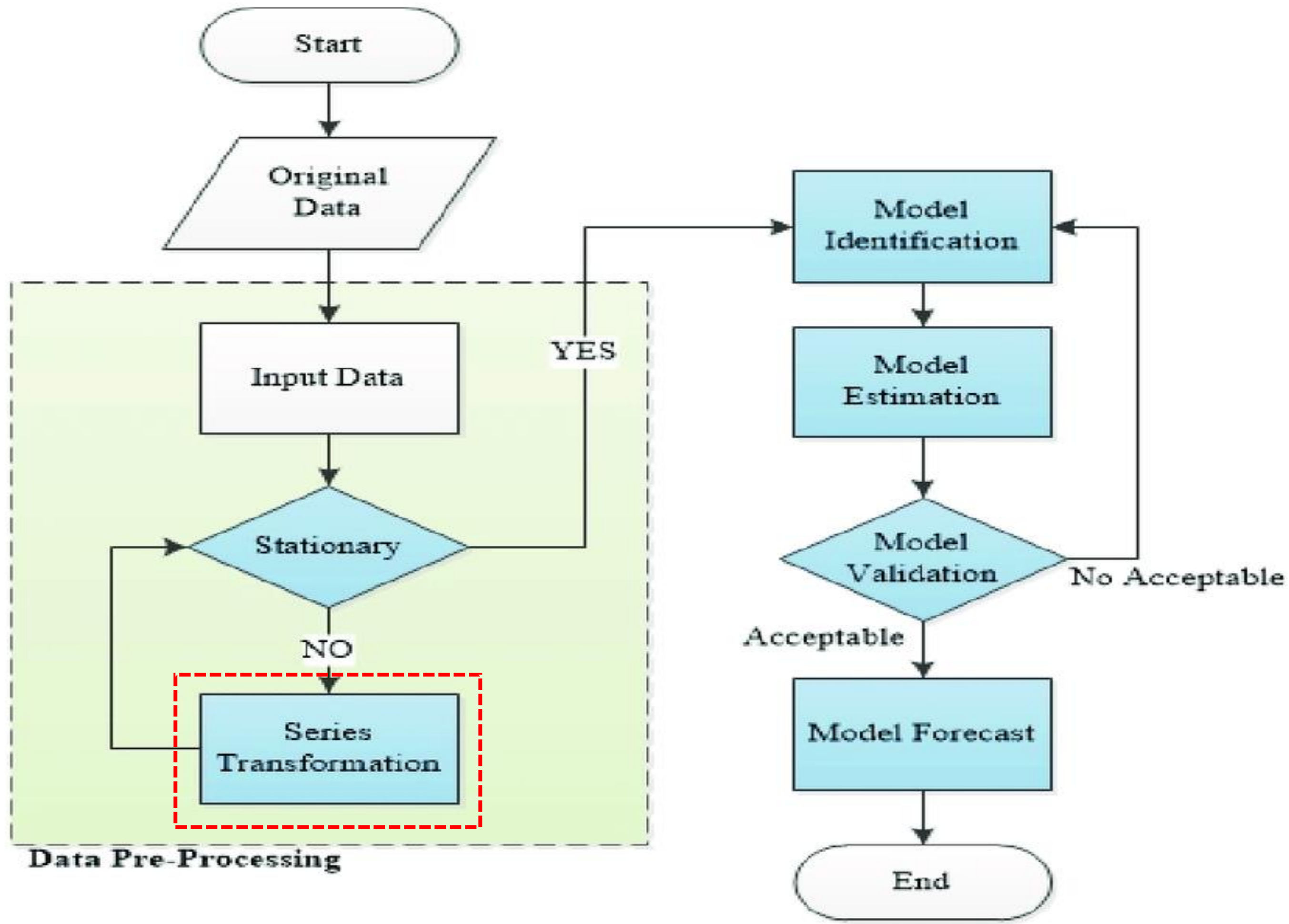
TIME SERIES

Time Series Analysis Flowchart



CONTENT

- Time series definition and application.
- Time series components.
- White noise, Stationarity, Autocorrelation, ACF and PACF.
- AR model.
- MA model.
- ARMA model.
- ARIMA model.

TIME SERIES

Time series is a series of statistical observations arranged in chronological order.

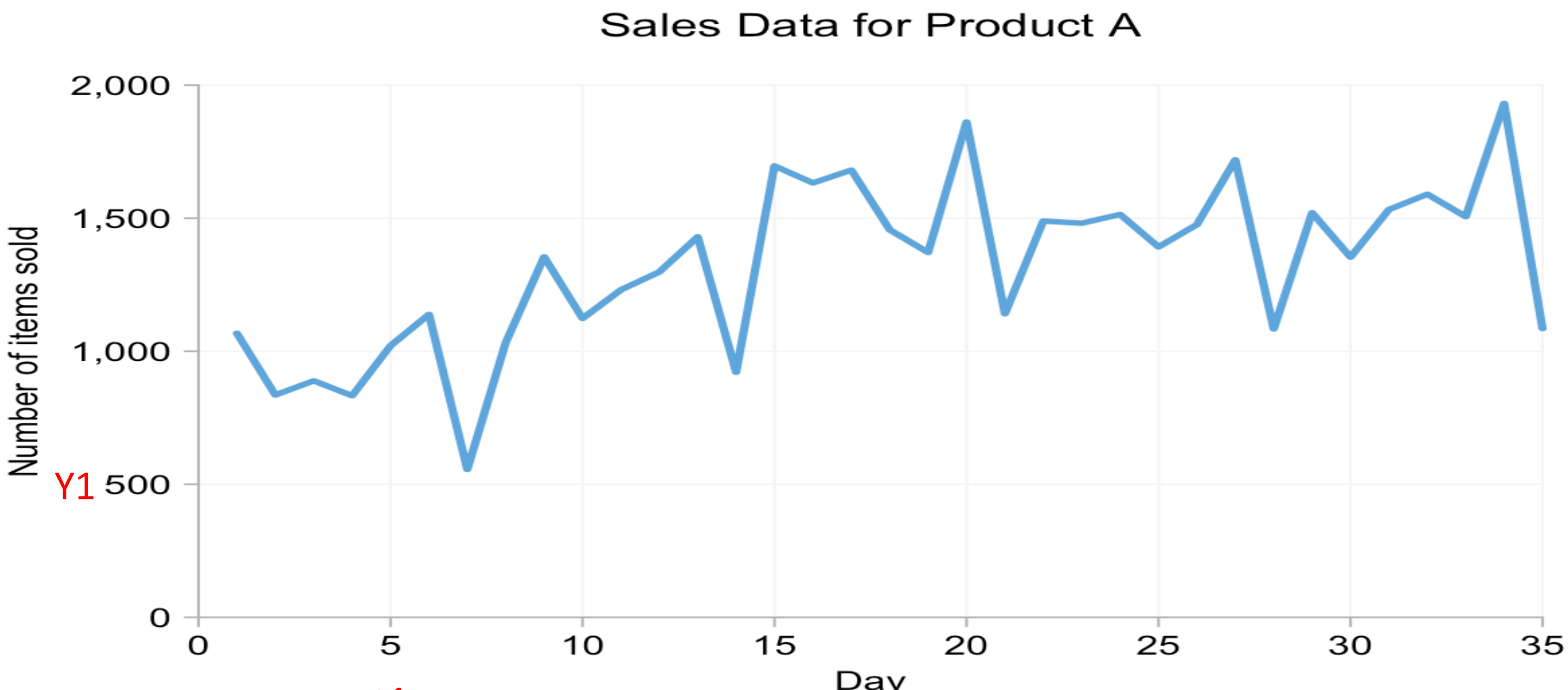
Examples:

- Daily price of gold
- Monthly deposits in a certain bank
- Daily record of maximum temperature in a city

TIME SERIES

- Number of Items sold per day or a period of 35 days

Y



Y1

t1

t

WHAT IS TIME SERIES?

- In logistics regression or linear regression algorithm we deal with 2 or more variables.
- In **Time Series** we deal with a single variable which is dependent on time. Time is the independent variable.
- x-axis has time and y-axis has other variable
- The scale of x-axis has the same unit - seconds, minute, hours, days etc.

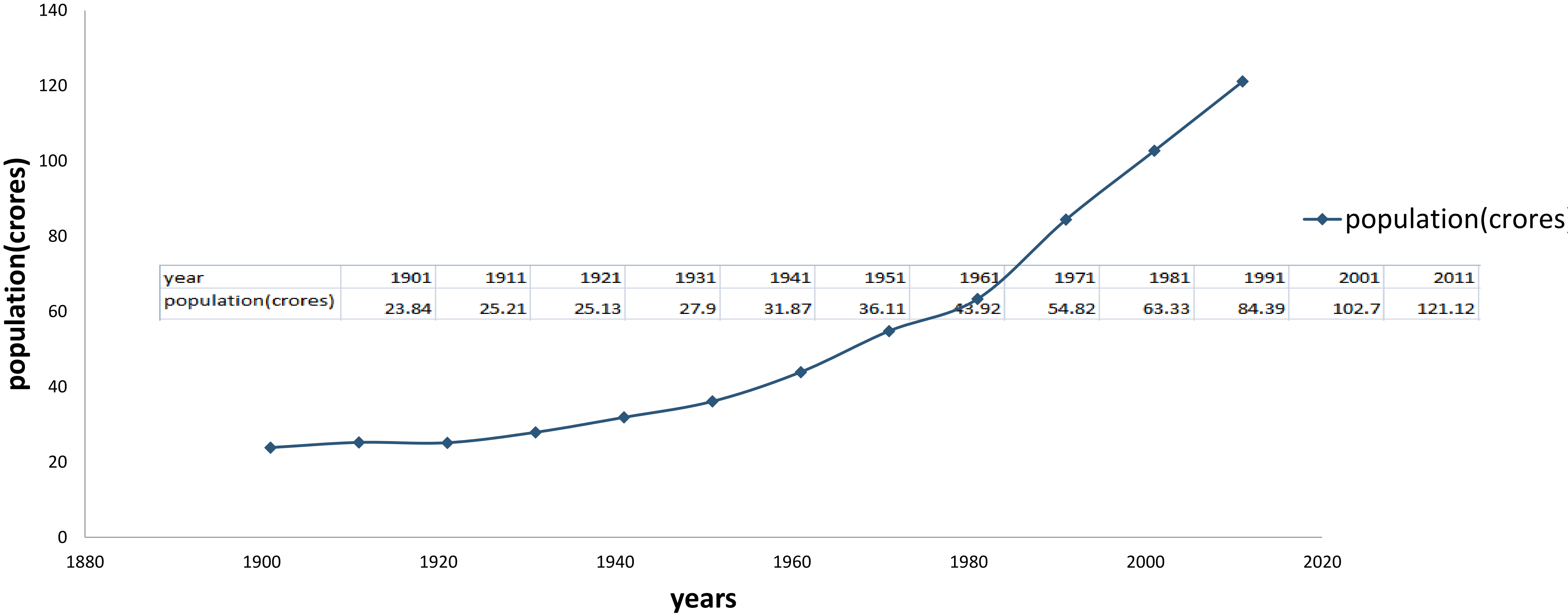
WHAT IS TIME SERIES ANALYSIS?

- A time series is a series of data points indexed (or listed or graphed) in time order. Most commonly, a time series is a sequence taken at successive equally spaced points in time. Time series forecasting is the use of a model to predict future values based on previously observed values.

WHAT IS TIME SERIES?

- A time series can be defined as a set of data dependent on time
- Acts as an independent variable to estimate dependent variable
- Mathematically , it is a set of observation taken at specified time(usually at equal intervals)
- A time series defined by the values Y1, Y2…. Of a variable Y at times t1, t2…is given by ,Y=F(t)

TIME SERIES - EXAMPLE OF POPULATION OF INDIA IN CENSUS YEARS

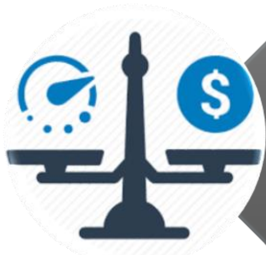


year	1901	1911	1921	1931	1941	1951	1961	1971	1981	1991	2001	2011
population(crores)	23.84	25.21	25.13	27.9	31.87	36.11	43.92	54.82	63.33	84.39	102.7	121.12

IMPORTANCE OF TIME SERIES ANALYSIS

Business forecasting

Understanding past behavior



Comparison of related time series

COMPONENTS OF TIME SERIES ANALYSIS

Trend

- The data goes up or down is trend.
- If data is horizontal it is called stationary trend

Seasona lity

- Fluctuations in the pattern due to seasonal determinants over a short period.

Irregular

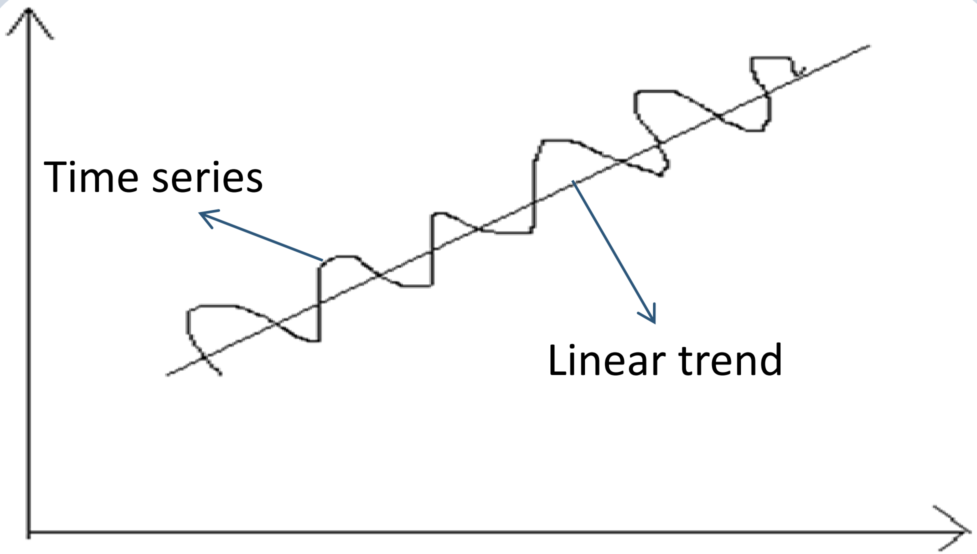
- Instability due to random factors that do not repeat in the pattern. White Noise

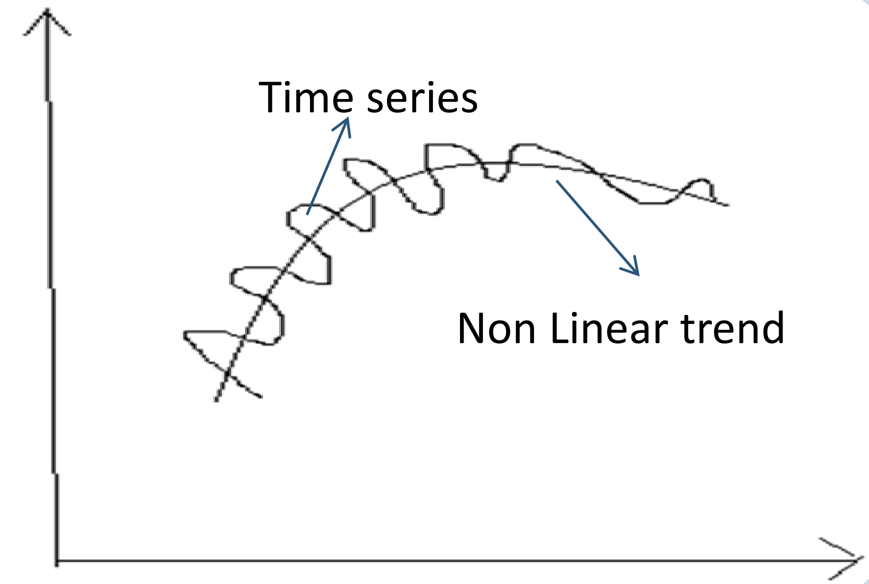
Cyclical patterns

- Variations occurring at irregular intervals due to certain circumstances.

TREND

- The trend describes the general behavior of a time series. If a time series manifests a positive long-term slope over time, it has an *upward* trend. If instead, it describes a general negative slope, it has a *downtrend*.





In linear trend the value of Y can be approximately by straight line.

