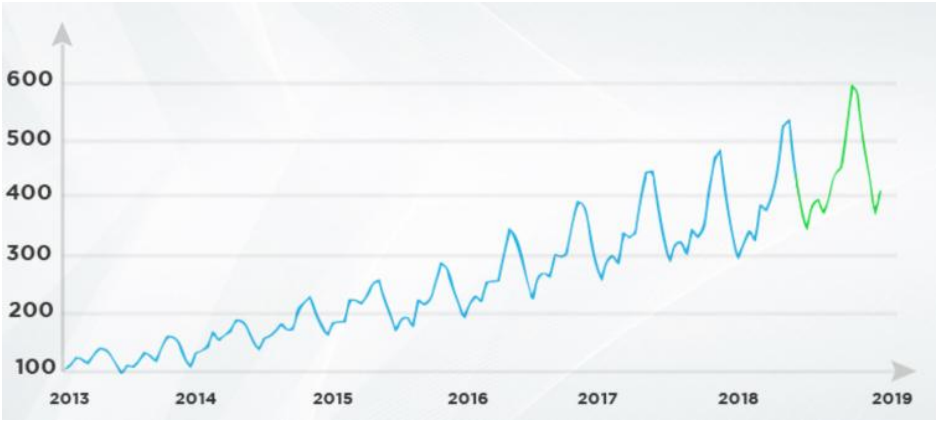
Time Series Analysis

- Its a series of data points listed in time order.
- A set of observations ordered in equal intervals of time.
- We are recording a data with respect to time.
- Eg. Temp, Sales, Stock, no of passengers, rainfall.

Time-series forecasting

It’s a process of using a model to predict future values based on previously observed values.



Applications:

- Retail stores forecasting sales
- Energy companies forecasting reserves, production, demand, and prices
- Educational institutions forecasting enrollments
- The government forecasting tax receipts and spending
- International financial organizations such as the World Bank and International Monetary Fund forecasting inflation and economic activities

- Passenger transport companies use time-series to forecast future travel

- **We must keep the data in order** so as not to violate the relationship of the TS data which claims that the future values are a function of past values.
- How is TS different from Regression?
- In regression we never bothered about the order of the data but in TS its required.
- Time Series sometimes do not have features but regression have it.

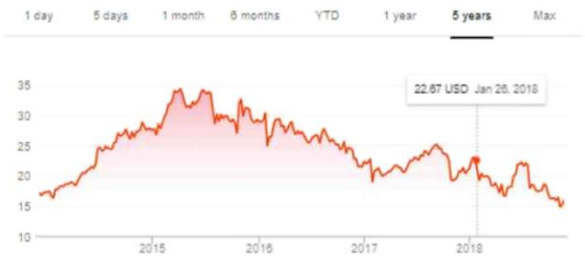
Jan	123
Feb	234
Mar	145
Apr	151

- Decomposition of TS is a process of separating a time series into its components: **trend, seasonality and residuals**.

Trend:

