e.g in the above figure the values 1,2, 3 up to 12 displays the total error(ACF) of count in pastries current month w.r.t the given the lag t by considering all the in-between lags between time t and current month. If we consider two significant values above the threshold then the model will be termed as MA(2).

**ARMA (Auto Regressive Moving Average) Model**



* This is a model that is combined from the AR and MA models. In this model, the impact of previous lags along with the residuals is considered for forecasting the future values of the time series. Here β represents the coefficients of the AR model and α represents the coefficients of the MA model.
* **Yt = β₁\* yₜ-₁ + α₁\* Ɛₜ-₁ + β₂\* yₜ-₂ + α₂ \* Ɛₜ-₂ + β₃ \* yₜ-₃ + α₃ \* Ɛₜ-₃ +………… + βₖ \* yₜ-ₖ + αₖ \* Ɛₜ-ₖ**
* Consider the above graphs where the MA and AR values are plotted with their respective significant values. Let's assume that we consider only 1 significant value from the AR model and likewise 1 significant value from the MA model. So the ARMA model will be obtained from the combined values of the other two models will be of the order of ARMA(1,1).

**ARIMA (Auto-Regressive Integrated Moving Average) Model**



* We know that in order to apply the various models we must in the beginning convert the series into Stationary Time Series. In order to achieve the same, we apply the differencing or Integrated method where we subtract the t-1 value from t values of time series. After applying the first differencing if we are still unable to get the Stationary time series then we again apply the second-order differencing.
* The ARIMA model is quite similar to the ARMA model other than the fact that it includes one more factor known as Integrated( I ) i.e. differencing which stands for I in the ARIMA model. So in short ARIMA model is a combination of a number of differences already applied on the model in order to make it stationary, the number of previous lags along with residuals errors in order to forecast future values.
* Consider the above graphs where the MA and AR values are plotted with their respective significant values. Let's assume that we consider only 1 significant value from the AR model and likewise 1 significant value from the MA model. Also, the graph was initially non-stationary and we had to perform differencing operation once in order to convert into a stationary set. Hence the ARIMA model which will be obtained from the combined values of the other two models along with the Integral operator can be displayed as ARIMA(1,1,1).

**Q2.What is ACF?**

* Correlation between time series with a lagged version of itself. The correlation between the observation at the current time spot and the observations at previous time spots.The autocorrelation function starts a lag 0, which is the correlation of the time series with itself and therefore results in a correlation of 1.
* We will be using the plot\_acf function from the statsmodels.graphics.tsaplots library.
* The ACF plot can provide answers to the following questions:
* Is the observed time series **white noise / random**?
* Is an observation related to an adjacent observation, an observation twice-removed, and so on?
* Can the observed time series be modeled with an **MA model**? If yes, what is the order?
* Import module for ACF

from statsmodels.graphics.tsaplots import plot\_acf

**Q3.What is PCF?**

* Additional correlation explained by each successive lagged term. The correlation between pbservations at two time spots given that we consider both observations are correlated to observations at other time spots.
* The partial autocorrelation at lag k is the autocorrelation between XtXt and Xt−kXt−k that is not accounted for by lags 1 through k−1k−1.
* We will be using the plot\_pacf function from the statsmodels.graphics.tsaplots library with the parameter method = "ols" (regression of time series on lags of it and on constant). (See [statsmodels.tsa.stattools.pacf](https://www.statsmodels.org/dev/generated/statsmodels.tsa.stattools.pacf.html))
* Sidenote: The default parameter for method is yw (Yule-Walker with sample-size adjustment in denominator for acovf). However, this default value is causing some implausible autocorrelations higher than 1 on the sample data. Therefore, we change the method parameter to one that is not causing this issue. ywmle would also work fine as suggested in this [StackExchange post](https://stats.stackexchange.com/questions/380196/what-do-very-high-pacf-values-10-mean)
* The PACF plot can provide answers to the following questions:

Can the observed time series be modeled with an **AR model**? If yes, what is the order?