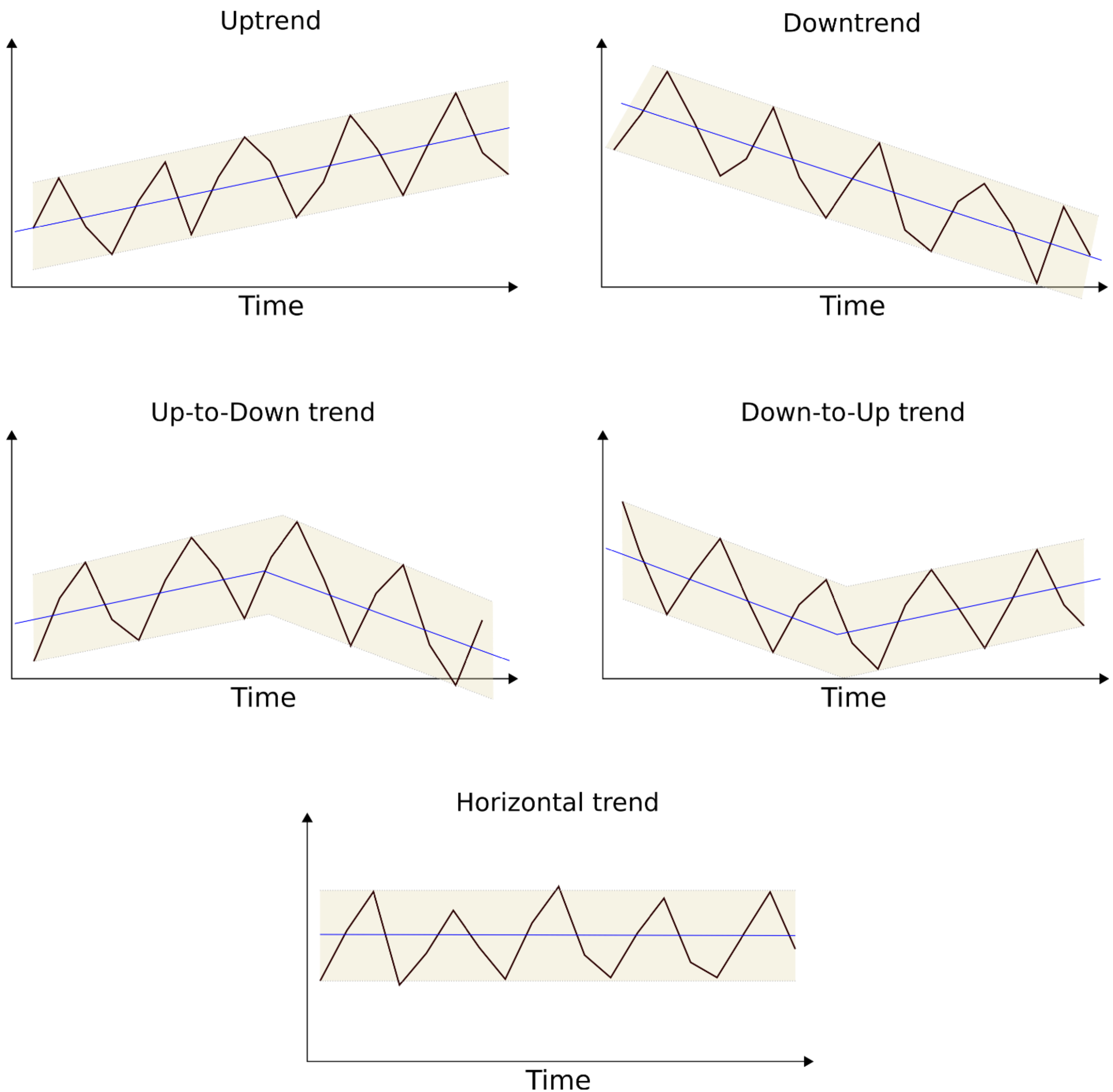
Linear trend is commonly used in business and economics

In non linear trend, the growth rate is different over different sector of time .

Non-linear trend is used in the study of population birth rates etc.

Do It Skills
do it. enjoy it.

TREND



Do It Skills
do it. enjoy it.

TREND

- The time series oscillates around trend. The trend may be go to upward to go downward or to remain stagnant. In some cases, Trend may happen for some time and disappear
- Trend is due to the reasons of following nature,

i) changes in population

ii) technological developments

iii) changes in economy

iv) changes in habits and tastes

of people.

Do It Skills
do it. enjoy it.

SEASONALITY

- A seasonal pattern is any kind of fluctuation (change) in a time series that is caused by calendar-related events.
- These events can be the time of year (like winter or summer), or the time of day or the week. Seasonality always has fixed frequencies. That is, a seasonal pattern always starts and ends in the same period of a week, year, etc.

SEASONALITY

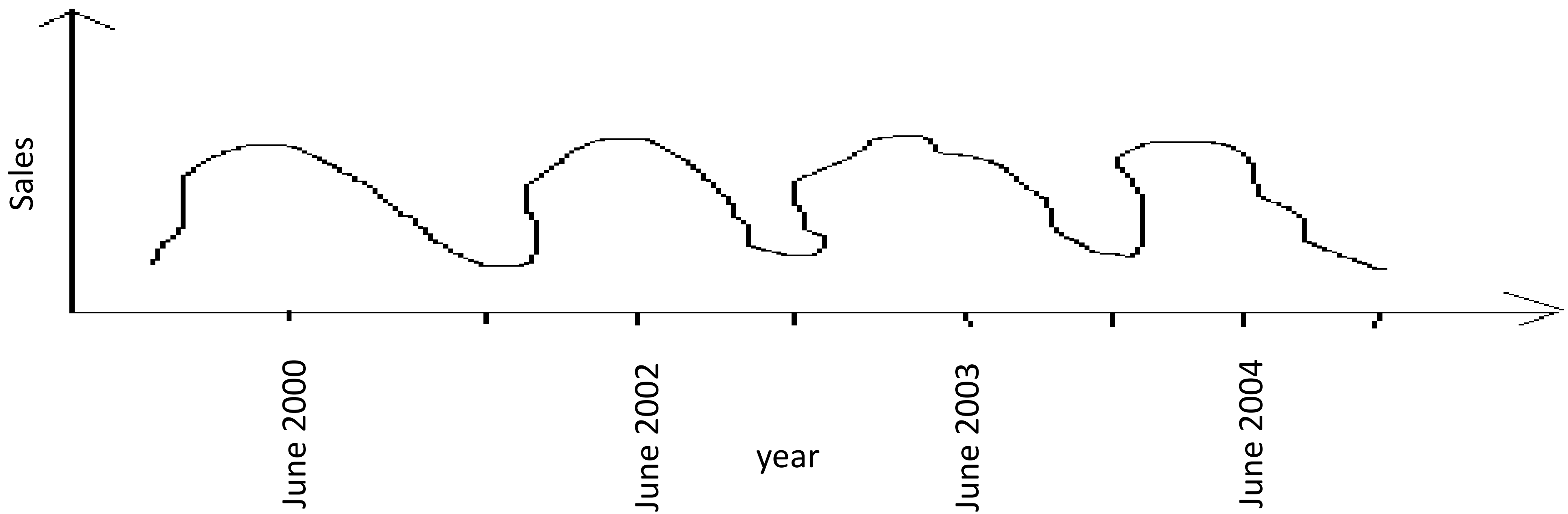
A seasonal pattern is any kind of fluctuation (change) in a time series that is caused by calendar-related events.

- These events can be the time of year (like winter or summer), or the time of day or the week. Seasonality always has fixed frequencies. That is, a seasonal pattern always starts and ends in the same period of a week, year, etc.
- When trend is stagnant, seasonal variation become predominant component.
- The amplitude may differ from cycle to cycle.
- Seasonal variations are evident due to festival, climate conditions, fashions or habits of people.
- These factors operate in a regular and periodic manner where the period of recurrence is generally one year.

SEASONALITY- EXAMPLES

- ✓Take a data center as an example. If we consider the cooling system as the primary source of energy consumption, it is easy to imagine that in the summer, energy costs probably go up, while winter might show a decrease in energy consumption.
- ✓Also, a clothing store that sells heavy coats might observe higher selling rates during winter, as opposed to the summer.
- ✓For example, Bank managers has to arrange for proper cash flow during the beginning of month, festivals .

SEASONALITY PLOT



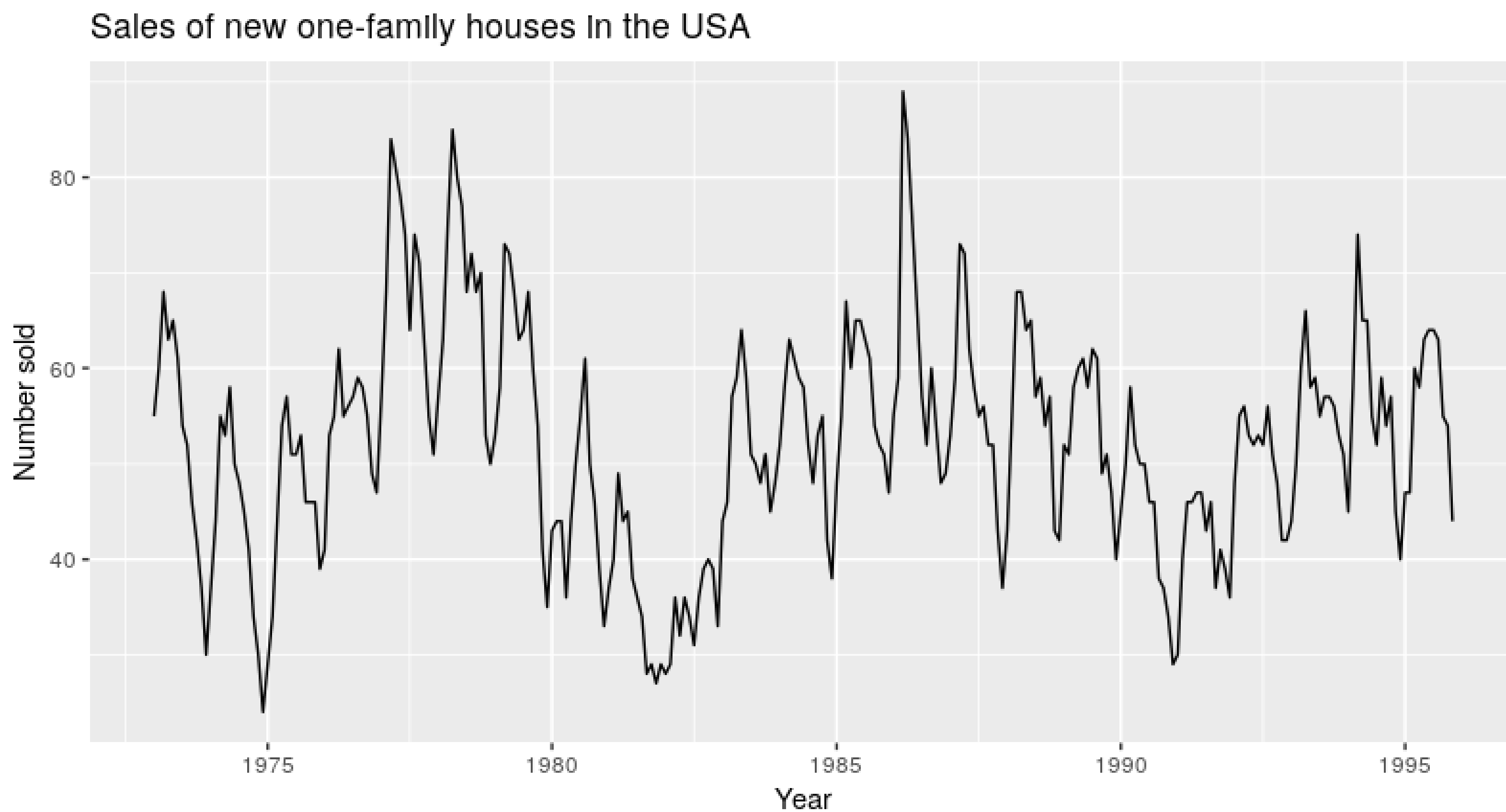
- Graph shows seasonal variations in the sales of a product. Note that the sales attain maximum in the month of June every year

CYCLICAL

- Cyclical pattern in a time series is a kind of change that is not related to seasonal factors. These are rises and falls with non-fixed magnitudes that can last for more than a calendar year. Cyclical patterns are not repetitive. Usually, they result from external factors which make them much harder to predict.
- Forecasting methods usually take advantage of these patterns to produce reliable predictions.

CYCLICAL

- Below, you can see time series data for Sales of new single-family houses in the USA. Note the strong seasonality. House sales are normally slow at the beginning of the year. Peaks occur around the months of June and July.
- Also, there are strong cycles that might range from six to ten years. Remember, cycles do not have fixed periods.



CYCLICAL VARIATION

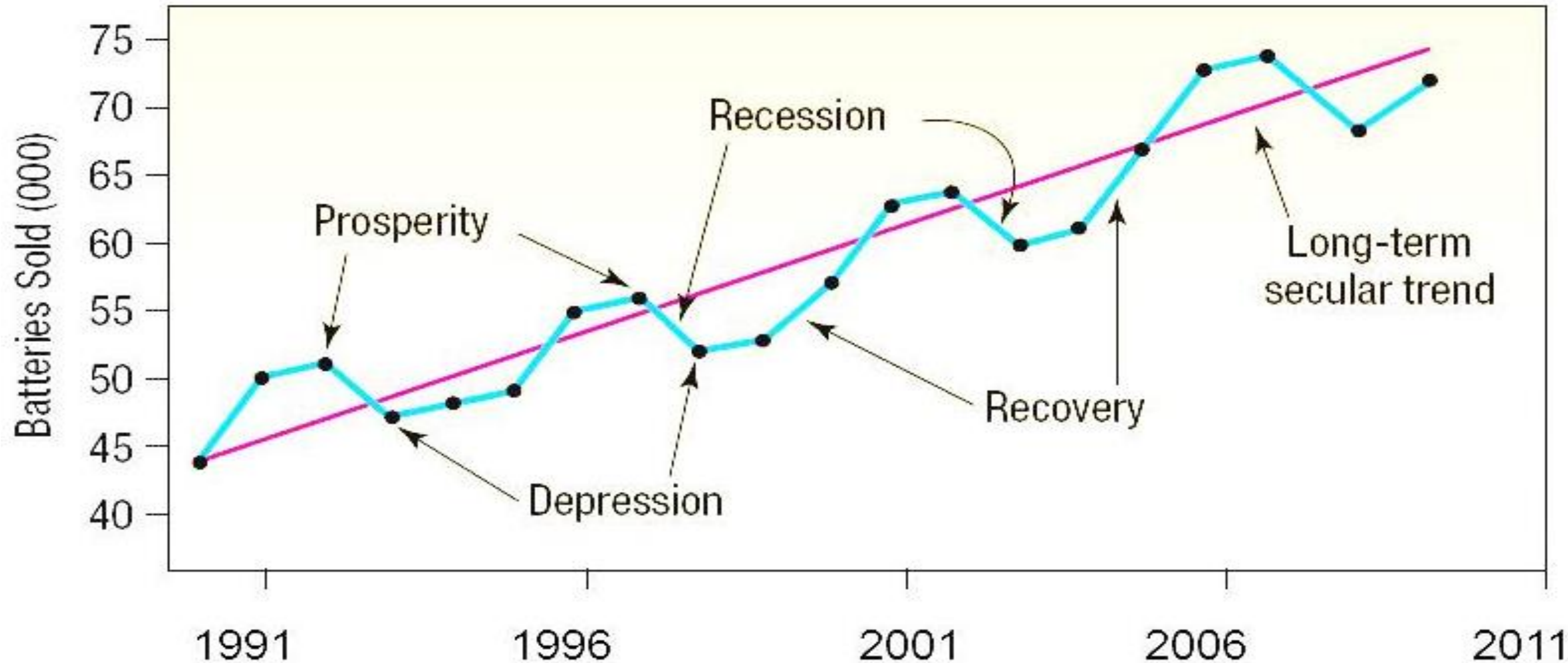
Cyclical variations may not be necessarily uniformly periodic. Period of cycle is about 7 to 10 years.

The ups and downs may occur at different intervals of time

These fluctuations are typically observed in business. Boom in business is followed by depression and vice-versa.

There are four phases in cycle.

- prosperity(boom)
- recession
- Depression
- recovery.

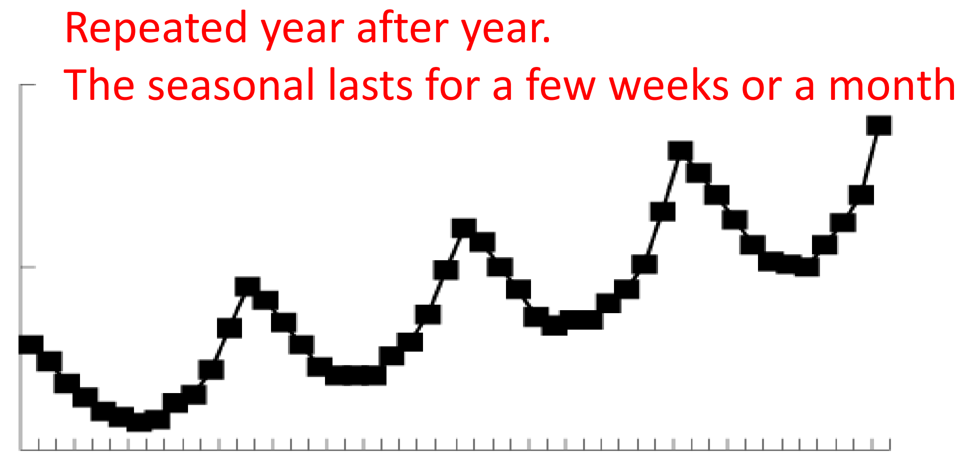


Do It Skills
do it. enjoy it.

IRREGULAR VARIATION

- Irregular variations are unpredictable and are the results of unforeseen forces or abnormal events.
- These events does not follow any pattern
- These variations are generally caused by calamities such as earthquakes, famines floods, epidemics or abnormalities such as war, lockouts etc.
- Irregular variations are mixed-up with seasonal and cyclical variations.

(a) Secular Trend



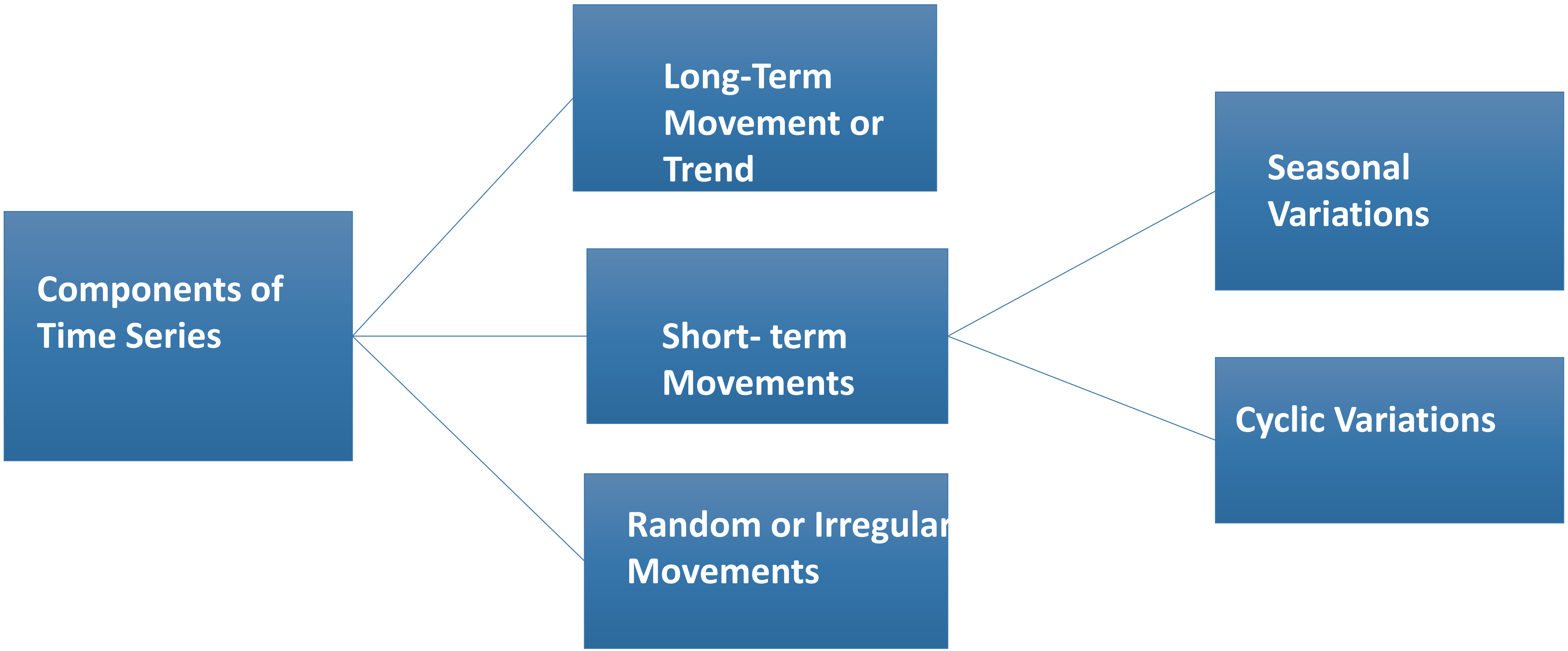
(b) Seasonal Variation

(c) Cyclical Variation

(d) Irregular Variation

TIMESERIES DATA MORE DETAILS

COMPONENTS OF TIME SERIES



Do It Skills
do it. enjoy it.

