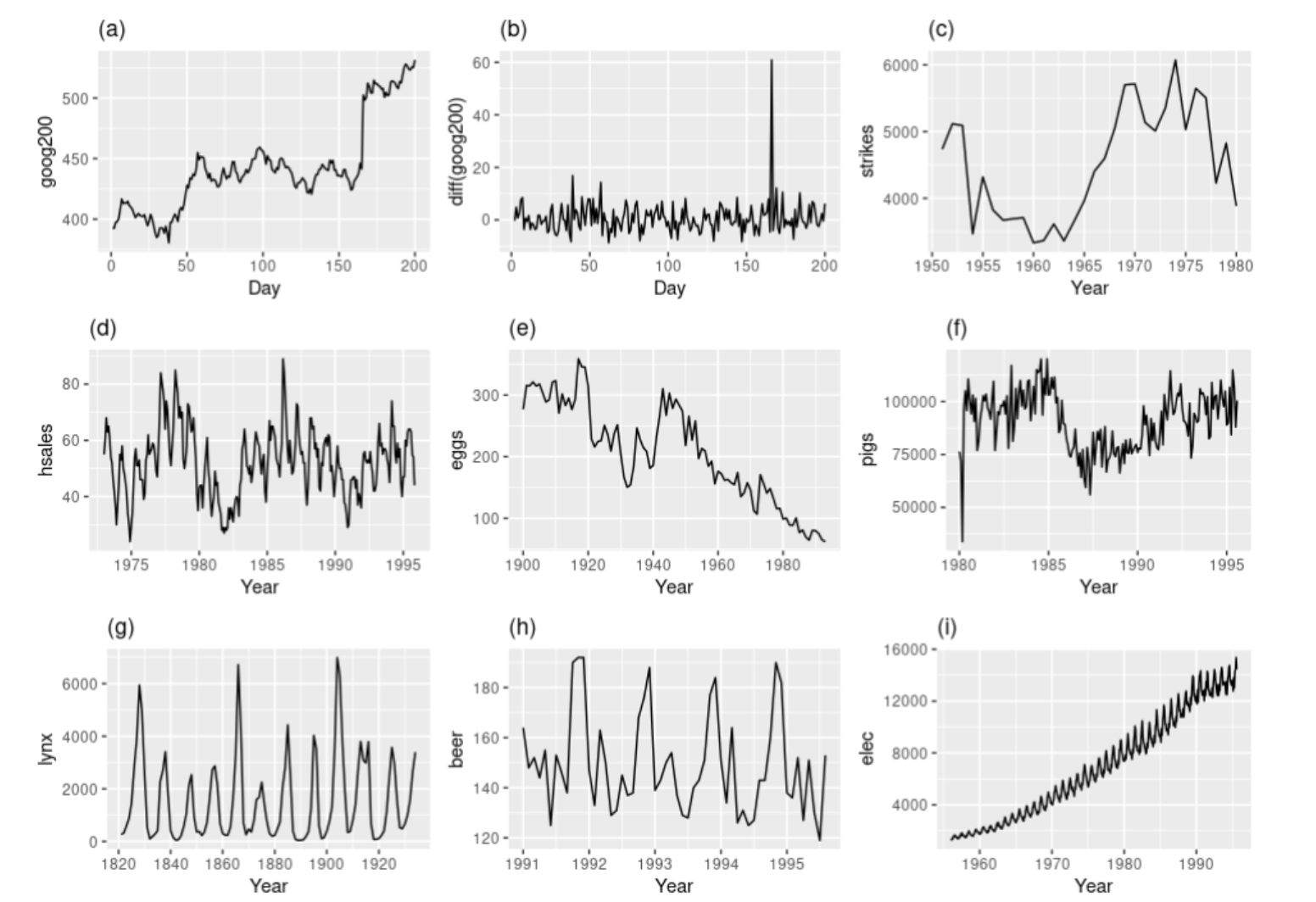
# **Stationarity**

* For a time-series model to be stationary, the parameters (like mean, variance, amplitude, and frequency) of the models should **not be dependent on time**.
* Thus, time series with trends, or with seasonality, are not stationary — the trend and seasonality will affect the value of the time series at different times.
* **For example** Heartbeats (mean=0; standard deviation=1) are stationary — it does not matter when you observe it, it should look much the same at any point in time.
* In general, a stationary time series will have no predictable patterns in the long term.
* If a model's parameter varies over time, then there is a complex relationship that needs to be modeled, which all models are not able to account for.
* Many models assume the series to be stationary to be able to give useful results.
* So, either we want to have a stationary time series, or convert to it. Categorizing a time series by just looking at it can be a little subjective.



* Plots a, c, e, and f, are not stationary they either have a trend or mean changing with time.
* Plots d and h are not stationary as the plots have seasonality
* Plot i is not stationary, it has a trend, seasonality and variance is also not stable.
* Plot b is stationary, there is 1 outlier, and can't say anything about mean seems to be just noise.
* Plot g is stationary, predicting this is dicey, so we assume it to be stationary, and try building a model, and seeing if it performs. Looks like a cyclic time series. But these are not at regular intervals. So, even though there is some seasonality, it can't be predicted.
* There are a few methods to check whether the time series is stationary or not.
  + - Can be checked visually using the time series plot.
    - Statistical test (i.e, Dickey-Fuller Test)