Uptrend

Smartphone sales for a 3 year period



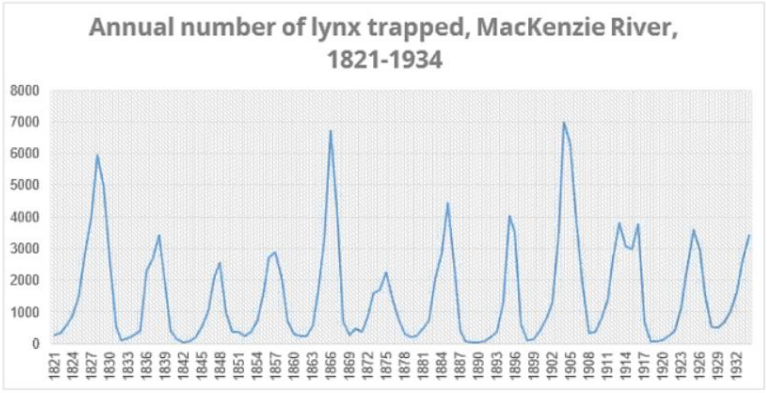
Downtrend

Stock Market price for a wall street company

Trend

A trend is a long-term increase or decrease in time series data

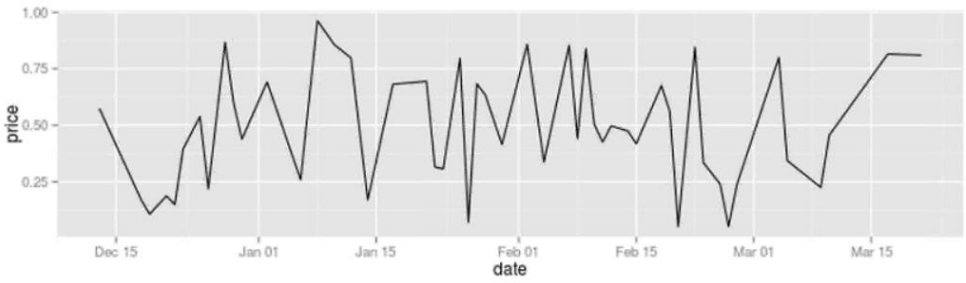
Seasonality

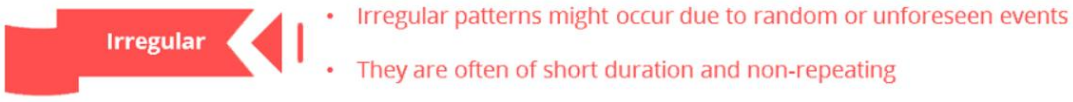


Seasonal

- When factors such as the time of the year or the day of the week affect the dependent variable, repetitive patterns are observed in the time series
- Seasonality is always of a fixed and known frequency

- **Residuals**.



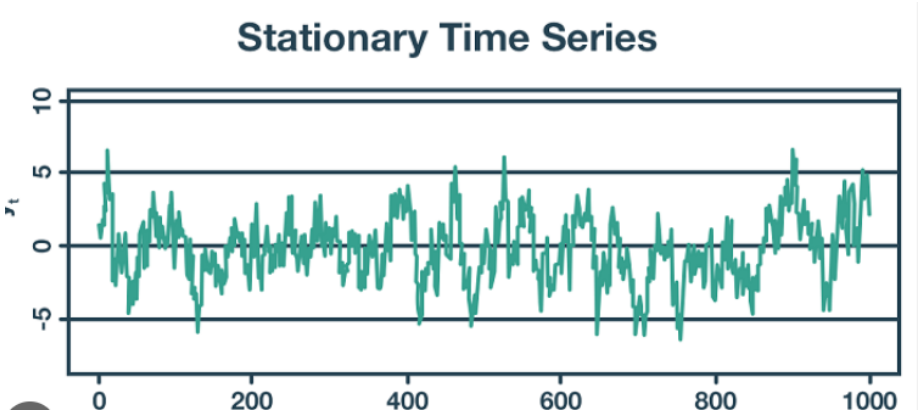


Steps for TS modeling

- Visualize the Time Series Data.
- Make the time series data stationary.
- Plot the Correlation and AutoCorrelation Charts.
- Construct the Model based on the data.
- Use the model to make predictions

Stationary :

- A stationary time series is one whose statistical properties are time invariant.
- The mean, variance and autocorrelation of time series do not change with time.
- This would mean that the model parameters should also change.
  And this makes impossible for a model to derive a function of future values as a function of past values.
- All the models assume the data to be stationary.



- If the data is non-stationary, then how do we make it stationary?
- A simple transformation called as differencing helps achieve stationary time series.

- What is differencing? Differencing involves calculating the series of change from one timestep to other. This can be formulated as

y(differenced) = y(t) - y(t-1)

Example :

1 2 3 4 5 6 7 8 9

Clearly the mean of the TS is changing? How

Mean (1 2 3) = 2
Mean (2 3 4) = 3
Mean (7 8 9) = 8

This TS is non stationary. Let us make it stationary. How ?

Difference it.

1 2 3 4 5 6 7 8 9

NAN 1 1 1 1 1 1 1 1

.

How do we test for stationarity?

- A common test is the Augmented Dickey Fuller (ADF) test. It is based on hypothesis testing.

#Ho: It is non stationary

#H1: It is stationary

- ADF test generates p-value

If p val <0.05 then reject null hypothesis else accept it.

AutoCorrelation Function (ACF):

Correlation - extent of a linear relationship between two variables.

Auto Correlation - measure of linear relationship between lagged value of a time series.

Example :