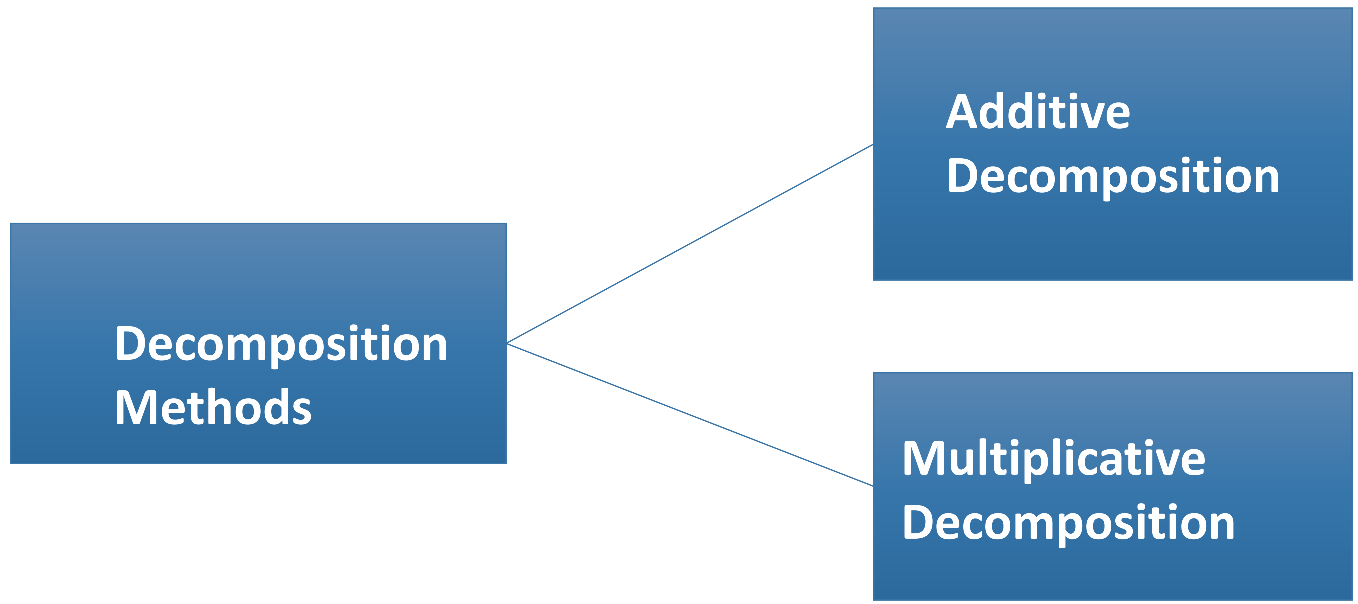
COMPONENTS OF TIME SERIES HELP IN MODEL SELECTION

- We need to analyze the various components of a time series and see how the components behave.
- Knowing the components of timeseries help us in machine learning pipeline — specially in model selection.

DECOMPOSITION

FOR FORECASTING, WE NEED TO FIRST USE DECOMPOSITION METHODS

- Choosing an appropriate forecasting method is one of the most important decisions an analyst has to make. While experienced data scientists can extract useful intuitions only by looking at a time series plot, time series decomposition is one of the best ways to understand how a time series behave.



DECOMPOSITION METHODS APPROACH

Depending on the Decomposition method we choose, we can apply each one of these components differently. A good way to get a feel of how each of these patterns behaves is to break the time series into many distinct components. Each component represents specific patterns, like trend, cycle, and seasonality.

Time series data 3 patterns(components)

- Trend
- Cycles
- Seasonality

Let us consider an example of International airline passengers' time series dataset. This contains 144 monthly observations from 1949 to 1960.

DECOMPOSITION TYPES

On Passenger data two types of decomposition: additive and multiplicative

- **Additive decomposition** assumes that a time series is composed of three additive terms.

$$y_t = S_t + T_t + R_t$$

- **Multiplicative decomposition** combines the terms through multiplication.

$$y_t = S_t \times T_t \times R_t$$

S represents the Seasonal variation

T encodes Trend plus Cycle

R describes the Residual or the Error component.

ADDITIVE

The first step is to extract the trend component of the international airline passenger time series. There are many ways to do this. Here, we compute it as the centered moving average of the data. The moving average smoother averages the nearest ***N periods*** of each observation.

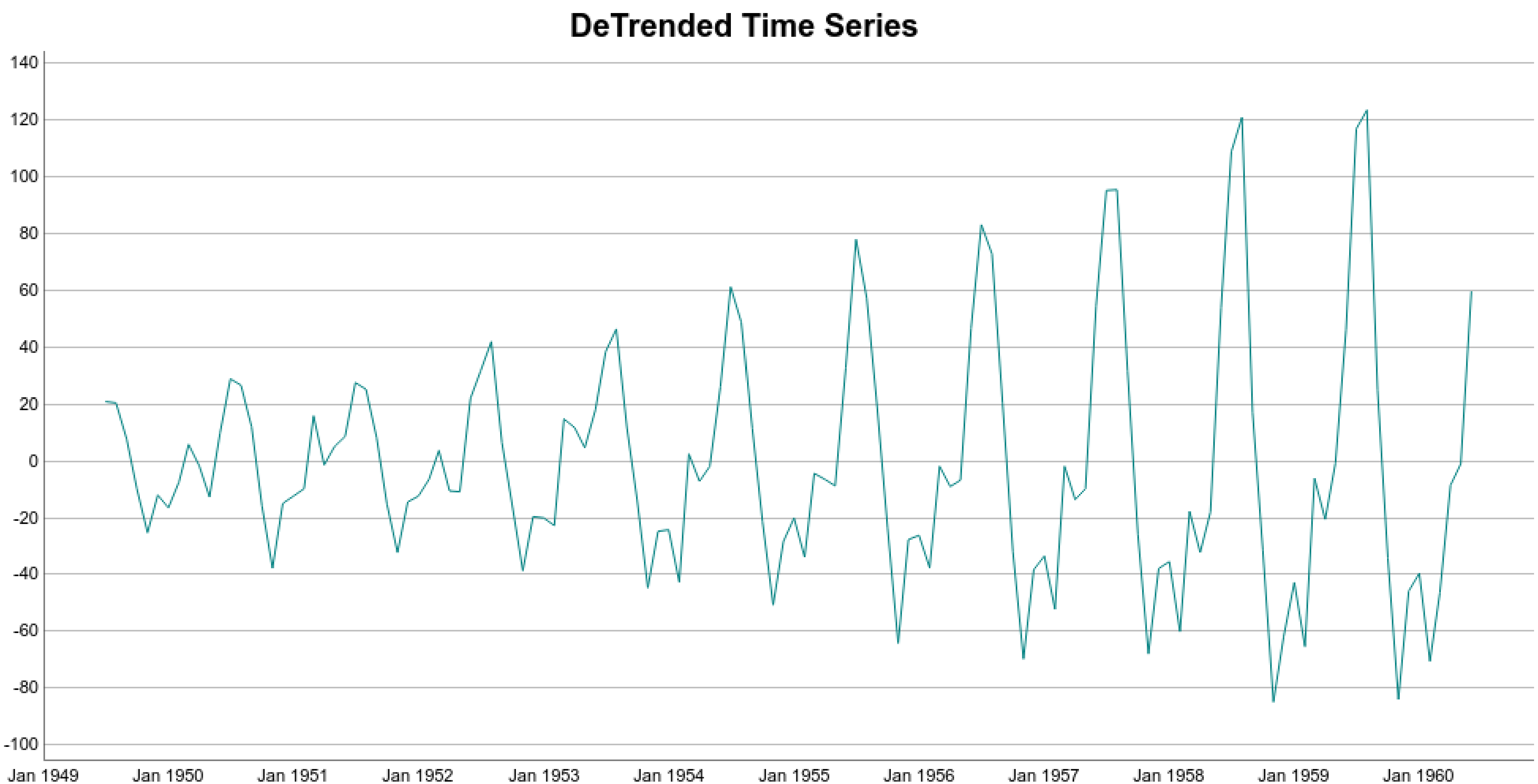
- **We can use the ma() function** from the forecast package. One important tip is to define the order parameter equal to the frequency of the time series. In this case, frequency is 12 because the time series contains monthly data.

We can see the trend over the original time series below:

STEP1 DeTRend(Is DeTrend and Smoothening): Once we have the trend component, we can use it to remove the trend variations from the original data. In other words, we can DeTrend the time series by subtracting the Trend component from it.

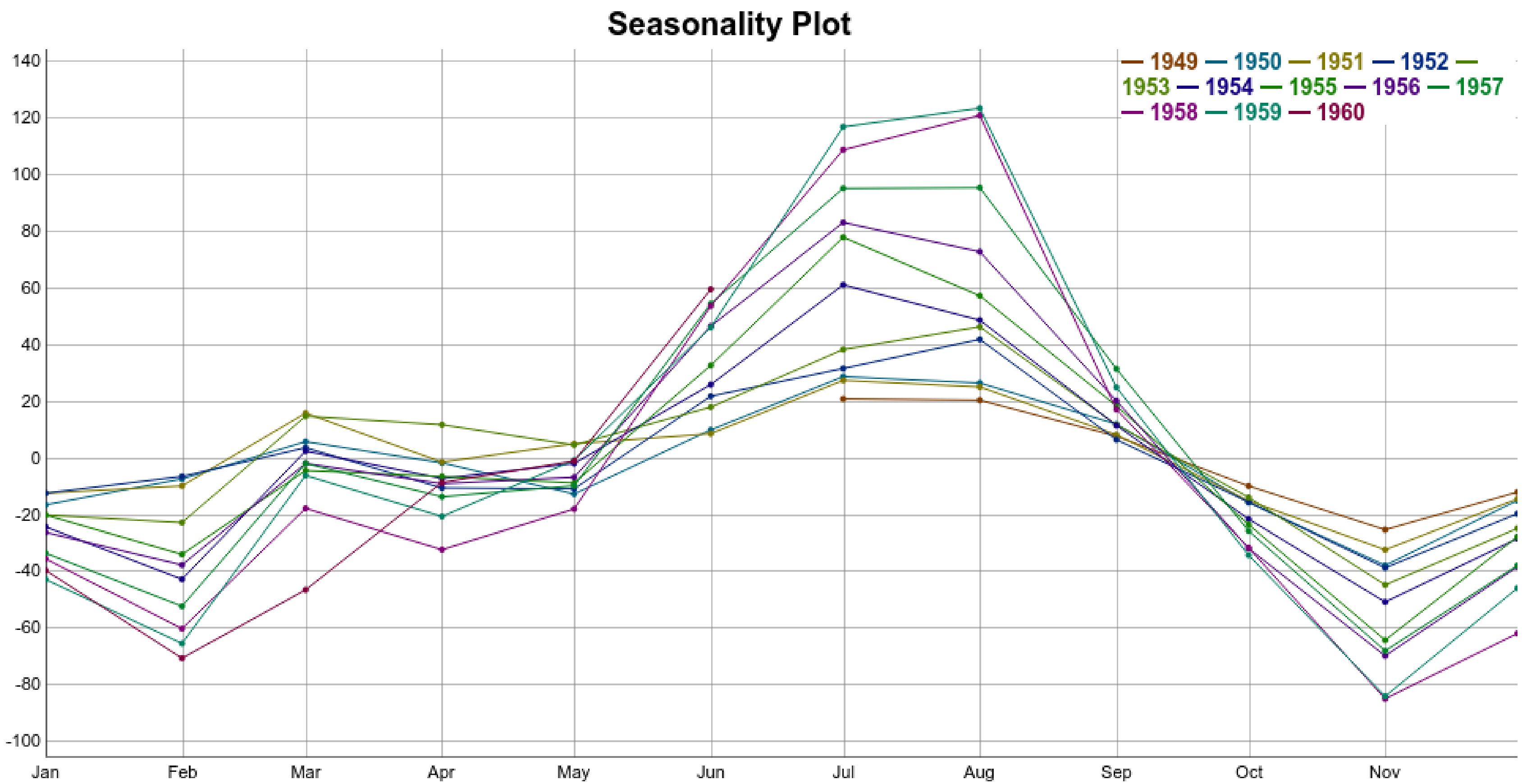
$$detrend = timeseries - trend$$

We can see that seasonality occurs regularly. Moreover, its magnitude increases over time.



STEP2 Extract Seasonality(Average the Monthly DeTrended Data): The next step to extract the seasonal component of the data. The approach here is to average the monthly DeTrended data for every year. This can be seen more easily in a seasonality plot.

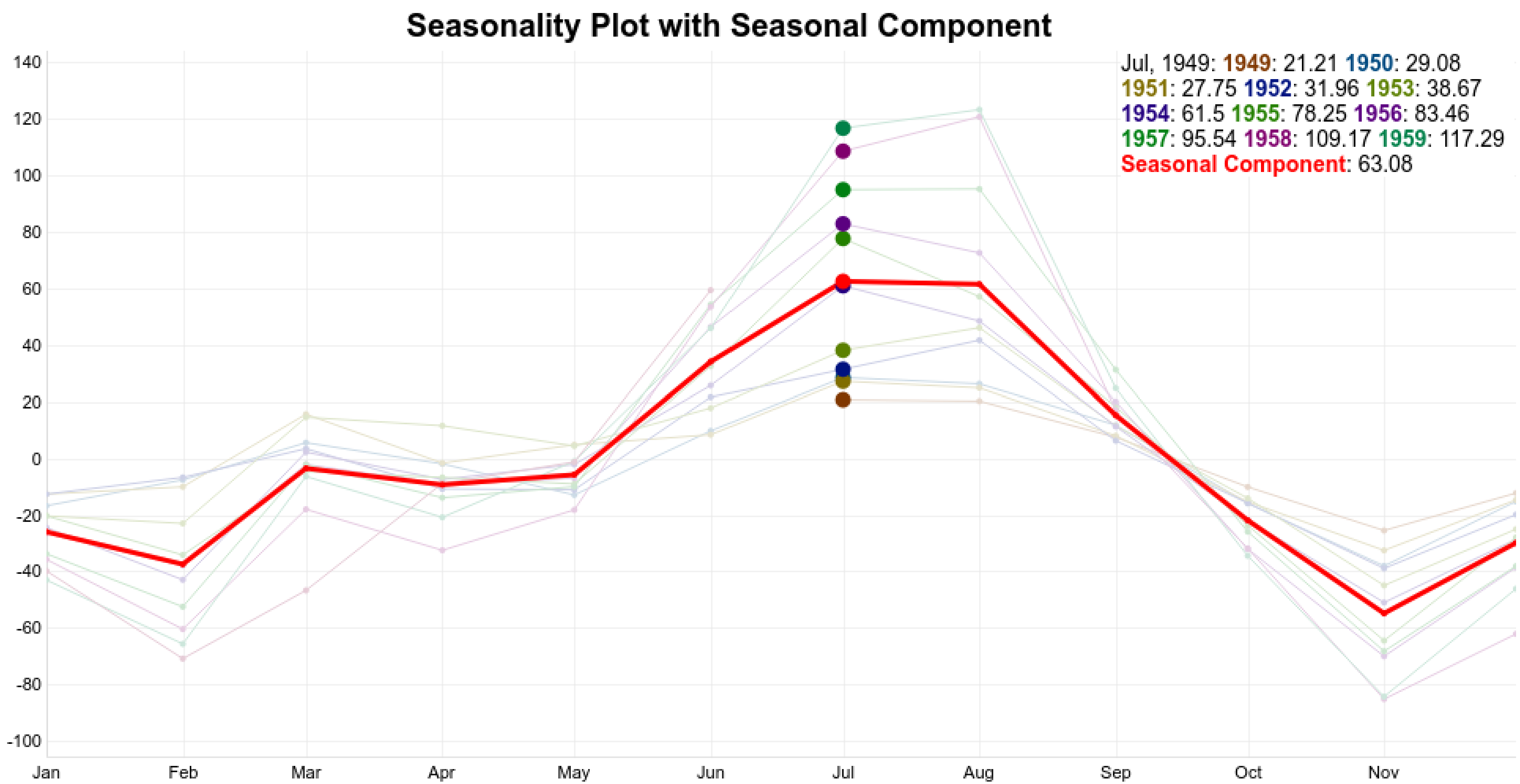
Each line in the seasonality plot corresponds to a year of data (12 data points). Each vertical line groups data points by their frequency. In this case, the x-axis groups data points for each specific month. For instance, for our dataset, the seasonal component for February is the average of all the detrended February values in the time series.



do it. enjoy it.