## **Dickey-Fuller Test**

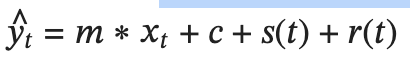
* There is a **statistical** method called the **Dickey-Fuller test**, which is designed for testing for stationarity.
* It fits an auto-regressive model and checks if it worked or not. If it did, then that means it was a stationary time series.
* There is a complicated mechanism to it. We don't need to know how it works. Just need to be aware of this test, as it can be handy.
* For this test,
  + - H0(Null- Hypothesis): The time series is non-stationary.
    - H1(Alternative- Hypothesis): The time series is stationary.
* **How to implement Dickey-Fuller Test?**
  + We can find this as a built-in function under the **statmodels** library as sm.tsa.stattools.adfuller().
* **How do we interpret the result of the Dickey-Fuller test?**
  + This test returns the **p-value**. In order for a time series to be stationary, the **p-value** should be less than 0.05

## **Converting a Non-stationary Time-Series to a Stationary Time-Series**

* As per basic intuition, if we remove trend and seasonality, from our time series, it should become stationary.
* If it's still not stationary, then maybe there's still some seasonality or trend component left in the series.
* Ideally, the trend gets removed in one step. However, seasonality can take multiple steps to be removed.
* These processes are called **Detrending** and **Deseasonalising** respectively.

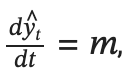
## **De-Trending**

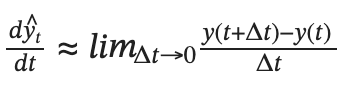
* The trend-seasonality decomposition that a time series can be written as:



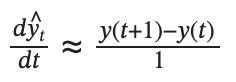
Where; represents the trend component.

* In order to remove this, we can **differentiate** theabove equation with respect to time t. This way, we'll get a stationary time series. This gives us:

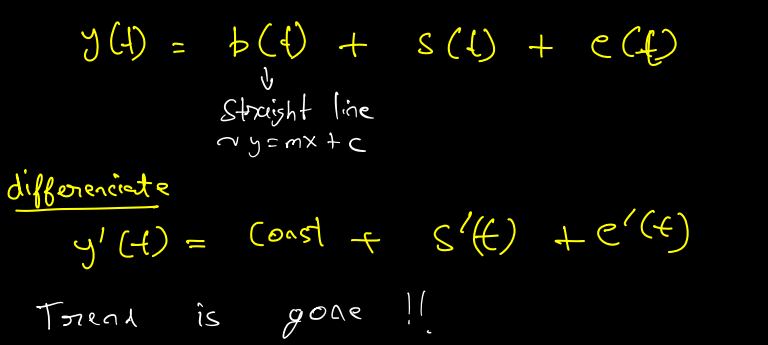




* The minimum value of that we have is 1, since we're talking in terms of time, and our minimum step is 1 month. So the equation becomes:

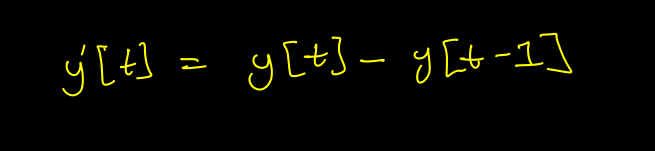


* This process is called **differencing**.
* **NOTE:**
  + Differencing gives a good approximation of differentiating
  + Differentiating gives us a De-trended time series.

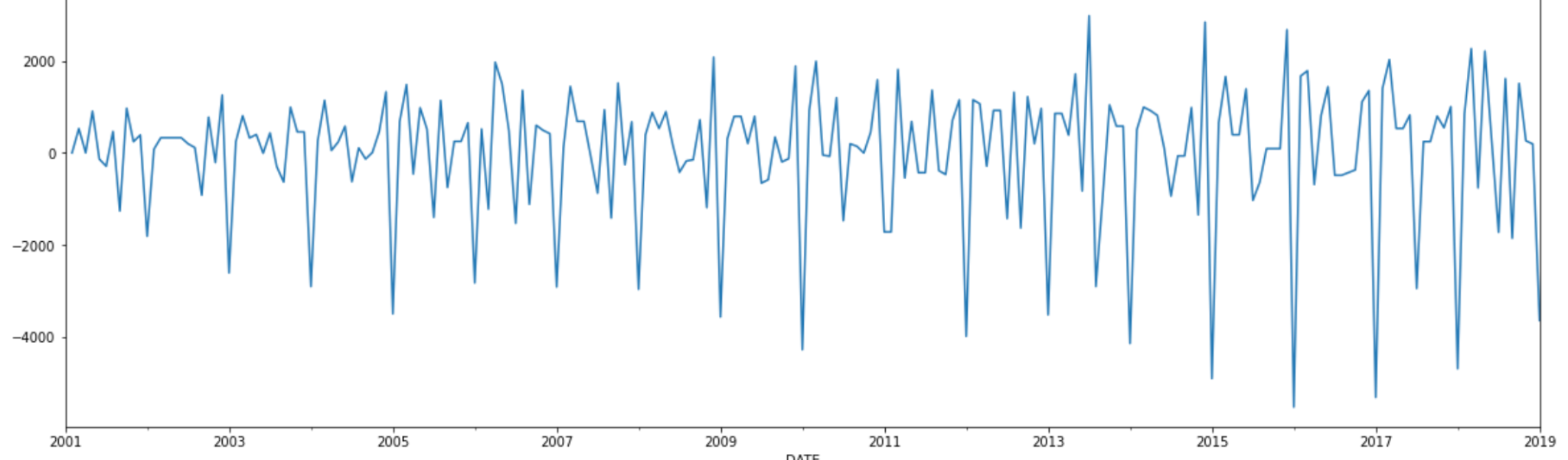


* If the time series has a non-linear trend, then we'll have to differentiate it multiple times, in order to finally achieve a stationary series.
* **NOTE:**
  + If the trend is an exponential function, then we'll not be able to convert it into stationary.
  + This is a very rare case.
  + Some exponential functions can be approximated by polynomials. In that case, we differentiate this polynomial to obtain a stationary time series.
* For applying differencing, a new series is constructed where the value at the current time step is calculated as the difference between the original observation and the observation at the previous time step.