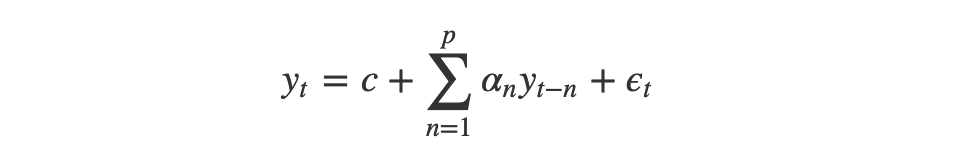
**Q4. What is ARIMA and ARIMAX model?**

**ARIMA**

The ARIMA model stands for "Auto-Regressive Integrated Moving Average" and can be broken down into AR, I, MA.

### **Autoregressive Component - AR(p)**

The autoregressive component of the ARIMA model is represented by AR(p), with the p parameter determining the number of lagged series that we use.



#### **AR(0): White Noise**

* If we set the p parameter as zero (AR(0)), with no autoregressive terms. This time series is just white noise.
* Each data point is sampled from a distribution with a mean of 0 and a variance of sigma-squared.
* This results in a sequence of random numbers that can't be predicted.
* This is really useful as it can serve as a null hypothesis, and protect our analyses from accepting false-positive patterns.

#### **AR(1): Random Walks and Oscillations**

* With the p parameter set to 1, we are taking into account the previous timestamp adjusted by a multiplier, and then adding white noise.
* If the multiplier is 0 then we get white noise, and if the multiplier is 1 we get a random walk.
* If the multiplier is between 0 < α₁ < 1, then the time series will exhibit mean reversion. This means that the values tend to hover around 0 and revert to the mean after regressing from it.

#### **AR(p): Higher-order terms**

* Increasing the p parameter even further is just means going further back and adding more timestamps adjusted by their own multipliers.
* We can go as far back as we want, but as we get further back it is more likely that we should use additional parameters such as the moving average (MA(q)).

### **Moving Average - MA(q)**

"This component is not a rolling average, but rather the lags in the white noise."

#### **MA(q)**

* MA(q) is the moving average model and q is the number of lagged forecasting error terms in the prediction.
* In an MA(1) model, our forecast is a constant term plus the previous white noise term times a multiplier, added with the current white noise term.
* This is just simple probability + statistics, as we are adjusting our forecast based on previous white noise terms.

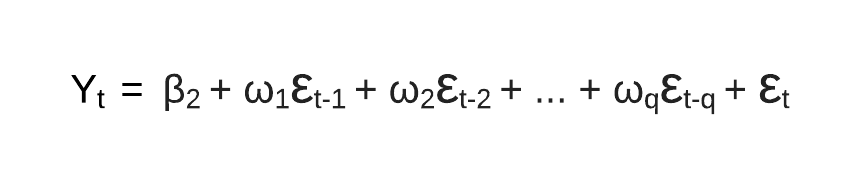
## ARMA and ARIMA Models

ARMA and ARIMA architectures are just the AR (Autoregressive) and MA (Moving Average) components put together.

#### **ARMA**

The ARMA model is a constant plus the sum of AR lags and their multipliers, plus the sum of the MA lags and their multipliers plus white noise. This equation is the basis of all the models that come next and is a framework for many forecasting models across different domains.

#### **ARIMA**



* The ARIMA model is an ARMA model yet with a preprocessing step included in the model that we represent using I(d).
* I(d) is the difference order, which is the number of transformations needed to make the data stationary.
* So, an ARIMA model is simply an ARMA model on the differenced time series.
* **ARIMA** stands for **Autoregressive Integrated Moving Average Model**. It belongs to a class of models that explains a given time series based on its own past values -i.e.- its own lags and the lagged forecast errors. The equation can be used to forecast future values. Any ‘non-seasonal’ time series that exhibits patterns and is not a random white noise can be modeled with ARIMA models.
* So, **ARIMA**, short for **AutoRegressive Integrated Moving Average**, is a forecasting algorithm based on the idea that the information in the past values of the time series can alone be used to predict the future values.