STEP2 Extract Seasonality: To extract the seasonality, we have to average the data points for each month. You can see the result in this plot:

The seasonality component is in **RED**. The idea is to use this pattern repeatedly to explain the seasonal variations on the time series.

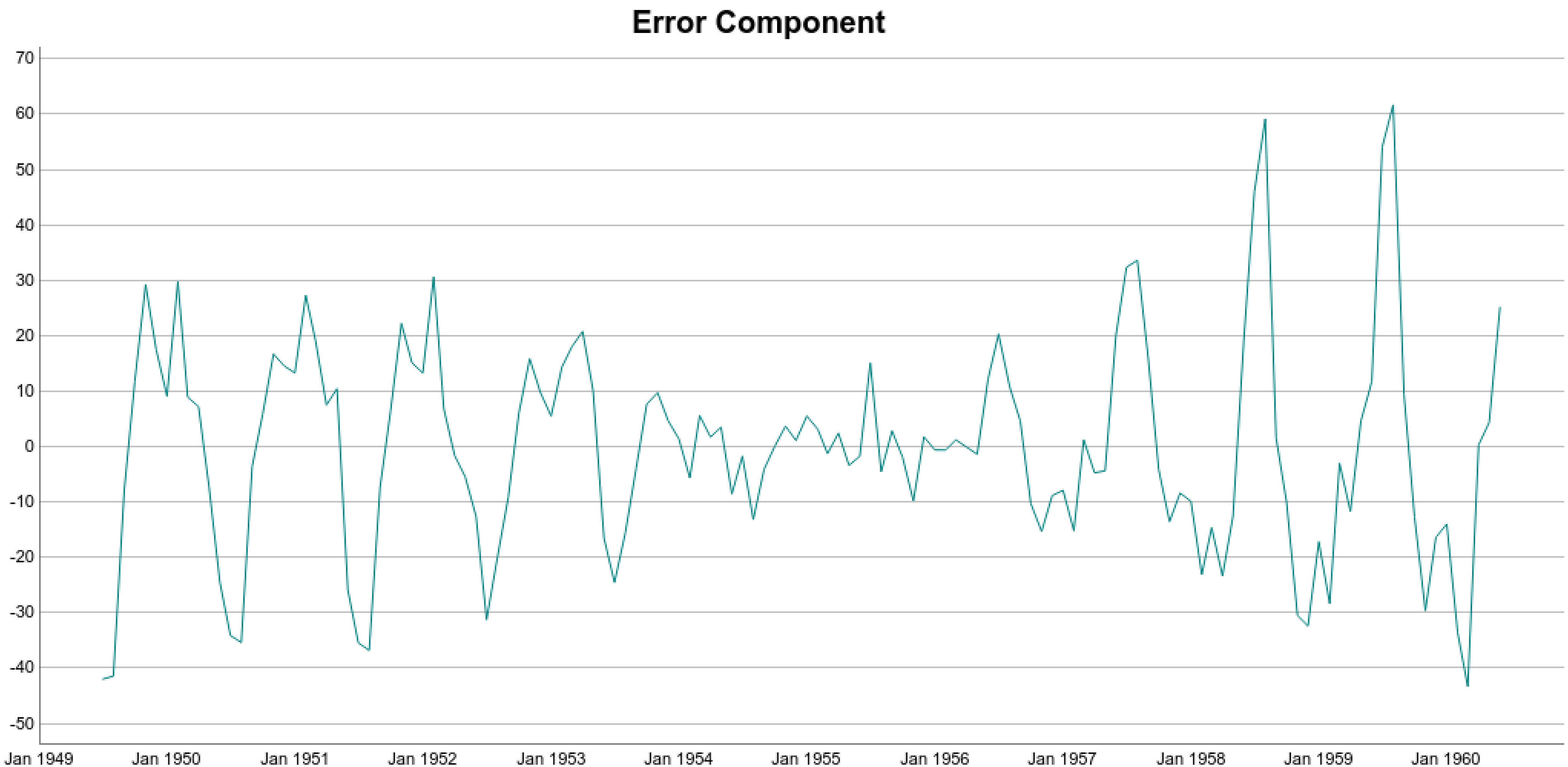


RESIDUALS

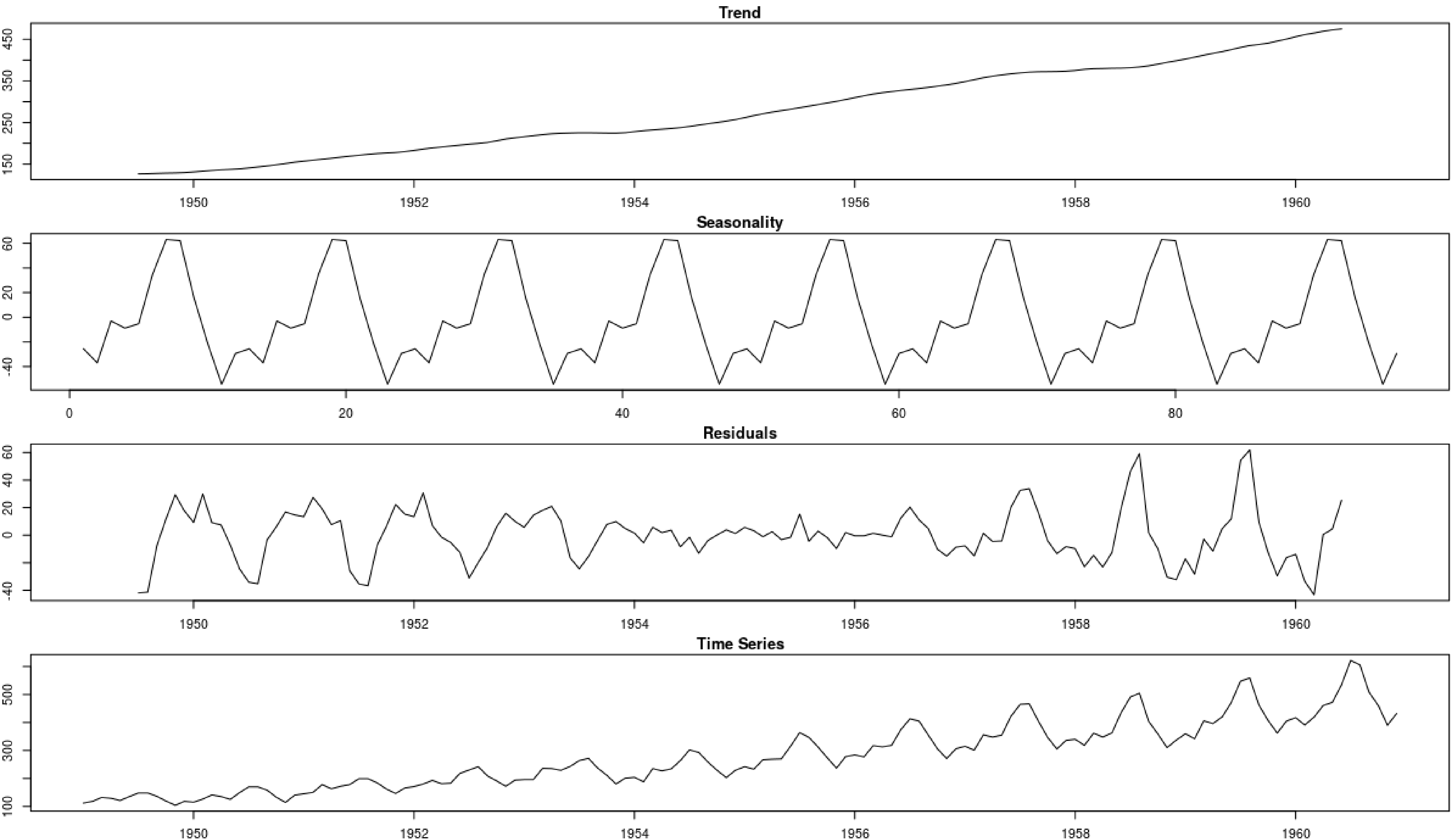
Ideally, trend and seasonality should capture most of the patterns in the time series.

The **residuals** represent what's left from the time series, **after trend and seasonal have been removed from the original signal.** We want the residuals to be as small as possible.

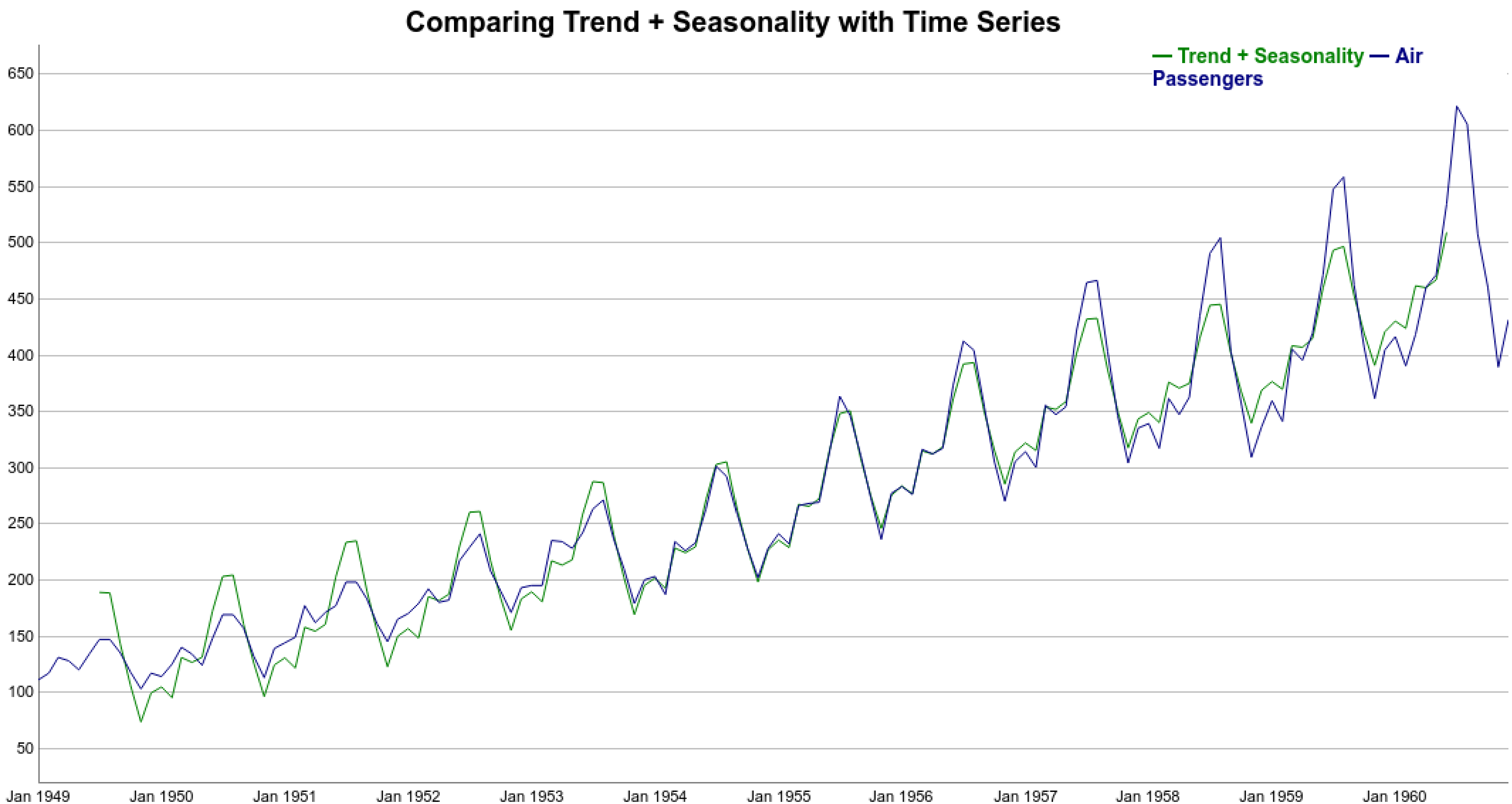
$$residuals = timeseries - trend - season$$



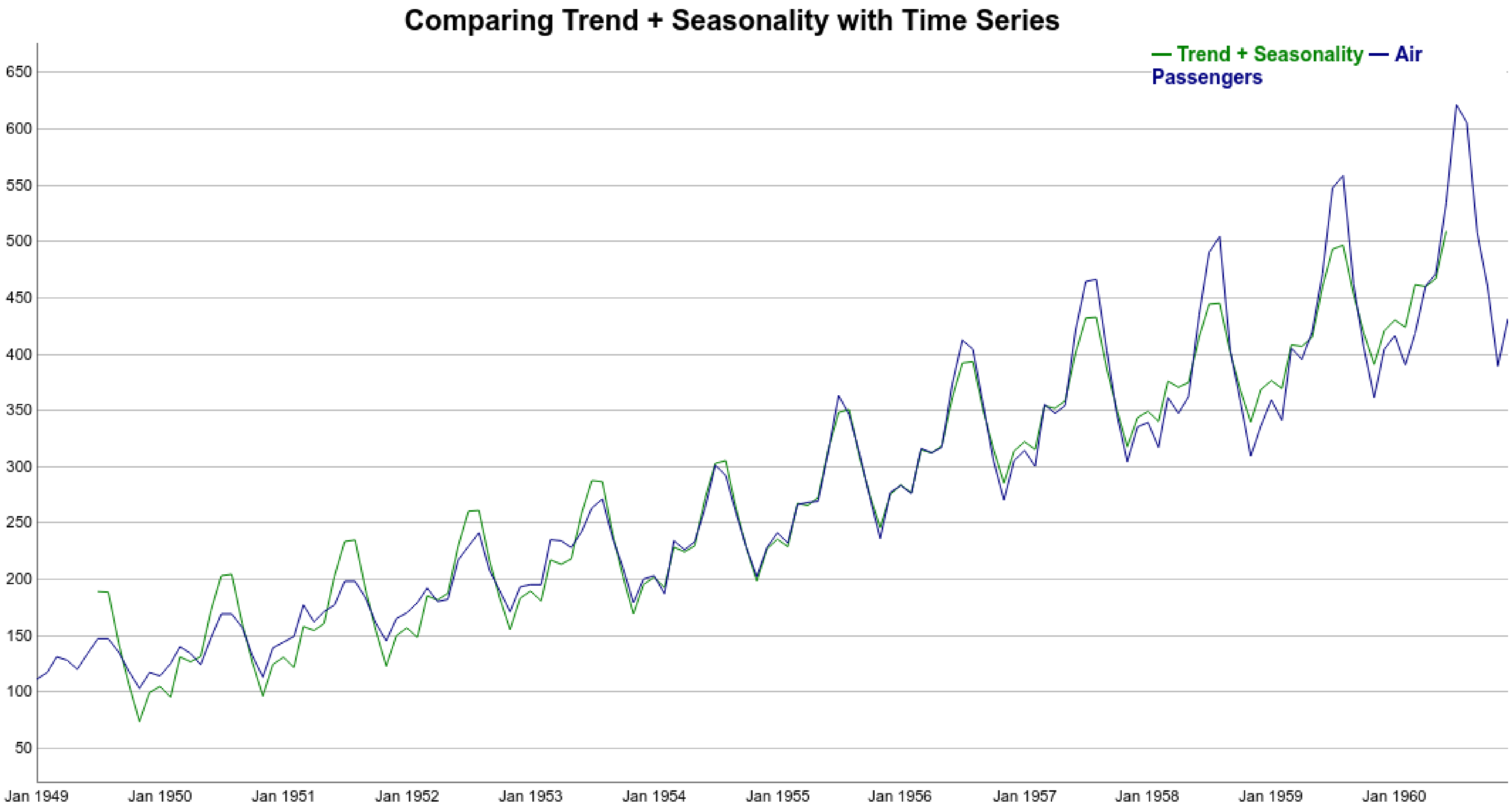
SEASONALITY PLOT (AFTER REMOVAL OF TREND, SEASON



Reconstruct the time series using the trend and seasonal components:

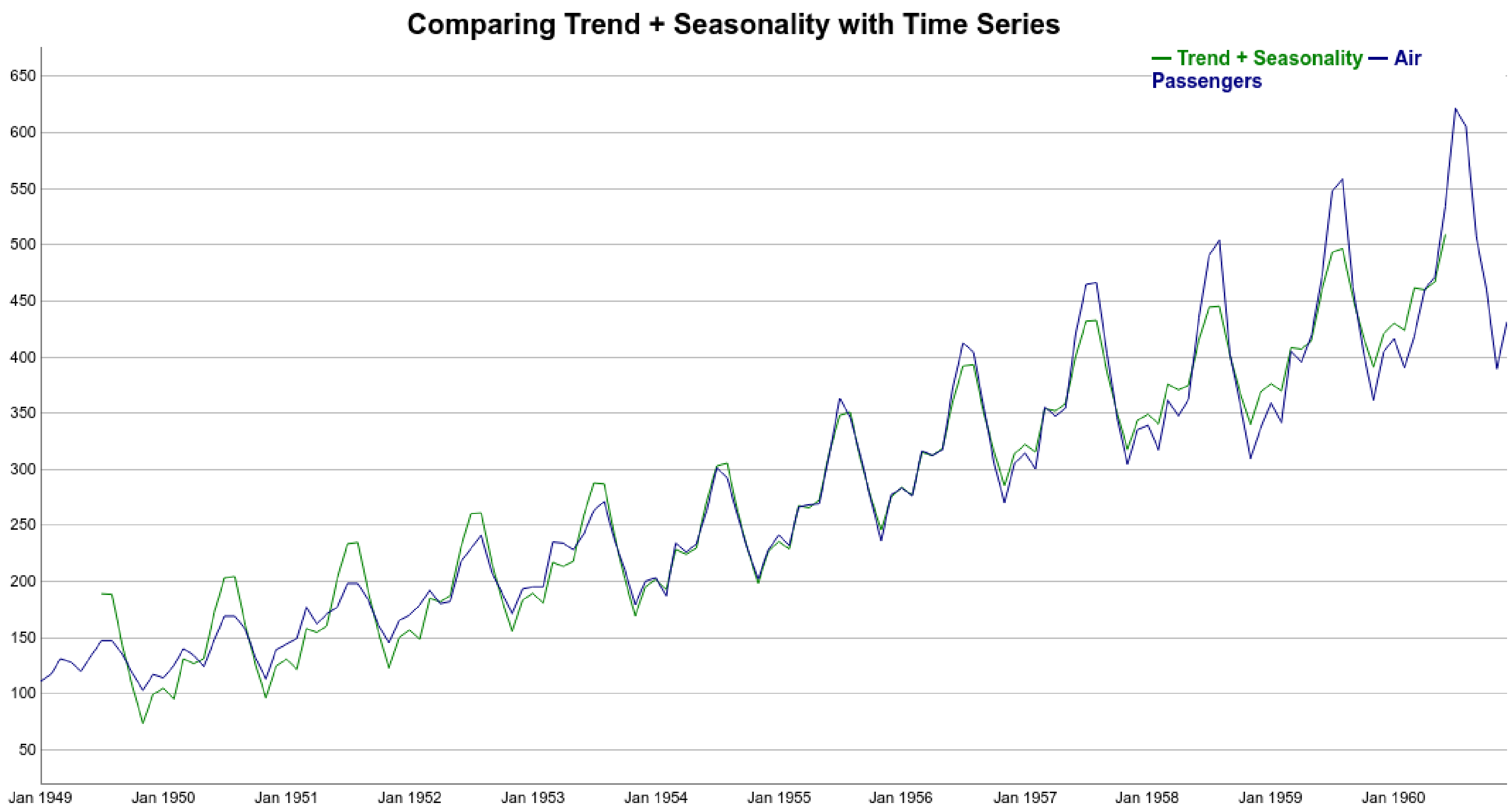


The new reconstructed signal (trend+seasonality) follows the increasing pattern of the air passenger data but in someplaces it is clear that the seasonal variations are not well represented. There's a discrepancy between the reconstructed (trend+seasonality) and original air passengers time series.