𝑣𝑎𝑙𝑢𝑒(𝑡)=𝑜𝑏𝑠𝑒𝑟𝑣𝑎𝑡𝑖𝑜𝑛(𝑡)−𝑜𝑏𝑠𝑒𝑟𝑣𝑎𝑡𝑖𝑜𝑛(𝑡−1)



* For this, We use the **diff()** method of **pandas**.
* This method calculates the difference between a Dataframe element compared with another element in the Dataframe (default is an element in the previous row, as the default value is 1).

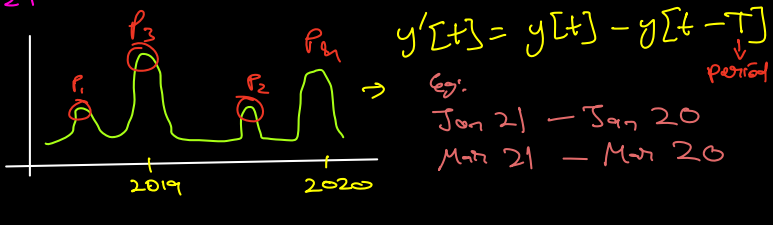


## **De-Seasonalising**

* We can use differencing, but instead of subtracting from the last point, we need to take a difference from the last mth point, where **m** is the period of the seasonality of the series.



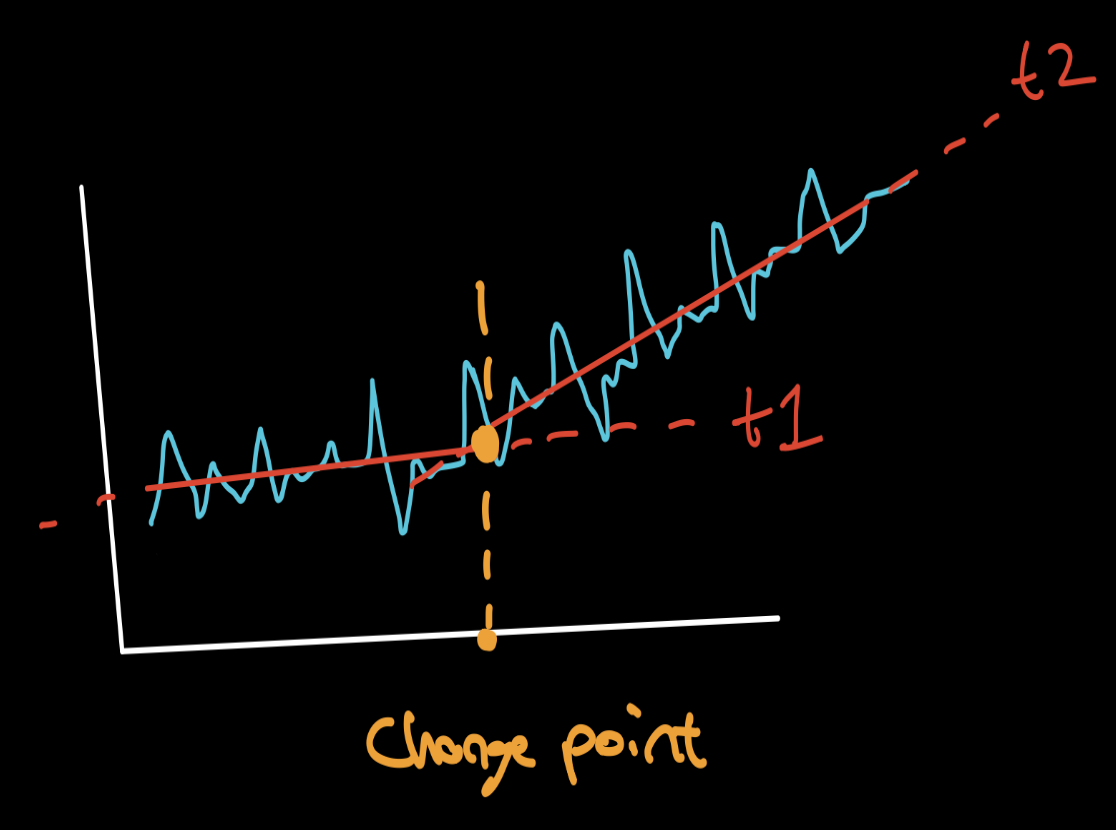
* This is called **m-differencing**.
* If there is a seasonal component at the level of one week, then we can remove it on an observation today by subtracting the value from last week.