**ARIMA Models** are specified by three order parameters: (p, d, q),

where,

* + p is the order of the AR term
  + q is the order of the MA term
  + d is the number of differencing required to make the time series stationary
* **AR(p) Autoregression** – a regression model that utilizes the dependent relationship between a current observation and observations over a previous period. An auto regressive (AR(p)) component refers to the use of past values in the regression equation for the time series.
* **I(d) Integration** – uses differencing of observations (subtracting an observation from observation at the previous time step) in order to make the time series stationary. Differencing involves the subtraction of the current values of a series with its previous values d number of times.
* **MA(q) Moving Average** – a model that uses the dependency between an observation and a residual error from a moving average model applied to lagged observations. A moving average component depicts the error of the model as a combination of previous error terms. The order q represents the number of terms to be included in the model.

## ****Types of ARIMA Model****

* **ARIMA** : Non-seasonal Autoregressive Integrated Moving Averages
* **SARIMA** : Seasonal ARIMA
* **SARIMAX** : Seasonal ARIMA with exogenous variables

**ARIMAX**

* An Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model can be viewed as a multiple regression model with one or more autoregressive (AR) terms and/or one or more moving average (MA) terms. This method is suitable for forecasting when data is stationary/non stationary, and multivariate with any type of data pattern, i.e., level/trend /seasonality/cyclicity.
* ARIMAX is related to the ARIMA technique but, while ARIMA is suitable for datasets that are univariate (see the article, entitled’ What is ARIMA Forecasting and How Can it Be Used for Enterprise Analysis?’). ARIMAX is suitable for analysis where there are additional explanatory variables (multivariate) in categorical and/or numeric format.
* To understand ARIMAX Forecasting, let’s look at an example of annual GDP values in India. As shown in the figure below, the plot of these data points suggests that this is non stationary data with an upward trend. This dataset is suitable for the ARIMAX algorithm because there is more than one variable affecting the GDP – in other words, the dataset is multivariate.

**Q5.Top five applications of Time Series.**

**1.Time series in Financial and Business Domain**

* Most financial, investment and business decisions are taken into consideration on the basis of future changes and demands forecasts in the financial domain.
* Time series analysis and forecasting essential processes for explaining the dynamic and influential behaviour of financial markets. Via examining financial data, an expert can predict required forecasts for important financial applications in several areas such as risk evolution, [option pricing & trading](https://www.analyticssteps.com/blogs/guide-option-trading-strategies-beginners), portfolio construction, etc.
* For example, time series analysis has become the intrinsic part of [financial analysis](https://www.analyticssteps.com/blogs/introduction-financial-analysis) and can be used in predicting interest rates, foreign currency risk, volatility in stock markets and many more. Policymakers and business experts use financial forecasting to make decisions about production, purchases, market sustainability, allocation of resources, etc.
* In investment, this analysis is employed to track the price fluctuations and price of a security over time. For instance, the price of a security can be recorded;
* For the short term, such as the observation per hour for a business day, and
* For the long term, such as observation at the month end for five years.
* Time series analysis is extremely useful to observe how a given asset, security, or economic variable behaves/changes over time. For example, it can be deployed to evaluate how the underlying changes associated with some data observation behave after shifting to other data observations in the same time period.

### **Time series in Medical Domain**

Medicine has evolved as a data-driven field and continues to contribute in time series analysis to human knowledge with enormous developments.

#### **Case study**

* Consider the case of combining time series with a medical method CBR (case-based reasoning) and data mining, these synergies are essential as the pre-processing for feature mining from time series data and can be useful to study the progress of patients over time.
* In the medical domain, it is important to examine the transformation of behaviour over time as compared to derive inferences depending on the absolute values in the time series. For example, to diagnose heart rate variability in occurrence with respiration based on the sensor readings is the characteristic illustration of connecting time series with case-based monitoring.
* However, time series in the context of the epidemiology domain has emerged very recently and incrementally as time series analysis approaches demand recordkeeping systems such that records should be connected over time and collected precisely at regular intervals.
* As soon as the government has placed sufficient scientific instruments to accumulate good and lengthy temporal data, healthcare applications using time series analysis have resulted in huge prognostication for the industry as well as for individuals’ health diagnoses.