Additive decomposition assumes seasonal patterns as periodic. In other words, the seasonal patterns have the same magnitude every year and they add to the trend.
But if we take a look again at the detrend time series plot, we see that it's not true. More specifically, the seasonal variations increase in magnitude over the years. In other words the number of passengers increased, year after year.

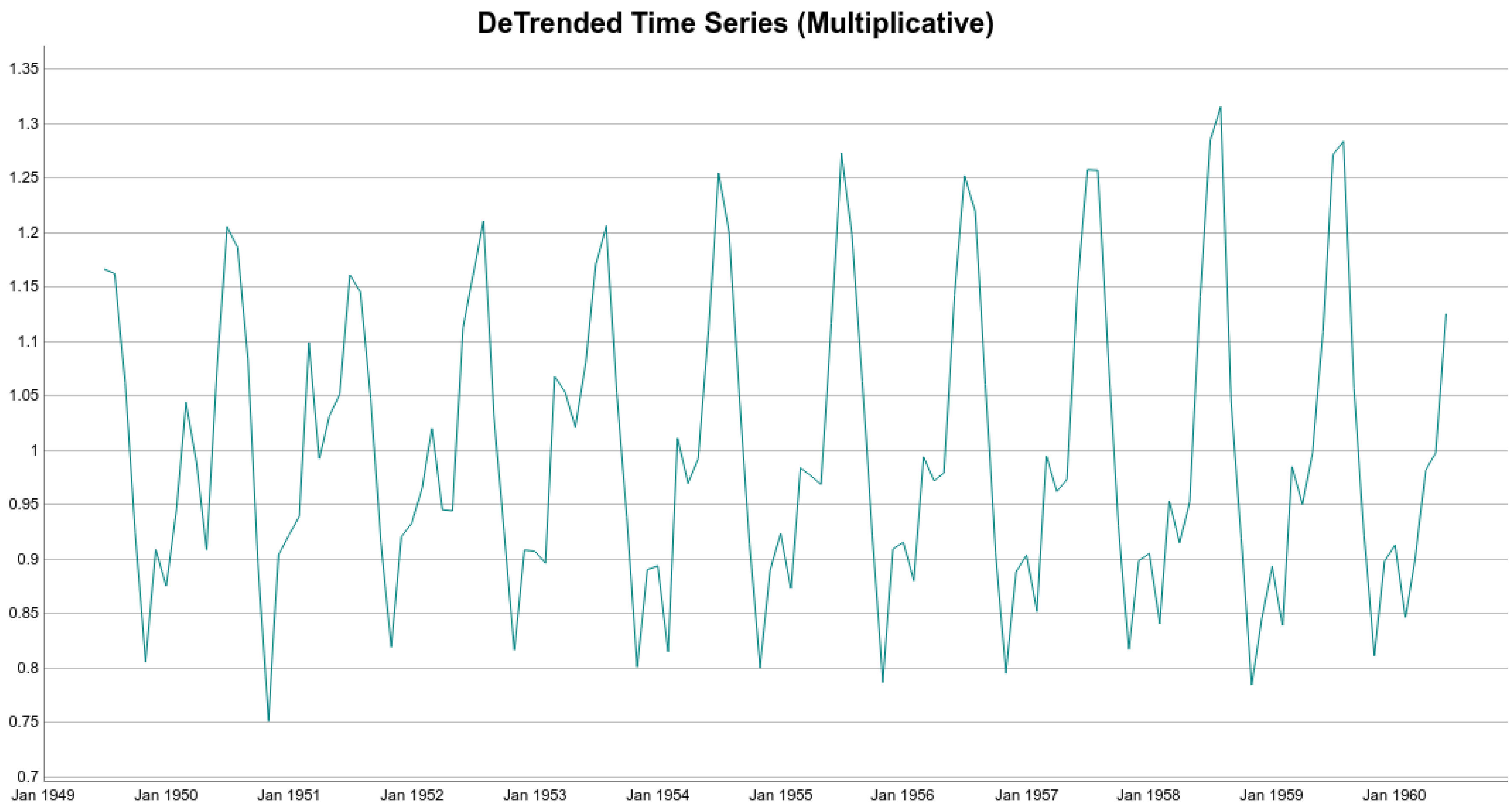
WE will try to overcome these problems using Multiplicative Decomposition



Multiplicative Decomposition

For multiplicative decomposition, the trend does not change. It’s still the centered moving average of the data. However, to detrend the time series, instead of subtracting the trend from the time series, we divide it:

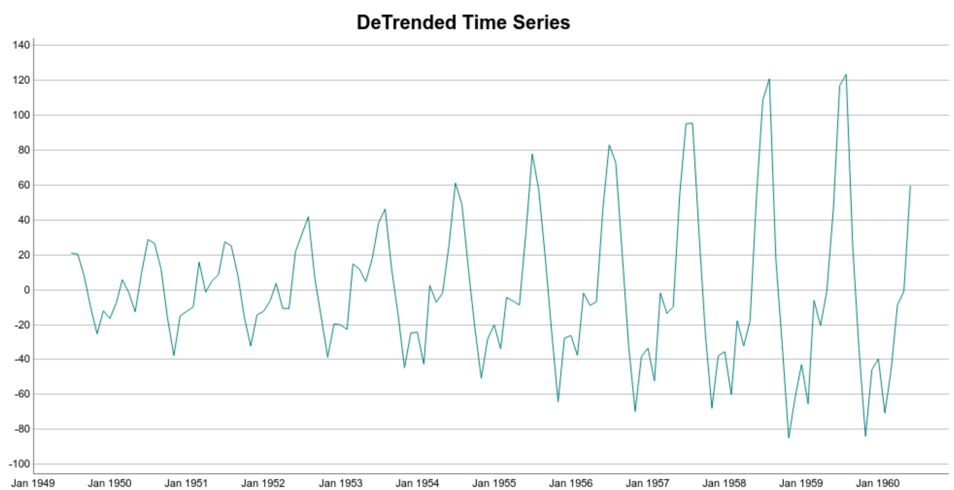
$$detrend = \frac{timeseries}{trend}$$

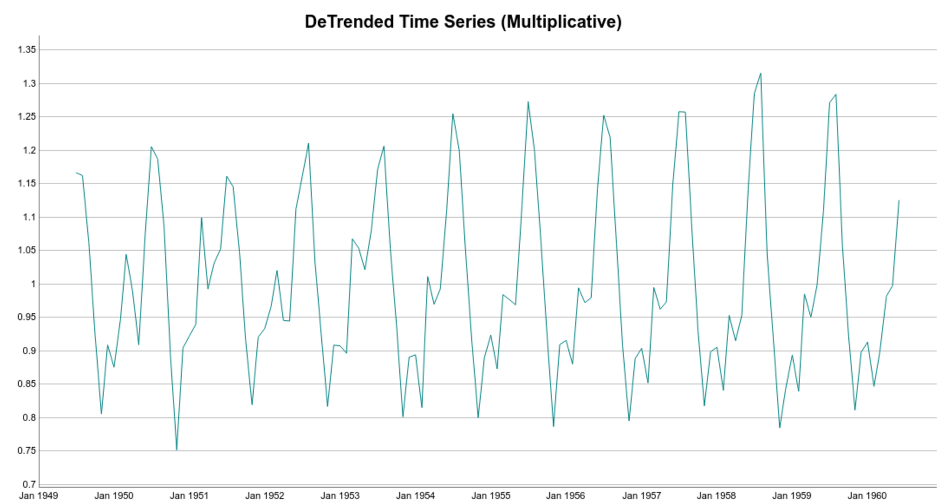


Multiplicative Decomposition

Note the difference between the detrended data for additive and multiplicative methods. For additive decomposition, the detrended data is centered at zero. That's because adding zero makes no change to the trend.

On the other hand, for the multiplicative case, the detrended data is centered at one. That follows because multiplying the trend by one has no effect either.

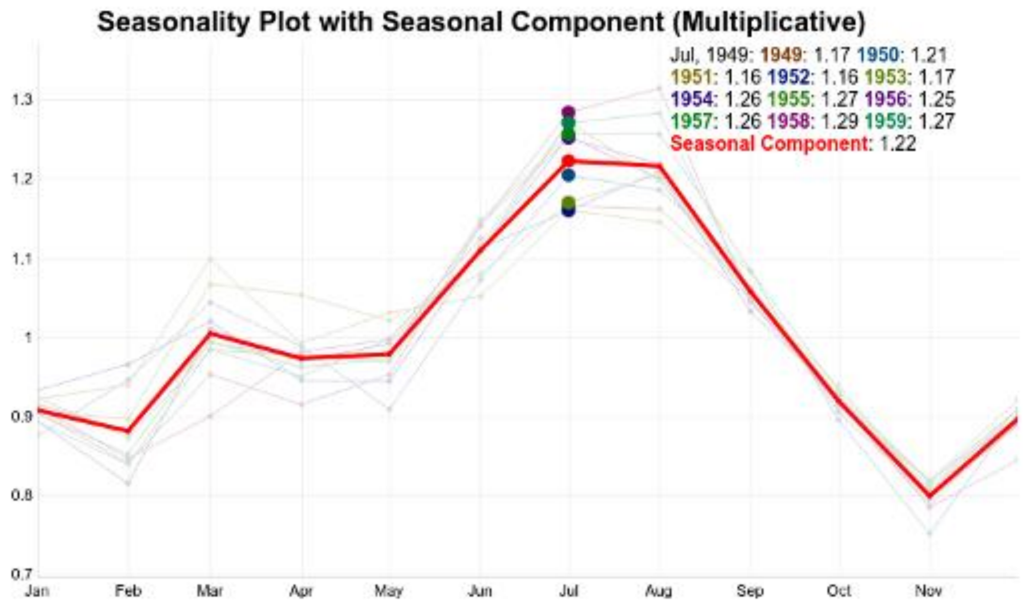




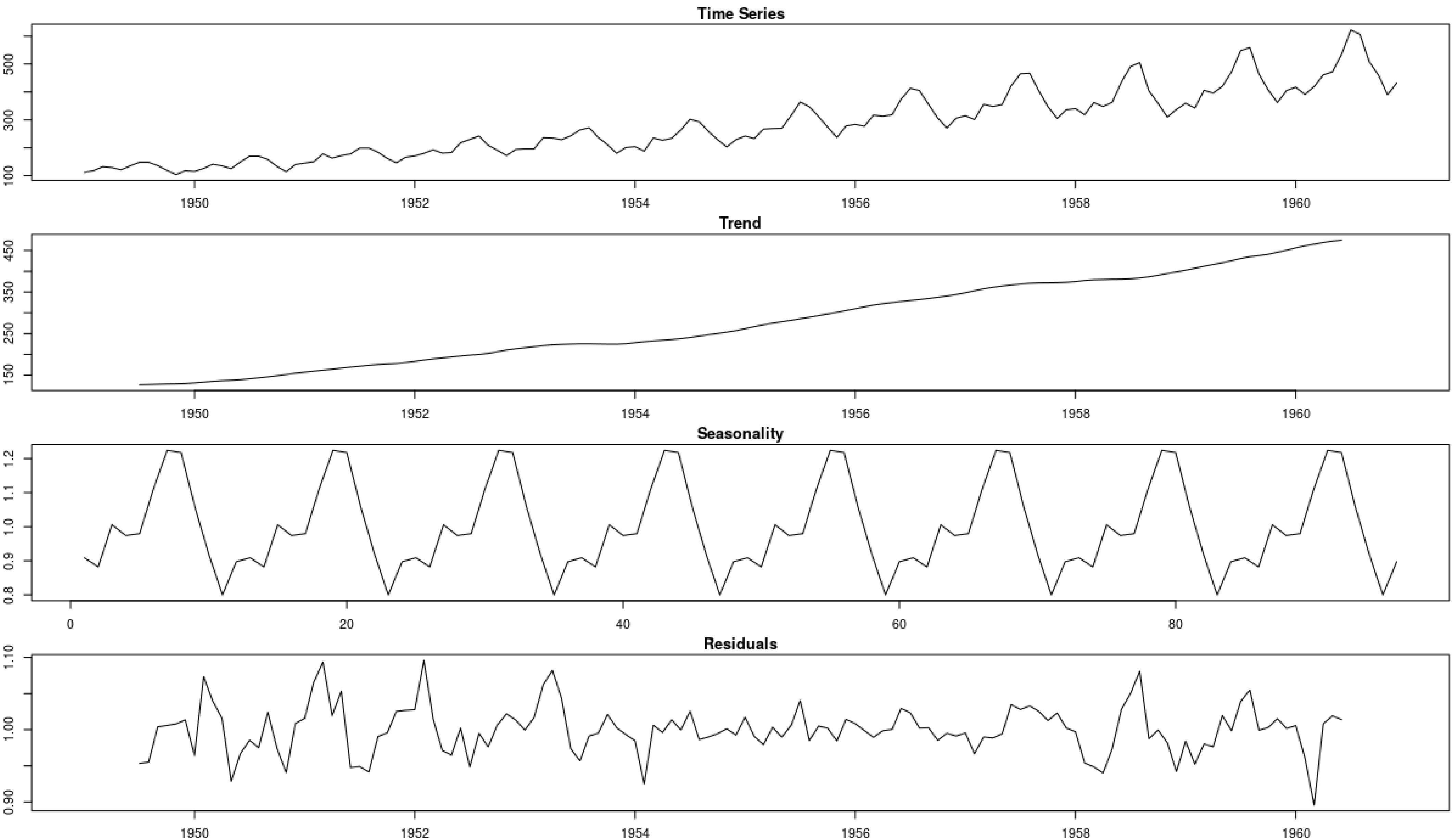
Do It Skills
do it. enjoy it.

Multiplicative Decomposition - Residuals

we can find *Seasonalilty* and *Error Residual* components: $detrend = \frac{timeseries}{trend}$



Below is complete multiplicative decomposition plot below. Since it is a multiplicative model, note that seasonality and residuals are both centered at one (instead of zero).

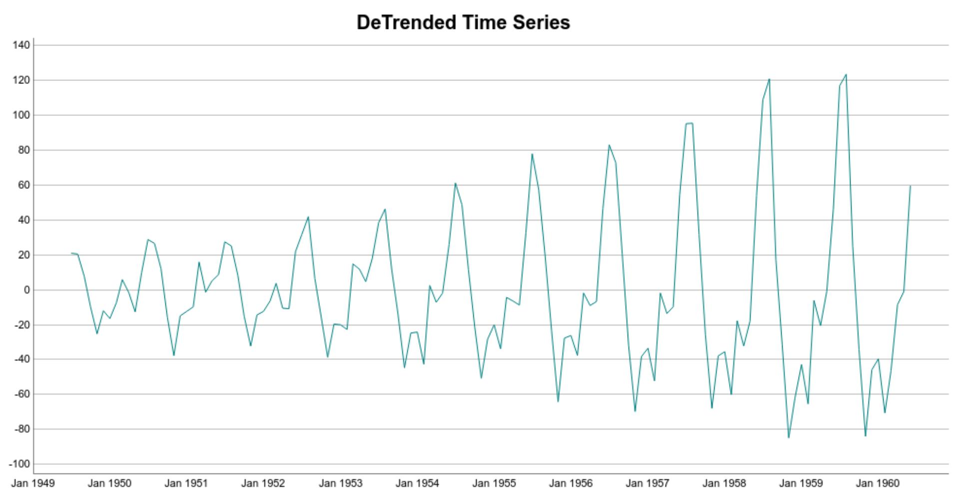


Multiplicative Decomposition

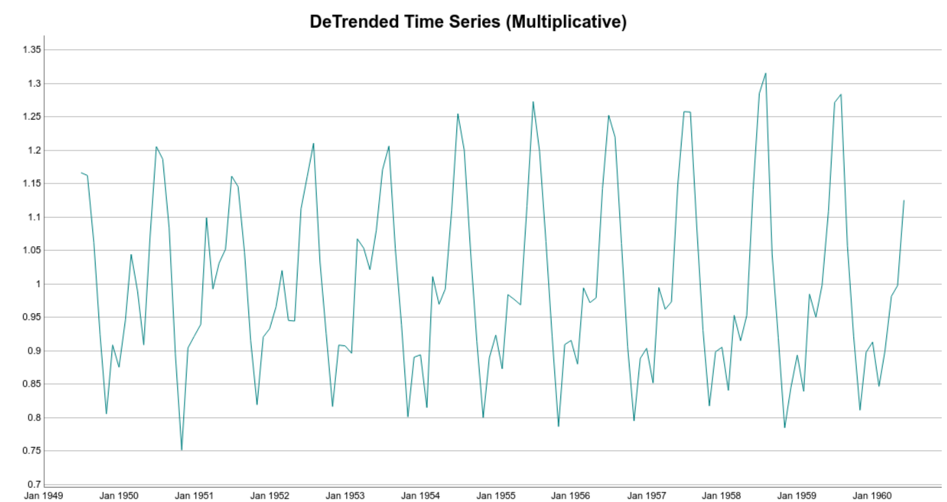
Note the difference between the detrended data for additive and multiplicative methods. For additive decomposition, the detrended data is centered at zero. That's because adding zero makes no change to the trend.