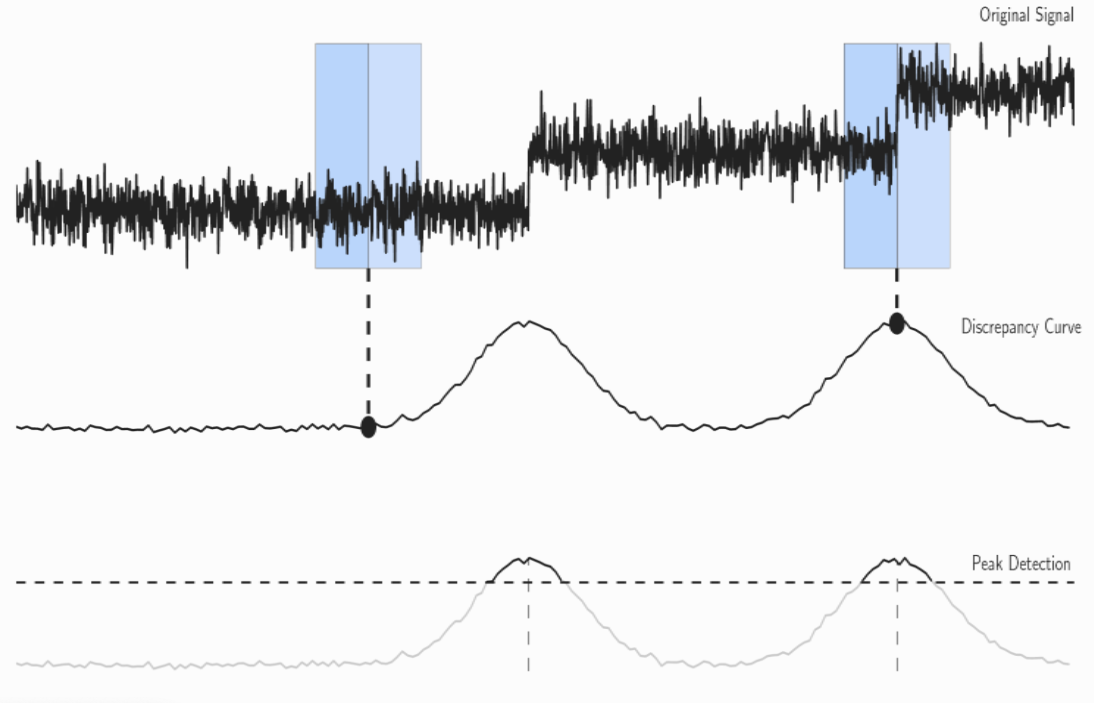
* To find the value of **m** here is tricky.
* Based on the seasonality in the data the m value for the differencing is chosen. For example,
  + - If the data have quarterly seasonality then m will be 4.
    - If the data have monthly seasonality then m will be 12.
* We use the same **diff()** method of **pandas** and first, we remove trend and then we remove seasonality from the series.

# **Change Points**

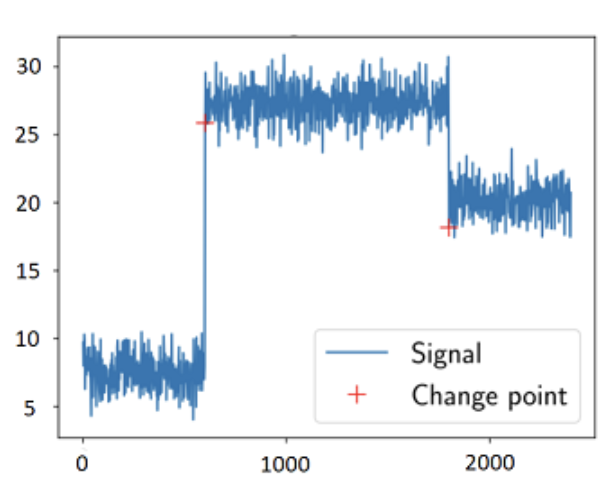
* A sudden change in a time series that changes the hidden trends and statistical characteristics of the data.
* A change point divides a time series into two segments with distinct statistical characteristics.



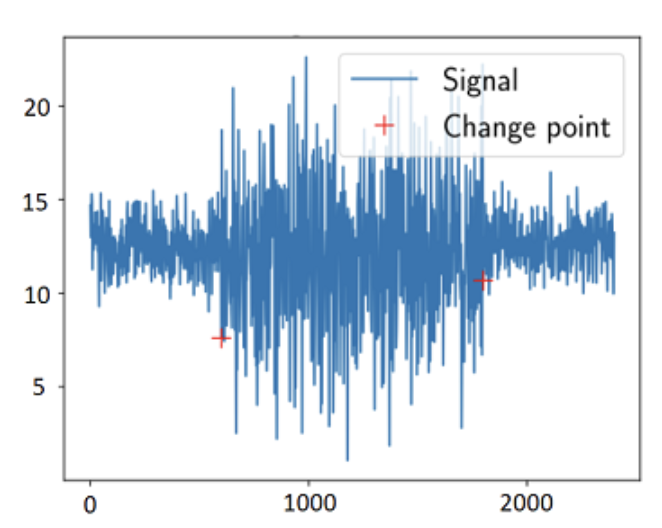
* To detect the change point:
  + - Walk through the series with a window of fixed size
    - For each step, compute the **cost** of all elements in the window. There are many options for this cost function.
    - Wherever the cost locally peaks, the center of the window can be considered as a changepoint.
    - We can add other conditions of threshold cost, etc to avoid making too many detections.
    - Cost is high than a certain threshold, change point is present, and vice versa.