#### **Medical Instruments**

* Time series analysis has made its way into medicine with the advent of medical devices such as
* Electrocardiograms (ECGs), invented in 1901: For diagnosing cardiac conditions by recording the electrical pulses passing through the heart.
* Electroencephalogram (EEG), invented in 1924: For measuring electrical activity/impulses in the brain.
* These inventions made more opportunities for medical practitioners to deploy time series for medical diagnosis.
* With the advent of wearable sensors and smart electronic healthcare devices, now persons can take regular measurements automatically with minimal inputs, resulting in a good collection of longitudinal medical data for both sick and healthy individuals consistently.

### **3.Time Series in Astronomy**

* One of the contemporary and modern applications where time series plays a significant role are different areas of astronomy and astrophysics,
* Being specific in its domain, astronomy hugely relies on plotting objects, trajectories and accurate measurements, and due to the same, astronomical experts are proficient in time series in calibrating instruments and studying objects of their interest.
* Time series data had an intrinsic impact on knowing and measuring anything about the universe, it has a long history in the astronomy domain, for example, sunspot time series were recorded in China in 800 BC, which made sunspot data collection as well-recorded natural phenomena.
* Similarly, in past centuries, time series analysis was used
* To discover variable stars that are used to surmise stellar distances, and
* To observe transitory events such as supernovae to understand the mechanism of the changing of the universe with time.
* Such mechanisms are the results of constant monitoring of live streaming of time series data depending upon the wavelengths and intensities of light that allows astronomers to catch events as they are occurring.
* In the last few decades, data-driven astronomy introduced novel areas of research as astroinformatics and astrostatistics; these paradigms involve major disciplines such as [statistics](https://www.analyticssteps.com/blogs/what-statistics-types-variance-bayesian-statistics), data mining, machine learning and computational intelligence. And here, the role of time series analysis would be detecting and classifying astronomical objects swiftly along with the characterization of novel phenomena independently.

### **Time series in Forecasting Weather**

* Anciently, the Greek philosopher Aristotle researched weather phenomena with the idea to identify causes and effects in weather changes. Later on, scientists started to accumulate weather-related data using the instrument “barometer” to compute the state of atmospheric conditions, they recorded weather-related data on intervals of hourly or daily basis and kept them in different locations.
* With the time, customized weather forecasts began printed in newspapers and later on with the advancement in technology, currently forecasts are beyond the general weather conditions.
* In order to conduct atmospheric measurements with computational methods for fast compilations, many governments have established thousands of weather forecasting stations around the world.
* These stations are equipped with highly functional devices and are interconnected with each other to accumulate weather data at different geographical locations and forecast weather conditions at every bit of time as per requirements.

### **Time series in Business Development**

* Time series forecasting helps businesses to make informed business decisions, as the process analyzes past data patterns it can be useful in forecasting future possibilities and events in the following ways;

* **Reliability:**When the data incorporates a broad spectrum of time intervals in the form of massive observations for a longer time period, time series forecasting is highly reliable. It provides elucidate information by exploiting data observations at various time intervals.
* **Growth:**In order to evaluate the overall financial performance and growth as well as endogenous, time series is the most suitable asset. Basically, endogenous growth is the progress within organizations’ internal human capital resulting in [economic growth](https://www.analyticssteps.com/blogs/difference-between-economic-growth-and-economic-development). For example, studying the impact of any policy variables can be manifested by applying time series forecasting.
* **Trend estimation:**Time series methods can be conducted to discover trends, for example, these methods inspect data observations to identify when measurements reflect a decrease or increase in sales of a particular product.
* **Seasonal patterns:**Recorded data points variances could unveil seasonal patterns & fluctuations that act as a base for data forecasting. The obtained information is significant for markets whose products fluctuate seasonally and assist organizations in planning product development and delivery requirements.