On the other hand, for the multiplicative case, the detrended data is centered at one. That follows because multiplying the trend by one has no effect either.

Additive

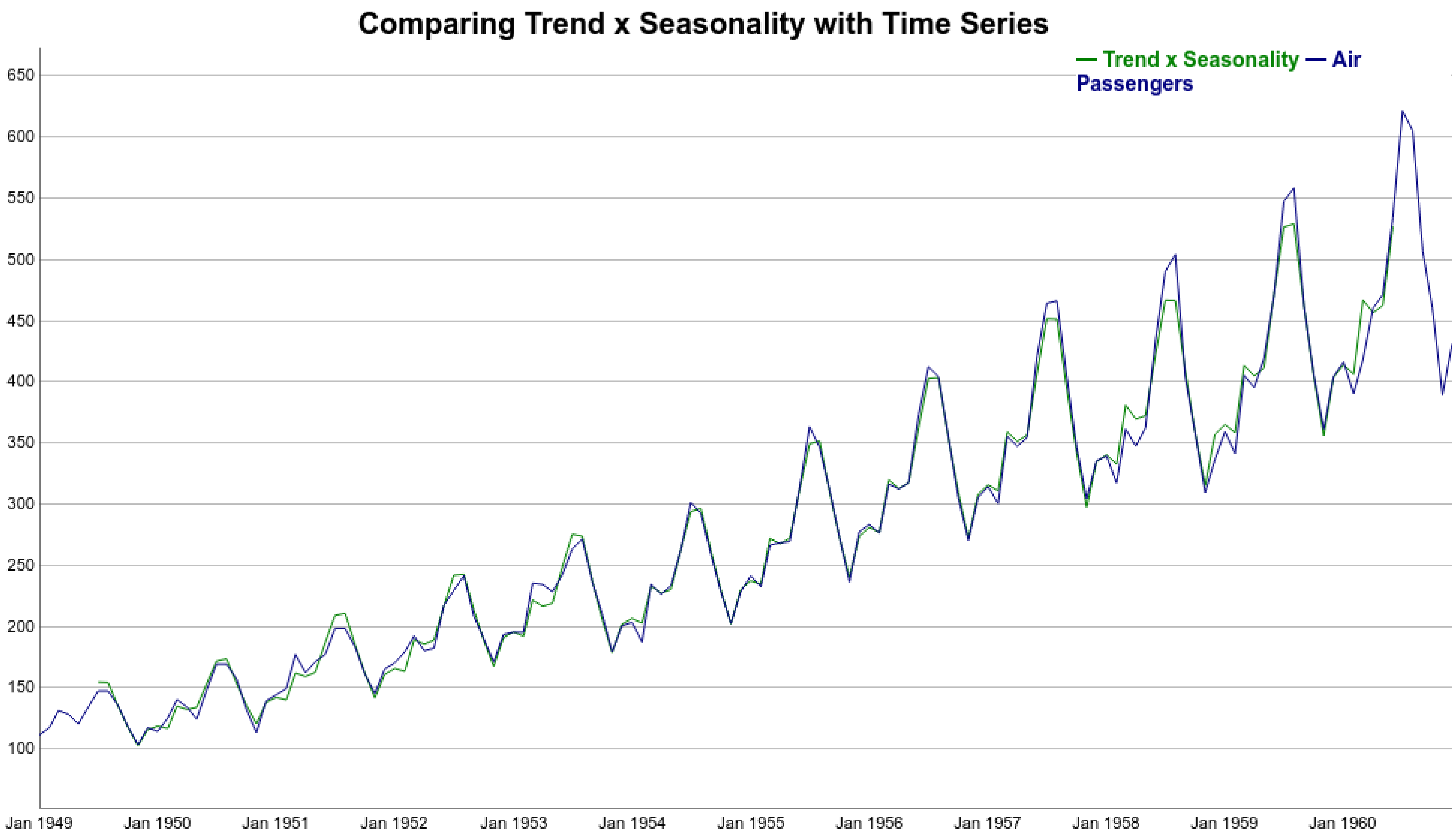


Multiplicative



Do It Skills
do it. enjoy it.

Multiplicative Decomposition

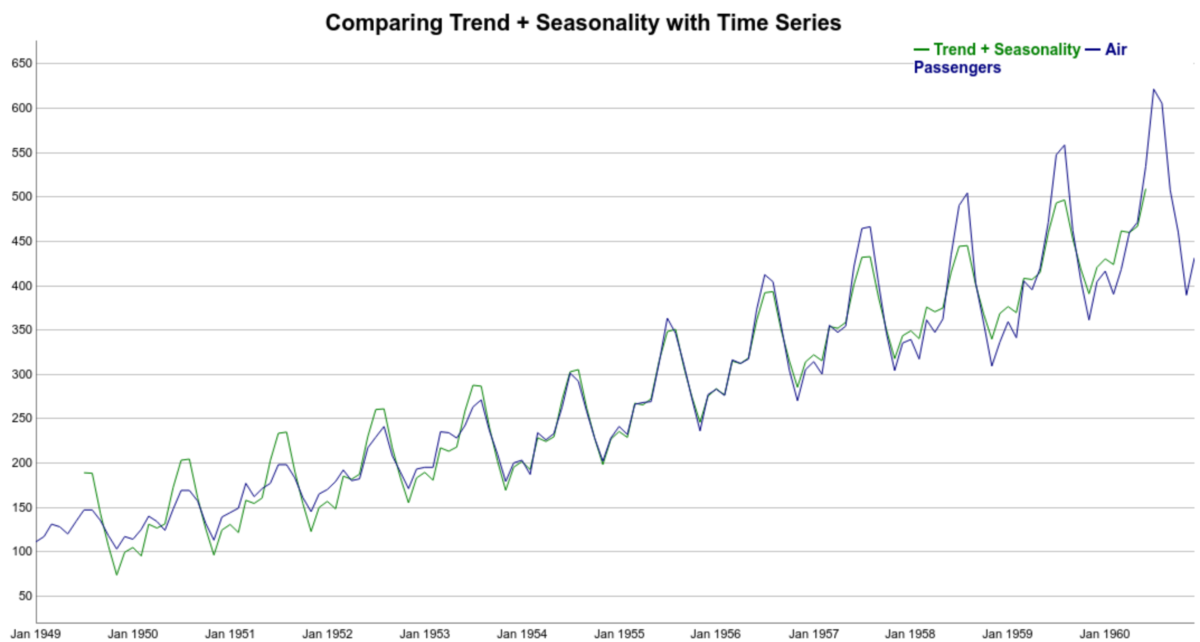


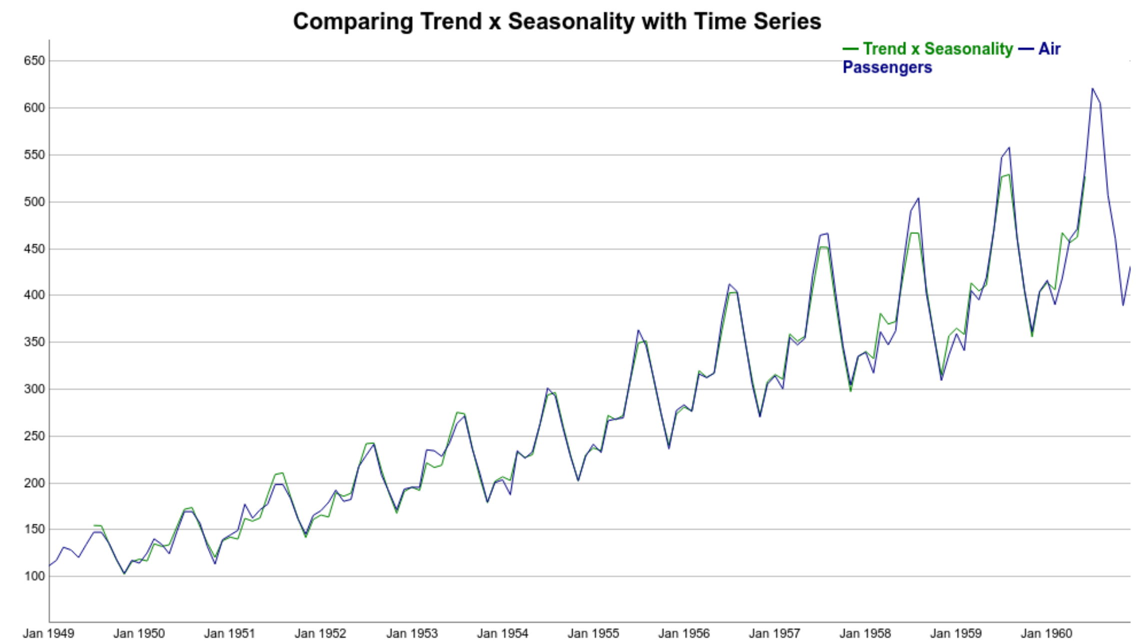
Multiplicative Decomposition

We reconstruct the time series using the Trend and Seasonal components. The multiplicative model better explains the variations of the original time series.
Here the seasonal component instead of adding to the trend, the seasonal component multiplies the trend.

Note how the seasonal swings follow the ups and downs of the original series. Also, if we compare additive and multiplicative residuals, we can see that the later is much smaller.

As a result, a multiplicative model (Trend x Seasonality) fits the original data much more closely.

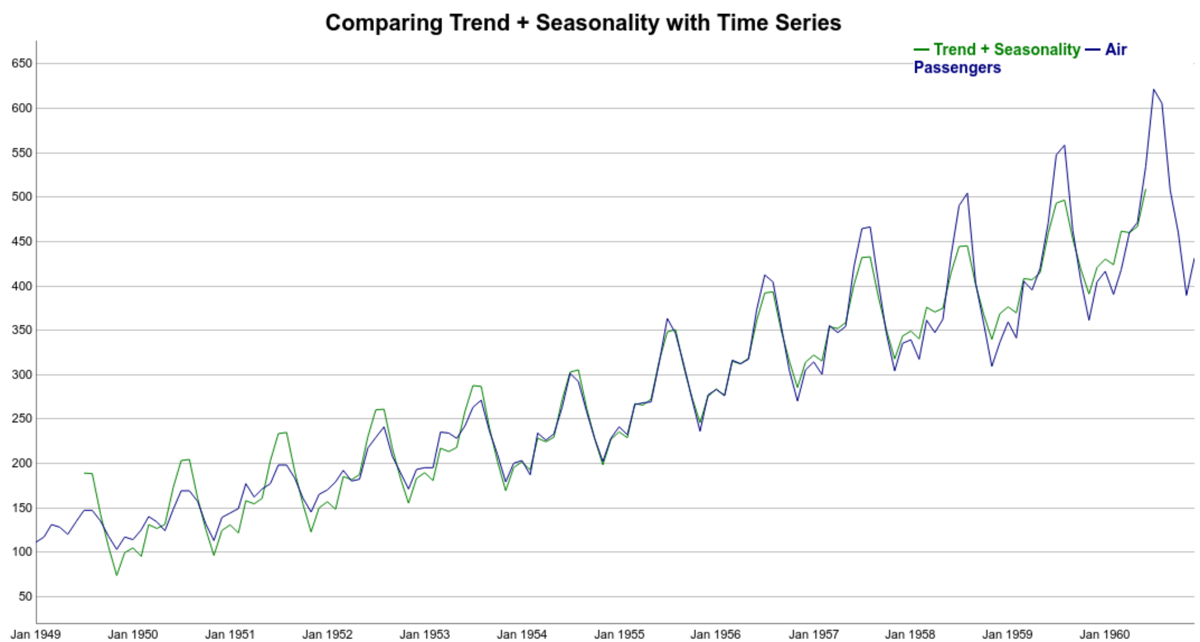


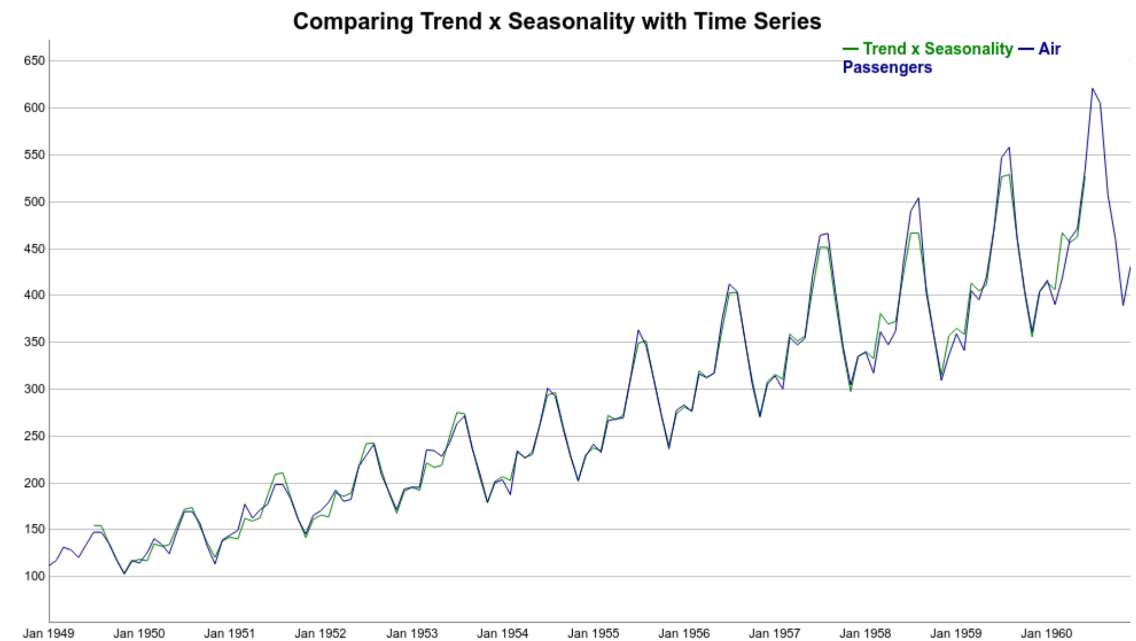


Disadvantages

First, using a moving average to estimate the trend+cycle component has some disadvantages. Specifically, this method creates missing values for the first few and last values of the series. For monthly data (frequency equal 12), we will not have estimates for the first and last six months. That is depicted on the Trend figure below.

The estimation of the seasonal pattern is assumed to repeat every year. This can be a problem for longer series where the patterns might change. You can see this assumption on both decomposition plots. Note how the additive and multiplicative seasonal patterns repeat over time.

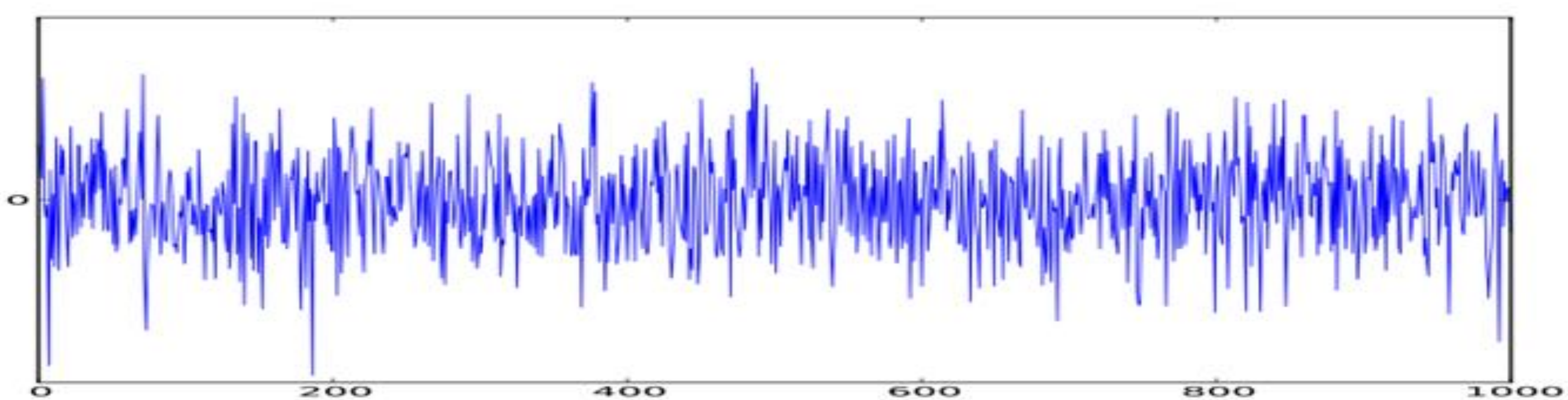




WHITE NOISE

WHITE NOISE (WHAT IS WHITE NOISE?)