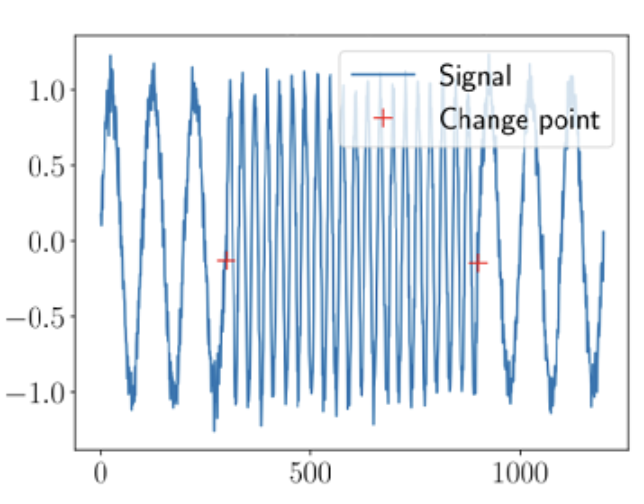
* There are various factors on which one can identify a change point. Some of them are:
  + - Mean
    - Variance
    - Periodicity
    - Pattern
* **Change in Mean:**
  + This is the most common and probably the easiest one to interpret and identify a change point.
  + Change in mean often occurs when a time series is the build-up of constant segments having different mean values.
  + Change points are defined as the first time step in each new segment starting with the second segment, so the number of change points is always one fewer than the number of segments.



* **Change in Variance**
  + Change in variance is another simple way of determining change points.
  + Here, the mean of the signal stays constant, but there are several segments with different variance values.
  + This can be interpreted as a sudden noise in the signal.



* **Change in Periodicity (Frequency)**
  + Change in periodicity (also called a change in frequency) is concerned with the changes occurring when the frequency changes suddenly.
  + Detection of this kind of change is usually done in the frequency domain, for example by using Fourier transform or wavelet transform.

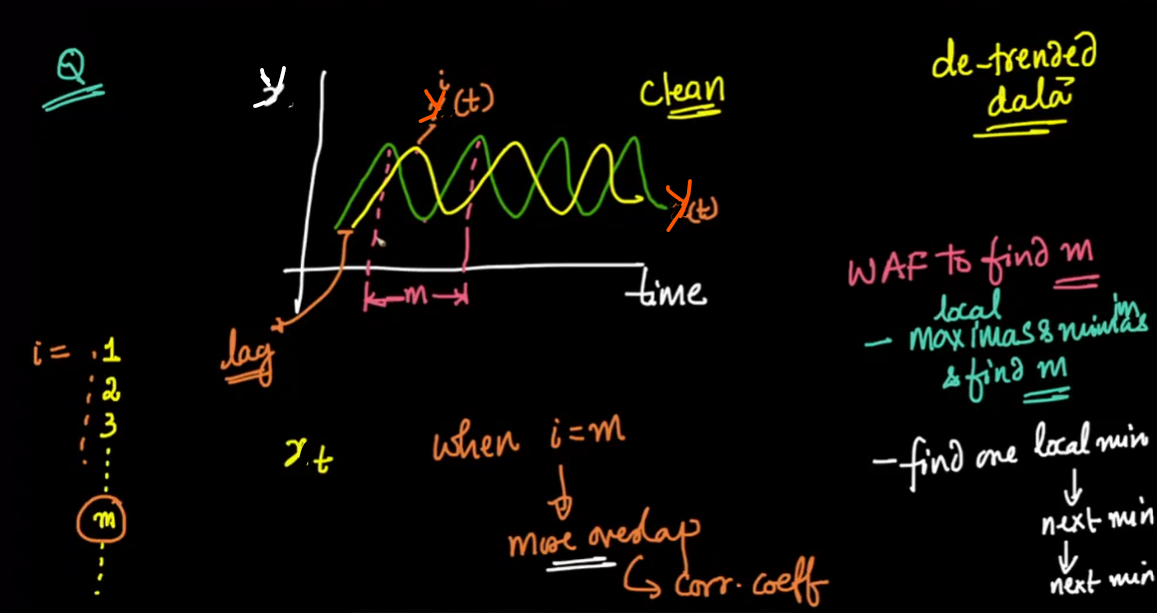


# **Autocorrelation**

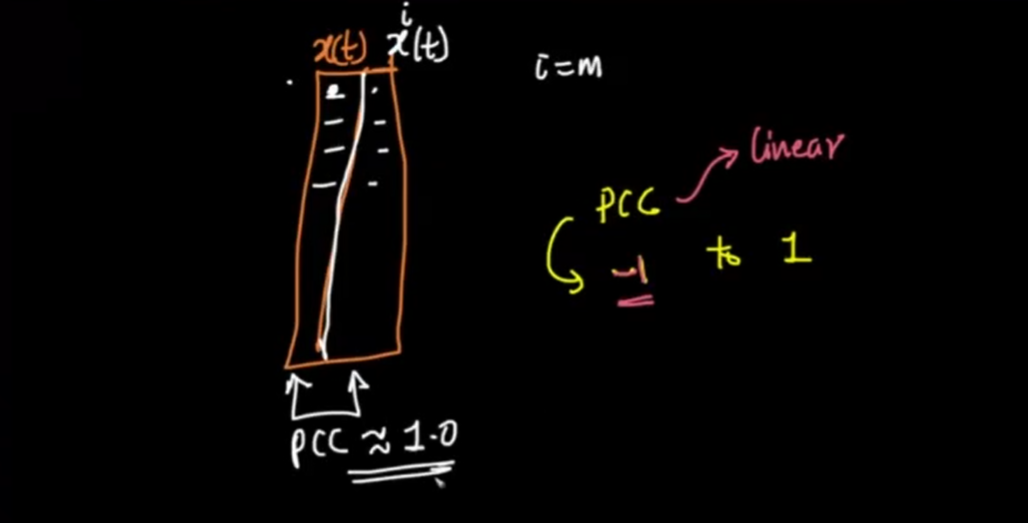
* One approach to finding the value **m** is that we find local maxima and minima in the time series, and try to analyze the intervals at which they are observed.
  + This approach makes sense.
* Another approach would be; given a time series 𝑦(𝑡),
  + What if we consider another time series where we introduce a **lag** of 1, i.e. **shift the series** by 1 unit of time: 𝑦`(𝑡)
  + We can then find the **correlation coefficient** between these two-time series: 𝑦(𝑡) and 𝑦`(𝑡).
  + Similarly, we find the correlation between the original time series 𝑦(𝑡) and a time series lagged by i units: 𝑦i(𝑡), where;

𝑖=1,2,3,... (i represents lag)

* + In doing so, we would find a value of 𝑖, where the lagged time series **roughly overlaps** over the original series.