y(t) and y(t-1) has a lag of 1 -> correlation coefficient1

y(t) and y(t-2) has a lag of 2 -> correlation coefficient2

Corelation:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

r = correlation coefficient

x_i = values of the x-variable in a sample

$\bar{x}$ = mean of the values of the x-variable

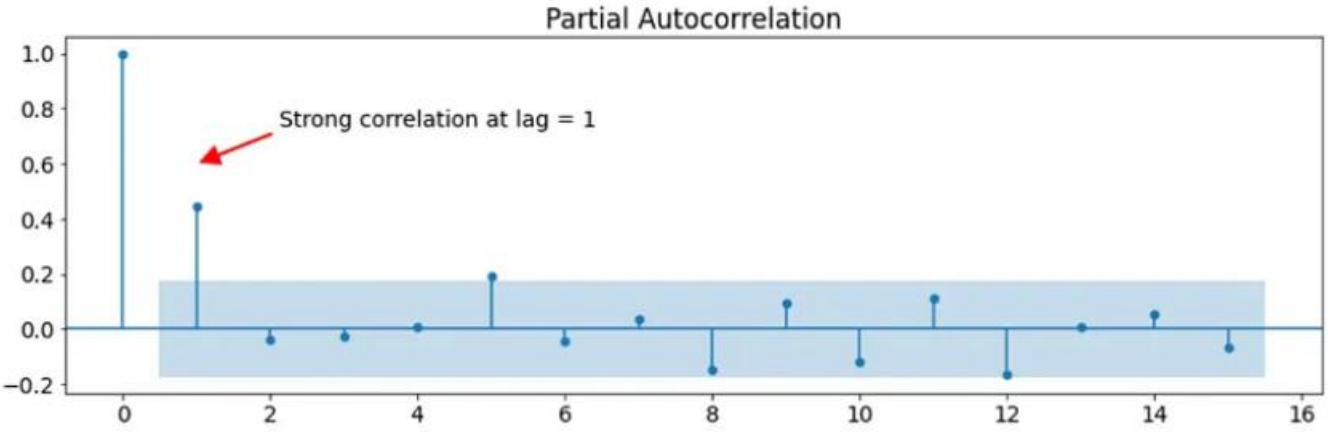
y_i = values of the y-variable in a sample

$\bar{y}$ = mean of the values of the y-variable

1. **Auto Regressive Model (AR)**

- It is use to predict future value using previous value.
- The current observation is a linear combination of previous observations
- The number of previous observations is given by the order or AR process
- The order (P) of AR process is identified by the **PACF plot**.

$$y(t) = C + B1.y(t-1) + B2.y(t-2) + ...Bp.y\ (t-p) + e$$



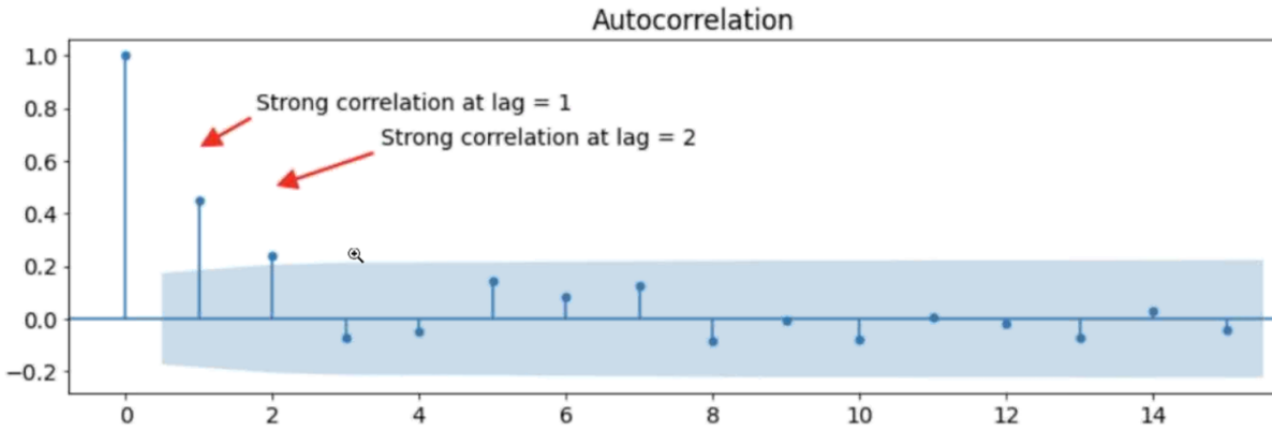
Here Order =1 (P=1)

2. Moving Average Model (MA)

- It is use to predict future value using past error.
- The current observation is a linear combination of current and previous error terms.
- The number of previous error terms is given by the order or MA process
- The **order(q)** of MA process is identifed by the **ACF plot**.

y(t) = mean + B0.E(t) +B1.E(t-1) + B2.E(t-2) + ...Bq.E(t-q)

Here E(t), E(t-1) ..... are nothing but normally distributed error terms i.e. **white noise**.



Here degree is 2 (q=2)

3. **ARMA**

- Combination of Autoregressive and Moving Average

4. ARIMA (p d q)

- Combination of differencing, auto regressive and moving average

AR I MA

AR: Auto regressive model (p) use PACF

I: Differencing to convert non stationary to stationary (d)

MA: Moving average (q) use ACF