- White Noise is a kind of Irregularity. Because there is NOISE in the Data our forecasting will not be good.
- Describes the assumption that each element in a series is a random draw from a population
- White noise for time t is denoted by W(t)
- If the value W(t) for any time t is a random variable that is statistically independent of its entire history before t
- Zero mean and constant variance
- Autoregressive (AR) and moving average (MA) models correct for violations of this white noise assumption



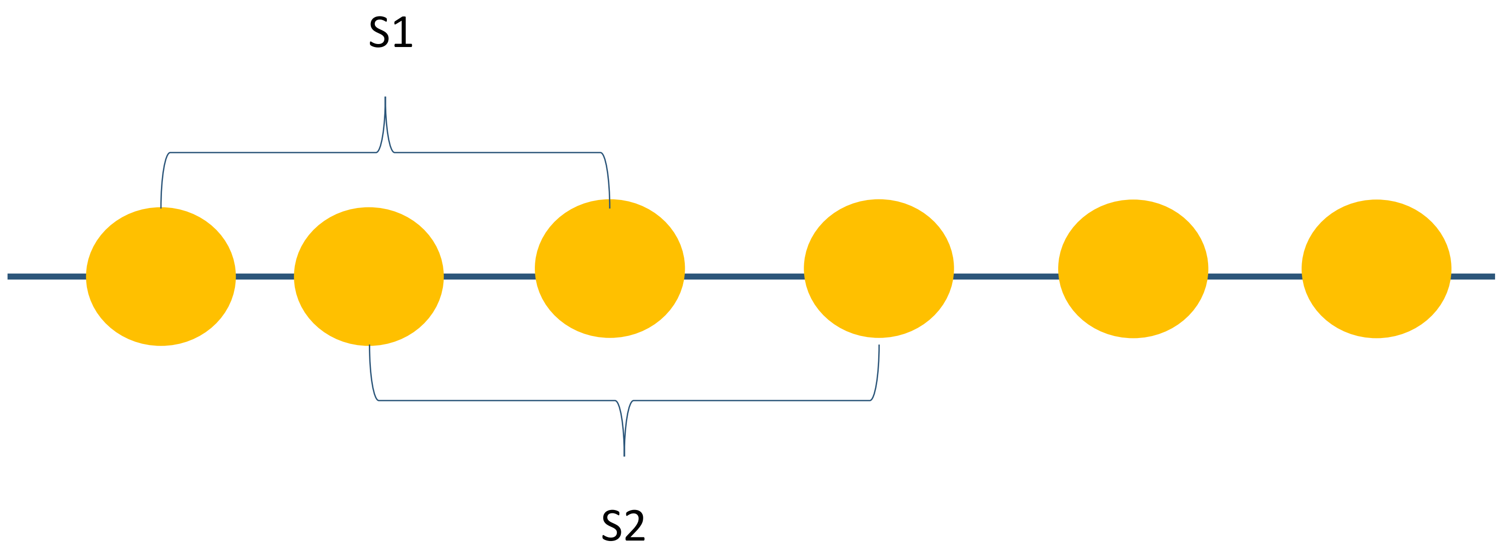
White Noise

Describes the assumption that each element in a series is a random draw from a population

- Zero mean and constant variance
- **Autoregressive** (AR) and **Moving Average** (MA) models correct for violations of this white noise assumption

STATIONARITY

What is stationarity of data?
Any stationarity is example of white noise distribution



COV(S1) = COV(S2)

What is Covariance?

Stationarity implies that taking consecutive sample of the data of same size should have identical covariance regardless of starting point.

Do It Skills
do it. enjoy it.

STATIONARITY ASSUMPTIONS

- Constant mean μ
- Constant variance σ
- Cov (X_n, X_{n+k}) = Cov (X_m, X_{m+k})

E.g. :
White noise
AS it has

- $\mu = 0$
- Constant σ^2
- Cov (X_n, X_{n+k}) = Cov (X_m, X_{m+k}) $\sigma_1 \sigma_2$ =0

As White noise follows all the assumptions it is completely stationary

Do It Skills
do it. enjoy it.

STATIONARITY

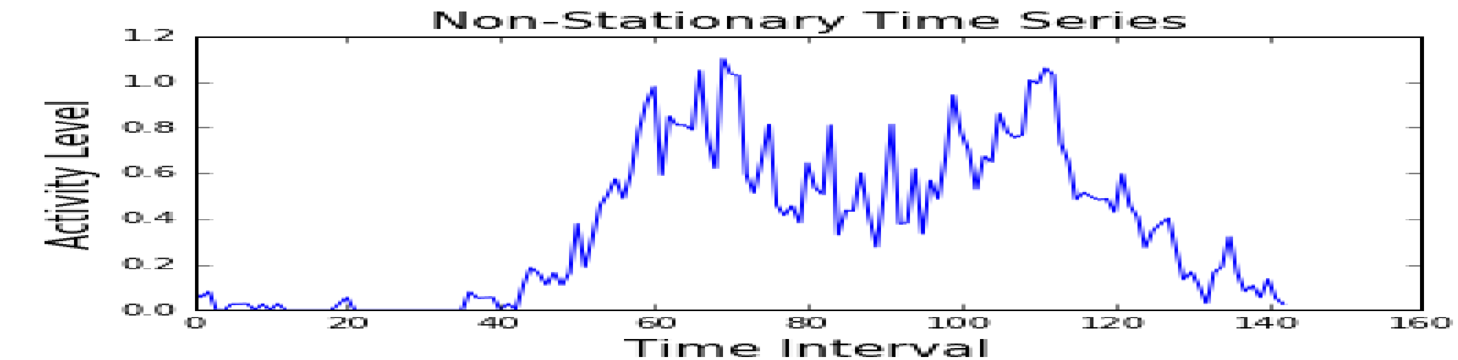
1. It is a process in which mean and variance remains unchanged or constant over the time . Such statistics are useful as descriptors of future behavior only if the **series is stationary**.

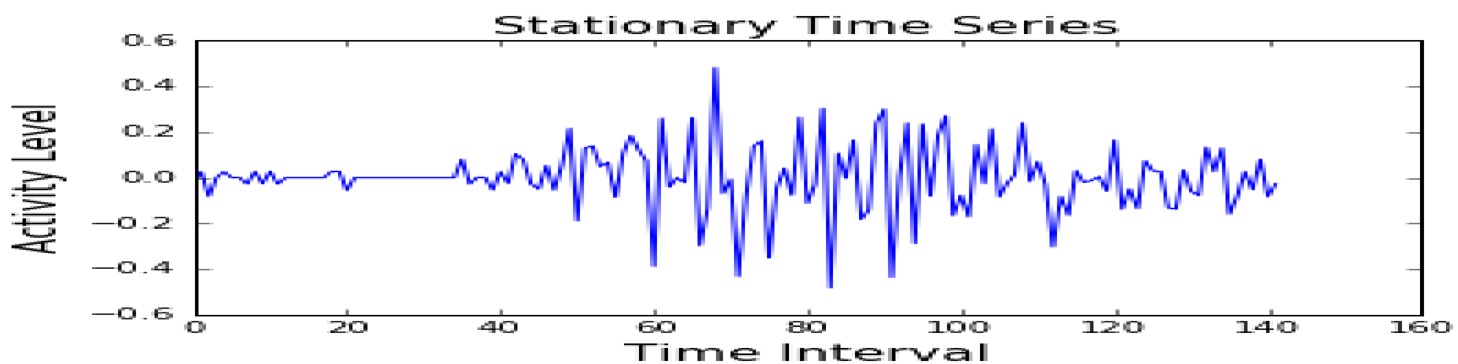
2. Using **non-stationary** time series **data** in financial models produces unreliable and spurious results and leads to poor understanding and forecasting. The solution to the **problem** is to transform the time series **data** so that it becomes **stationary**..

3. White noise is the simplest example of stationary time series, White Noise time series has zero mean, constant variance and zero covariance with lagged time series.

EXAMPLE OF STATIONARY TIME SERIES

Both are time series data. If the data is non stationary we cannot make correct predictions. So we need convert the non stationary data to stationary.





Do It Skills
do it. enjoy it.

IDENTIFYING STATIONARY TIME SERIES

METHODS

GRAPHICAL

CORRELOGRAM

ANALYTICAL

D-F-TEST

(Dickey fuller test)

WHERE DOES AUTO CORRELATION COME IN

AUTOCORRELATION

What is Auto Correlation ?

- correlation of a signal with a delayed copy of itself as a function of delay.
- The picture below illustrates the lag operation for lags 1 .

Date	Value	$Value_{t-1}$
1/1/2017	200	NA
1/2/2017	220	200
1/3/2017	215	220
1/4/2017	230	215
1/5/2017	235	230
1/6/2017	225	235
1/7/2017	220	225
1/8/2017	225	220
1/9/2017	240	225
1/10/2017	245	240

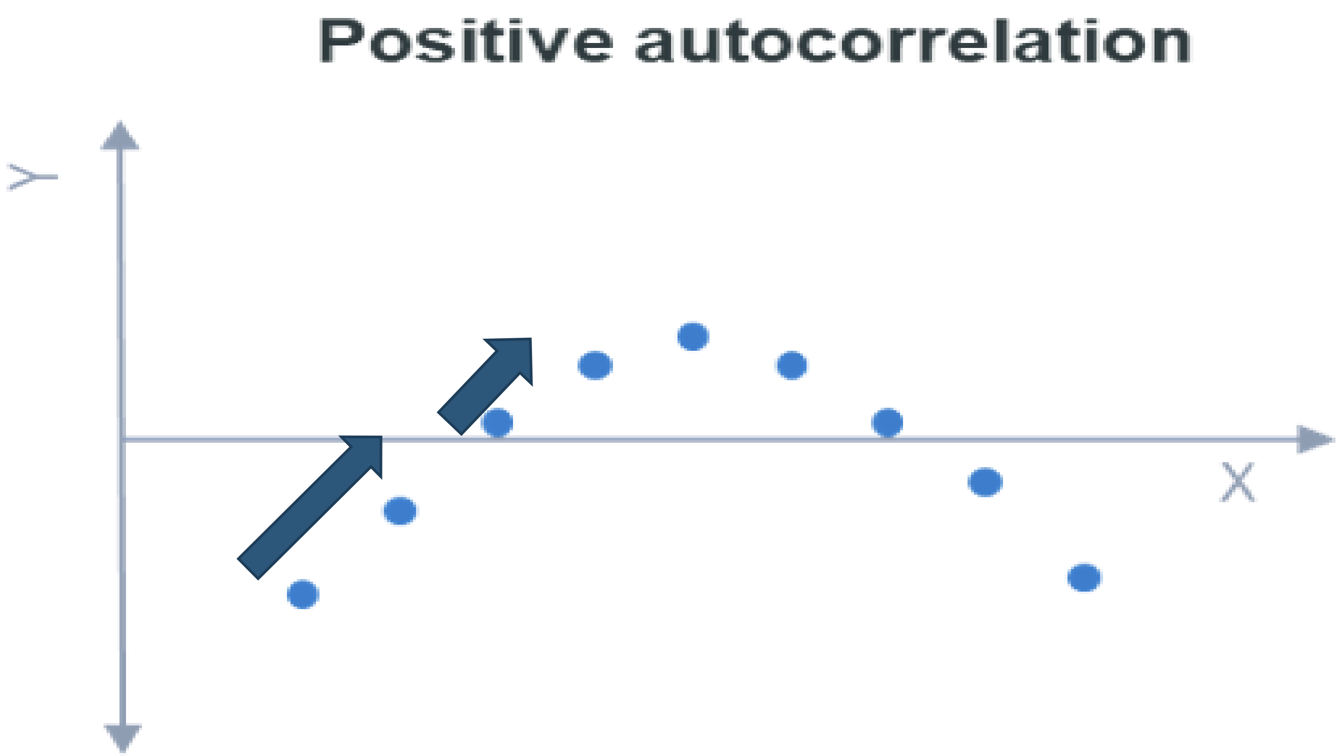
- Lag-one autocorrelation
- Also know as serial correlation

This is a linear combination of data
T-1

Auto correlation is used in the PCF and ACF models

Do It Skills
do it. enjoy it.