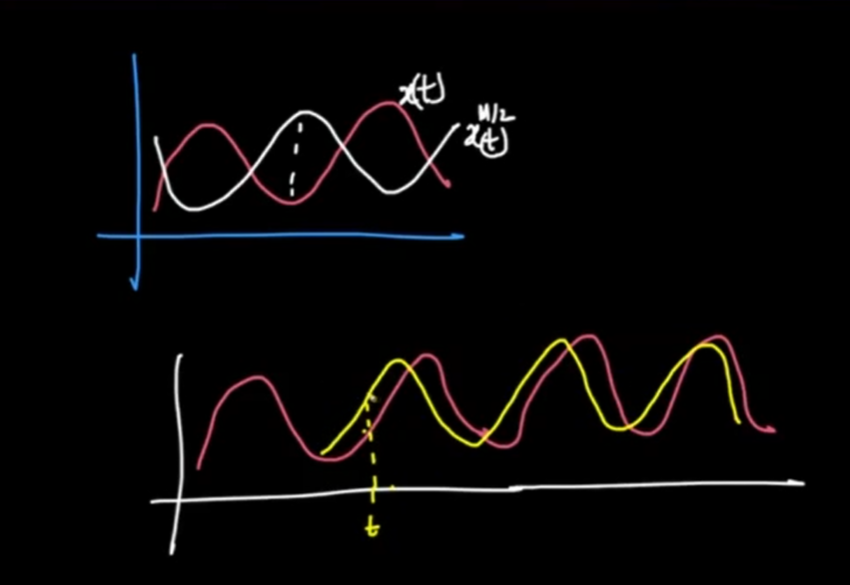
* Then, we create a table containing: 𝑦(𝑡) and 𝑦i(𝑡) time series values
* When ***i=m,*** (the lagged time series roughly overlaps over the original series), the Pearson correlation coefficient would be very close to 1.
* This value of 𝑖 would indicate the optimal value of 𝑚.



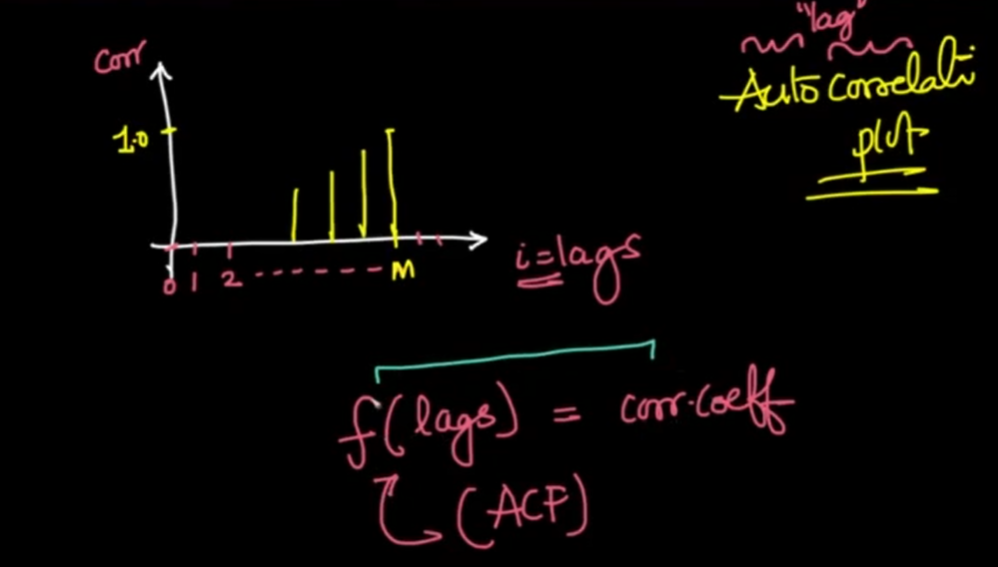
* **What are the advantages of using the concept of correlation for finding the optimal value of m?**
  + Easily interpretable
  + Value ranges from -1 to +1
  + Captures linear relationship

**Q. What will the plot between the correlation coefficient and lag (i) look like?**

* At 𝑖=𝑚, we would get a correlation value very close to 1.
* At 𝑖=𝑚/2, as the value of lagged time series increases, the value of the original series decreases, giving us a **strong correlation.**
* For a value of 𝑖 that is even close to 𝑚, though the correlation value would not be as strong as at 𝑚, it would be relatively strong.

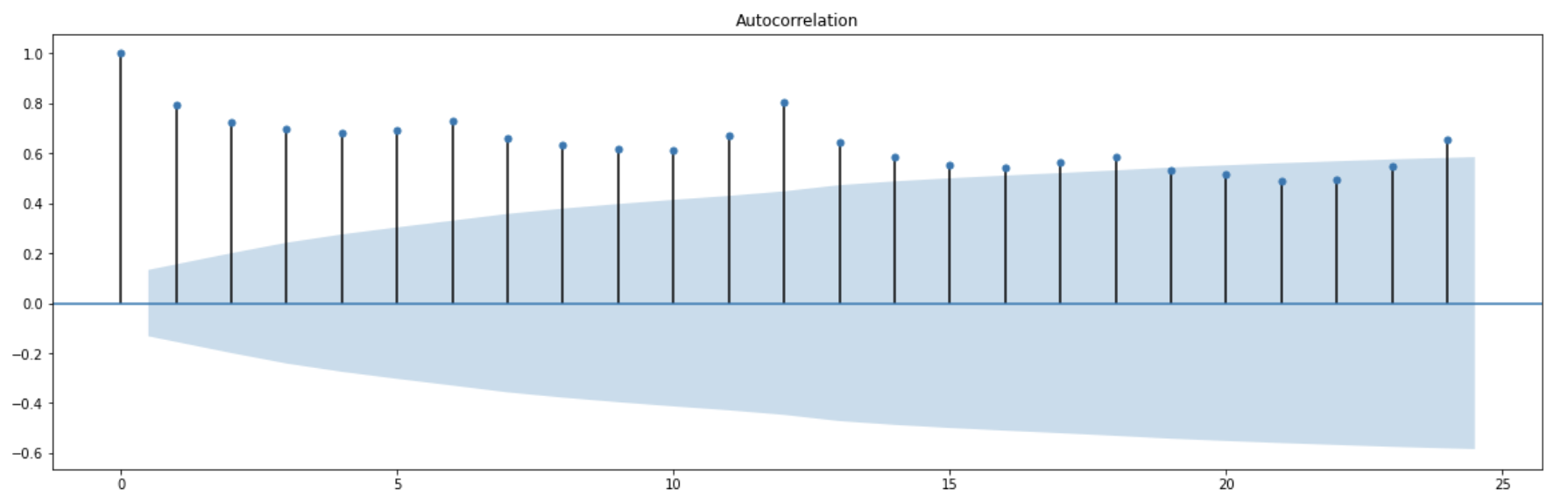


* So, the final plot between lag value (i) and correlation coefficient would look something like this:-



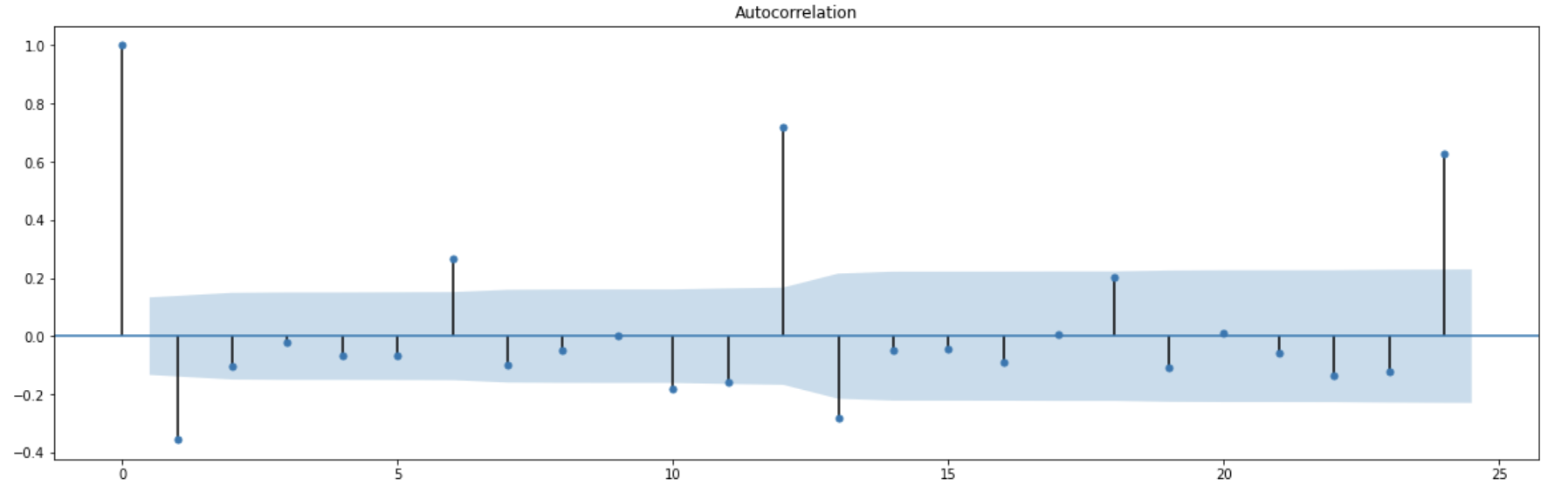
* This plot is called an **Autocorrelation Plot** because we are computing the correlation of the series with itself with various values of lag.
* This can be written as a function also. That is called the **Autocorrelation Function (ACF)**.
* The ACF plot shows both the direct and the indirect impact of the correlation values. It takes the average correlation between the time series original values and the lag values for different lags.
* If the time series is random the lag values will not have much autocorrelation with the original values.

### **ACF Plot of Trended Series:**



### 

### **ACF Plot of De-Trended Series:**



* After de-trending the series, we can see the correlation of the series is high at periods 6 and 12. This indicates there is some seasonality.
* Besides that, the correlation values are random and small.
* On removing the trend, we're **able to capture negative correlations** between the original series and lagged series with more ease.
* The blue color highlighted part is the **confidence interval** which gives the significance level of the correlation
* If the correlation is higher or outside of the blue highlighted area, that can be considered as a highly significant value.
* So the major values of lag we consider to study the seasonality pattern here are:-
* Correlation at lag=1
* Correlation at lag=6
* Correlation at lag=12
* Correlation at lag=24
* Other points lie within or very close to the blue shaded region of confidence intervals.
* This means that we are not very confident about these correlation values
* The confidence interval increases as we consider time series that lagged more.

# 

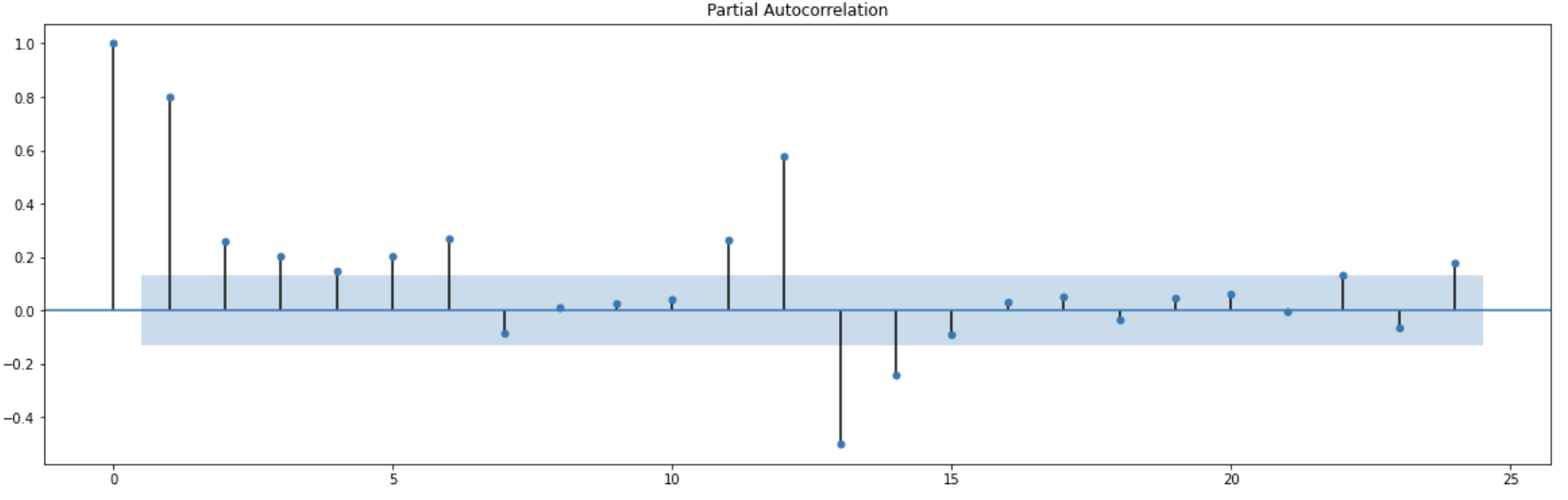
# **Partial Autocorrelation**

* This is similar to AutoCorrelation with only a small difference.
* We try to find a relationship between the original time series 𝑦(𝑡) and time series lagged with 𝑖 steps 𝑦𝑖(𝑡), where 𝑖=1,2,3,..., to find the optimal value of 𝑚.
* This plot only explains the values which have a direct impact on the original values.
* The difference is that: **All intermediate/indirect correlations are removed**.
  + The correlation between observations at successive time steps is a linear function of the indirect correlations.
  + These indirect connections are eliminated using the **partial autocorrelation function (PACF)**.

For example

* When we're considering the correlation between 𝑦(𝑡) and say, 𝑦12(𝑡),
  + Then, we do not want this correlation value to get corrupted by the correlations when 𝑖=1,2,3,...,11 (i.e. the intermediate correlations)
* The partial correlation for each lag is the **unique correlation** between the two observations after the intermediate correlations have been removed.
* This is also known as **Conditional AutoCorrelation**.

### **PACF Plot:**



* Here, one can see:
  + Value at 𝑖=1 is high: This means that given the values of time series 𝑦(𝑡), we can compute the values of a time series with lag 1, 𝑦1(𝑡) with ease.
  + There is a high value at 𝑖=12
    - This means that even if we ignore the information given by all the time series with 𝑙𝑎𝑔𝑠=1,2,...,11, the information carried in time series 𝑦12(𝑡) alone is very high.
    - This shows strong seasonality.
* We calculate PACF on the original time series, whereas ACF is plotted on stationary time series.

# **Correlation Vs Causation**

* If a variable x is correlated with another variable y, does that mean that x causes y?
  + A variable x may be useful for forecasting a variable y, but that does not mean x is causing y.

For example,

* + It is possible to model the number of drownings at a beach resort each month with the number of ice creams sold in the same period.
  + The model can give reasonable forecasts, not because ice-creams cause drownings, but because people eat more ice creams on hot days when they are also more likely to go swimming.
  + So the two variables (ice cream sales and drownings) are correlated, but one is not causing the other. They are both caused by a third variable (temperature). This is an example of **confounding**.
* Correlations are useful for forecasting, even when there is no causal relationship between the two variables.
  + For example, It is possible to forecast if it will rain in the afternoon by observing the number of cyclists on the road in the morning.
  + When there are fewer cyclists than usual, it is more likely to rain later in the day.
  + This model can give reasonable forecasts, not because cyclists prevent rain, but because people are more likely to cycle when there is less or no chance of rain.
  + In this case, there is a causal relationship, but in the opposite direction to our forecasting model.
* Though we can get good forecasts based on correlated variables, if we try and understand the causality behind those variables, we can identify better features, thereby creating a better model.
  + A better model for the example of the drowning will probably include temperatures and visitor numbers and exclude ice cream sales.
  + A good forecasting model for rainfall will not include cyclists, but it will include atmospheric observations from the previous few days.