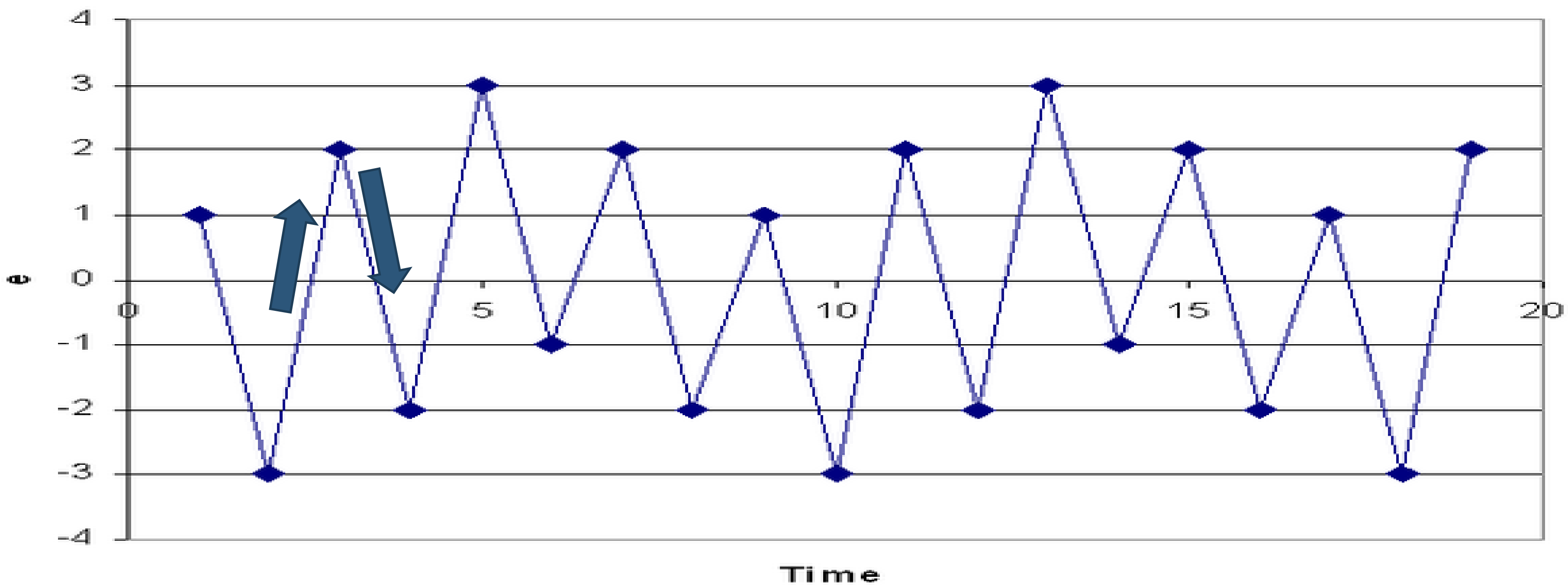
POSITIVE AUTOCORRELATION

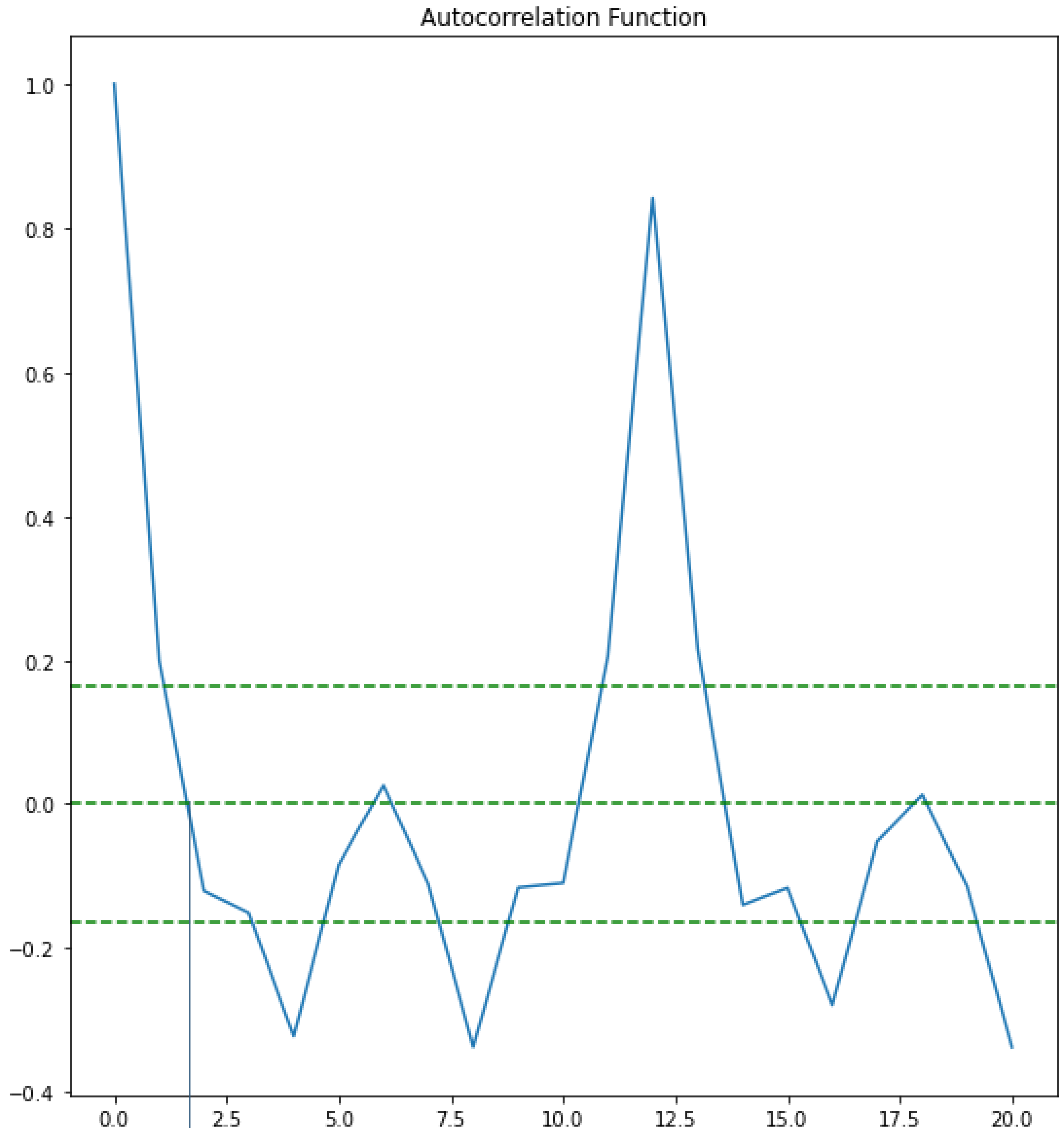


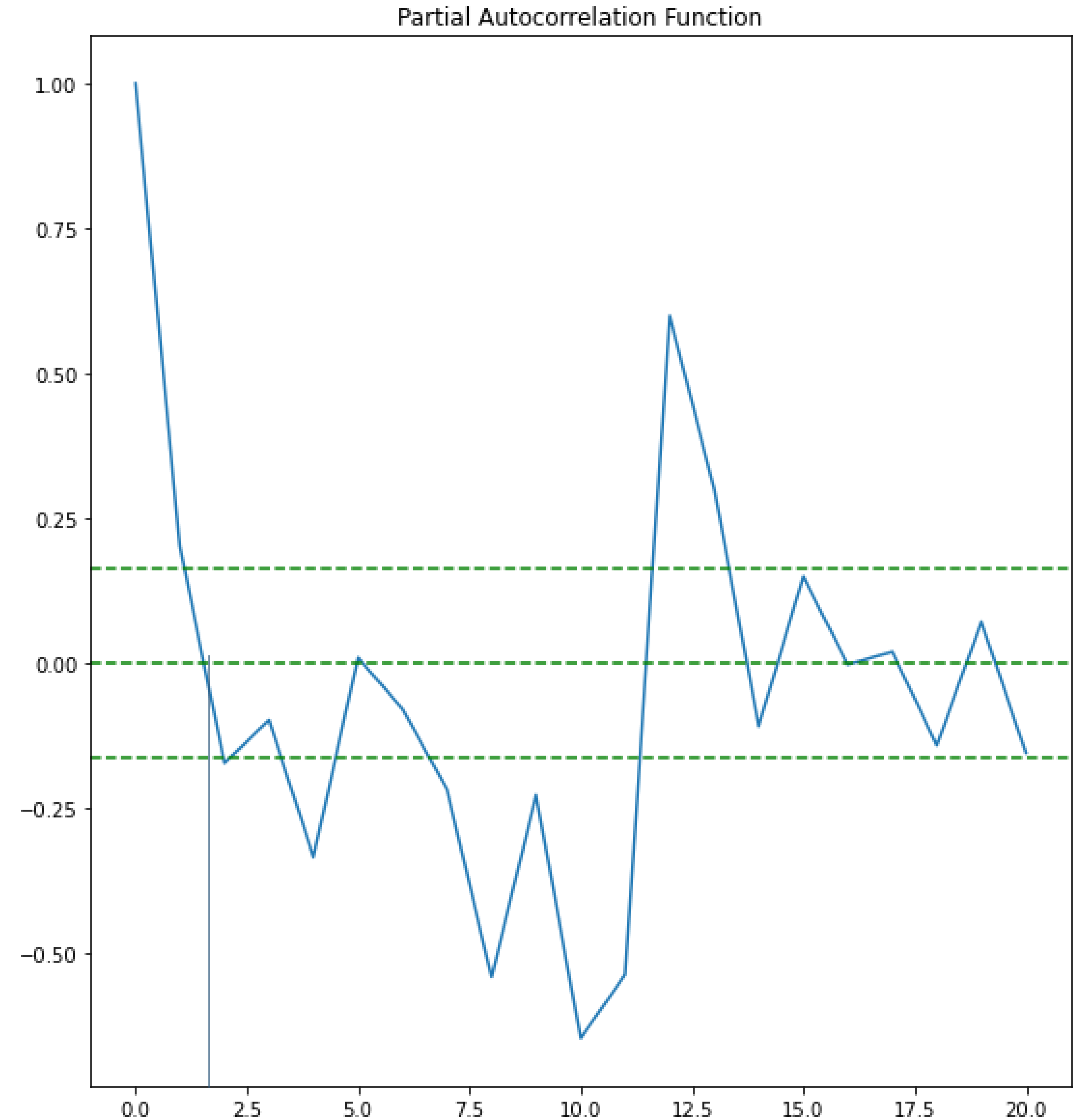
If autocorrelation is present, positive autocorrelation is the most likely outcome. *Positive autocorrelation* occurs when an error of a given sign tends to be followed by an error of the same sign. For example, positive errors are usually followed by positive errors, and negative errors are usually followed by negative errors.

NEGATIVE AUTOCORRELATION

Negative autocorrelation occurs when an error of a given sign tends to be followed by an error of the opposite sign. For instance, positive errors are usually followed by negative errors and negative errors are usually followed by positive errors.




Autocorrelation Function
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0


Partial Autocorrelation Function
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

ACF VS PACF

- The main difference is that how ACF and PACF works
- PACF cancel outs addition channel in series of lagged periods.
- PACF will directly find relation from P_{t-2} to $P_{t.}$ And will cancels out $P_{t-1.}$
- ACF find the relation for both P_{t-2} to P_{t-1} to P_t. and P_{t-2} to P_t
- However we must note that first lag in ACF and PACF must be similar.

AR MODEL

AUTO-REGRESSIVE MODEL

AUTO REGRESSION (AR):

- An autoregressive model is when a value from a time series is regressed on previous values from that same time series.
- Autoregressive also means a relationship with itself

AUTO REGRESSION (AR):

- We can predict the value for the next time step (t+1) given the observations at the last two time steps (t-1 and t-2).
- The order of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the AR model below is a second-order autoregression, written as AR(2).

$$\boxed{X_{(t+1)} = b_0 + b_1 X_{(t-1)} + b_2 X_{(t-2)} + E_{(t)}}$$

$$x_t = b_0 + b_1 x_{t-1} + \varepsilon_t$$

where:

x_t = value of time series at time t
b0 = intercept at the vertical axis (y-axis)
b1 = slope coefficient
x_{t-1} = value of time series at time $t-1$
ε_t = error term (or residual term or disturbance term)
t = time; t = 1, 2, 3…T

Do It Skills
do it. enjoy it.

AR Model

Auto Regressive Model takes the P parameters
P lags means:
Sales of 4th Jan are effecting 8 th Jan, 12th Jan …4 days lag

01 Value of a variable in one period is related to its values in previous periods

02 AR(p) is an autoregressive model with p lags: $y_t = \mu + \sum_{i=1}^{p} \gamma_i y_{t-i} + \epsilon_t$, where μ is a constant, γ_p is the coefficient for the lagged variable, y_t is the dependent variable at time t, y_{t-i} is the independent variable at previous time period and ϵ_t is the error at time t

03 AR(1) is expressed as:
$y_t = \mu + \gamma y_{t-1} + \epsilon_t = \mu + \gamma(Ly_t) + \epsilon_t$ or (1-γL)y_t= $\mu + \epsilon_t$

Mu + Y(gama)yt-1-epsilon

MOVING AVERAGE (MA):

- A moving average term in a time series model is a past error (multiplied by a coefficient).
- In MA model, the past error is nothing but white noise (wt) which is identically and independently distributed, each with a normal distribution having mean 0 and the same variance.
- The qth order moving average model, denoted by MA(q) is:

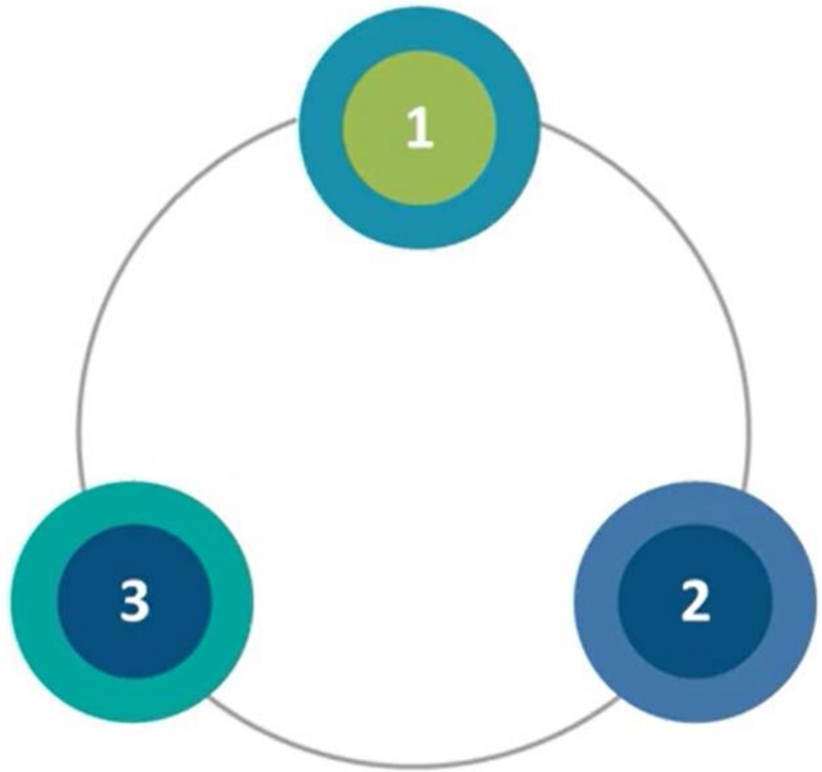
$$x_{(t)} = w_t + \theta_1 w_{(t-1)} + \theta_2 w_{(t-2)} + \ldots \theta_q w_{(t-q)}$$

Do It Skills
do it. enjoy it.

MA Model

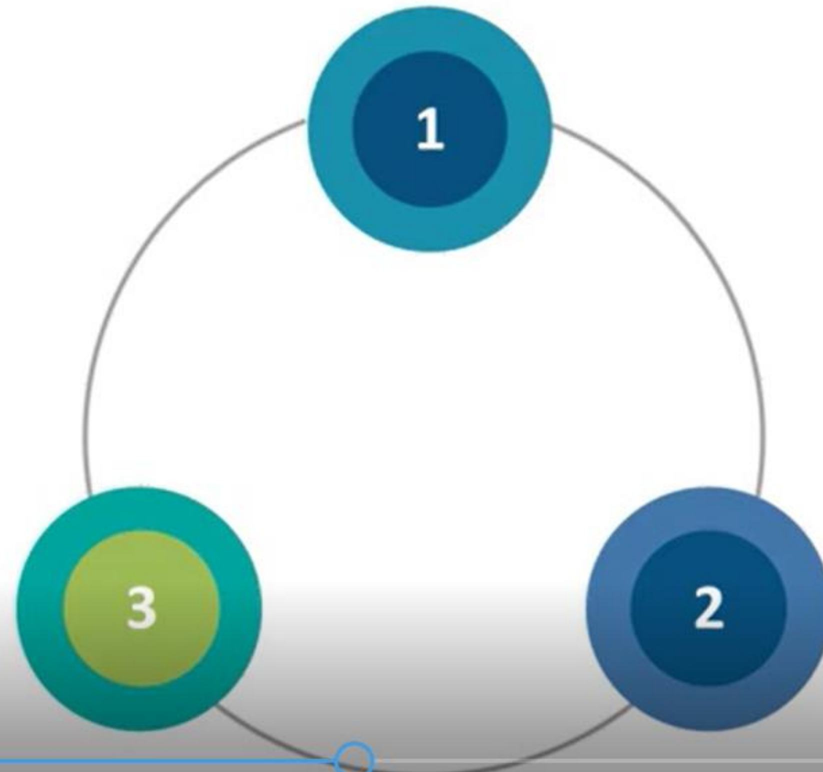
Moving Average Model takes the Q parameters, it is usually used for noise

Accounts for the possibility of a relationship between a variable and the residuals from previous periods



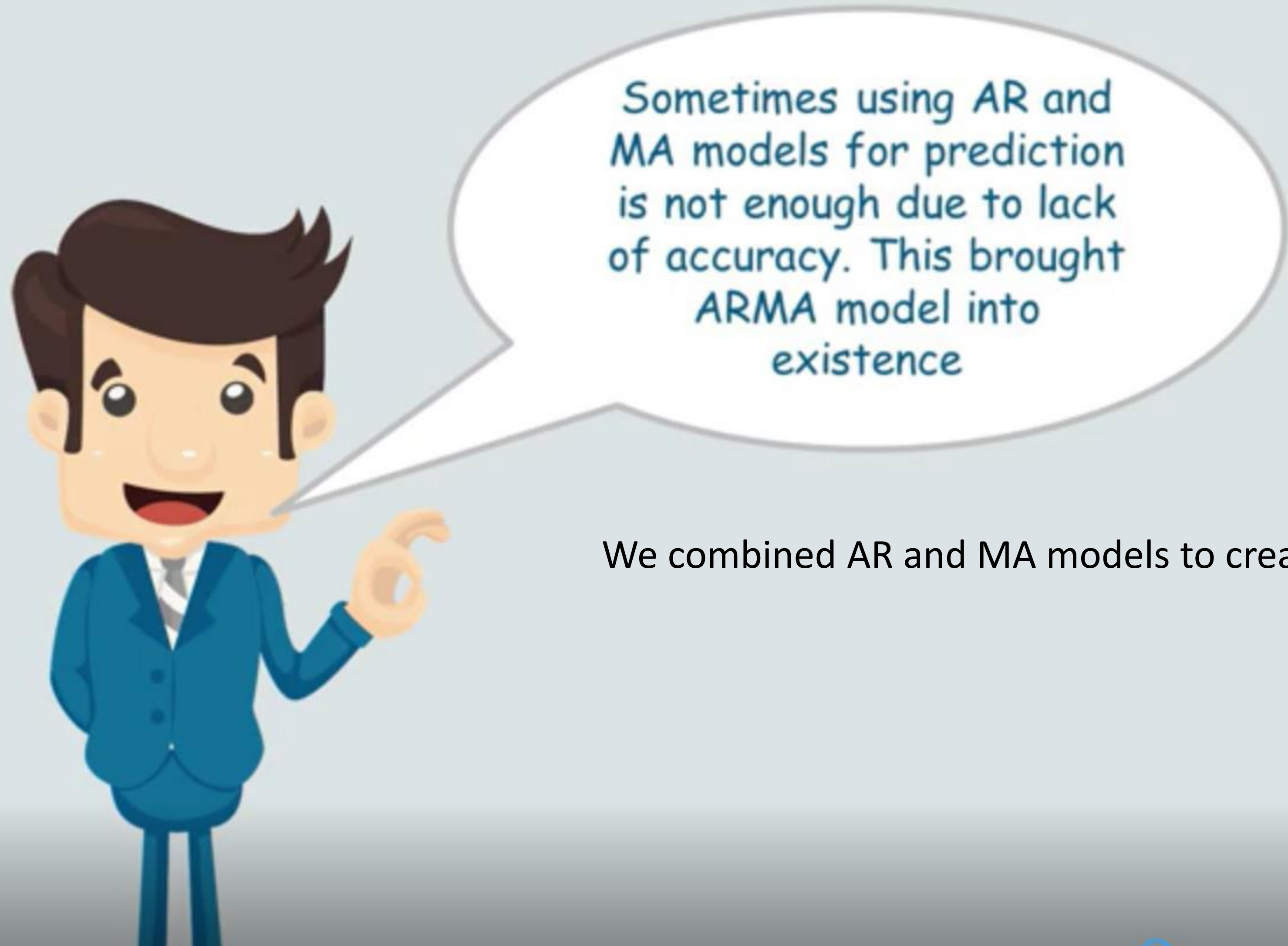
MA Model

Accounts for the possibility of a relationship between a variable and the residuals from previous periods



MA(q) is a moving average model with q lags:
$y_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i}$, where θ_q is the coefficient for the lagged error term in time t-q

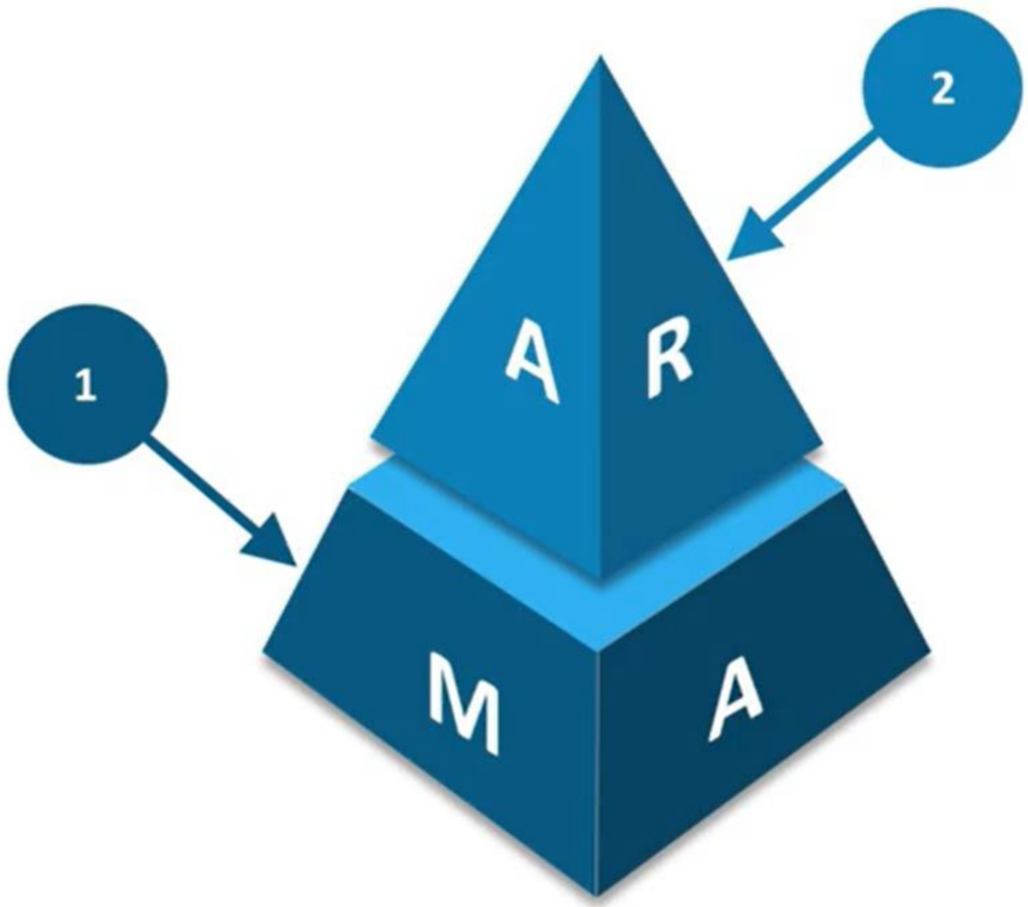
MA(1) model is expressed as:
$y_t = \mu + \epsilon_t + \theta_i \epsilon_{t-i}$



We combined AR and MA models to create ARMA

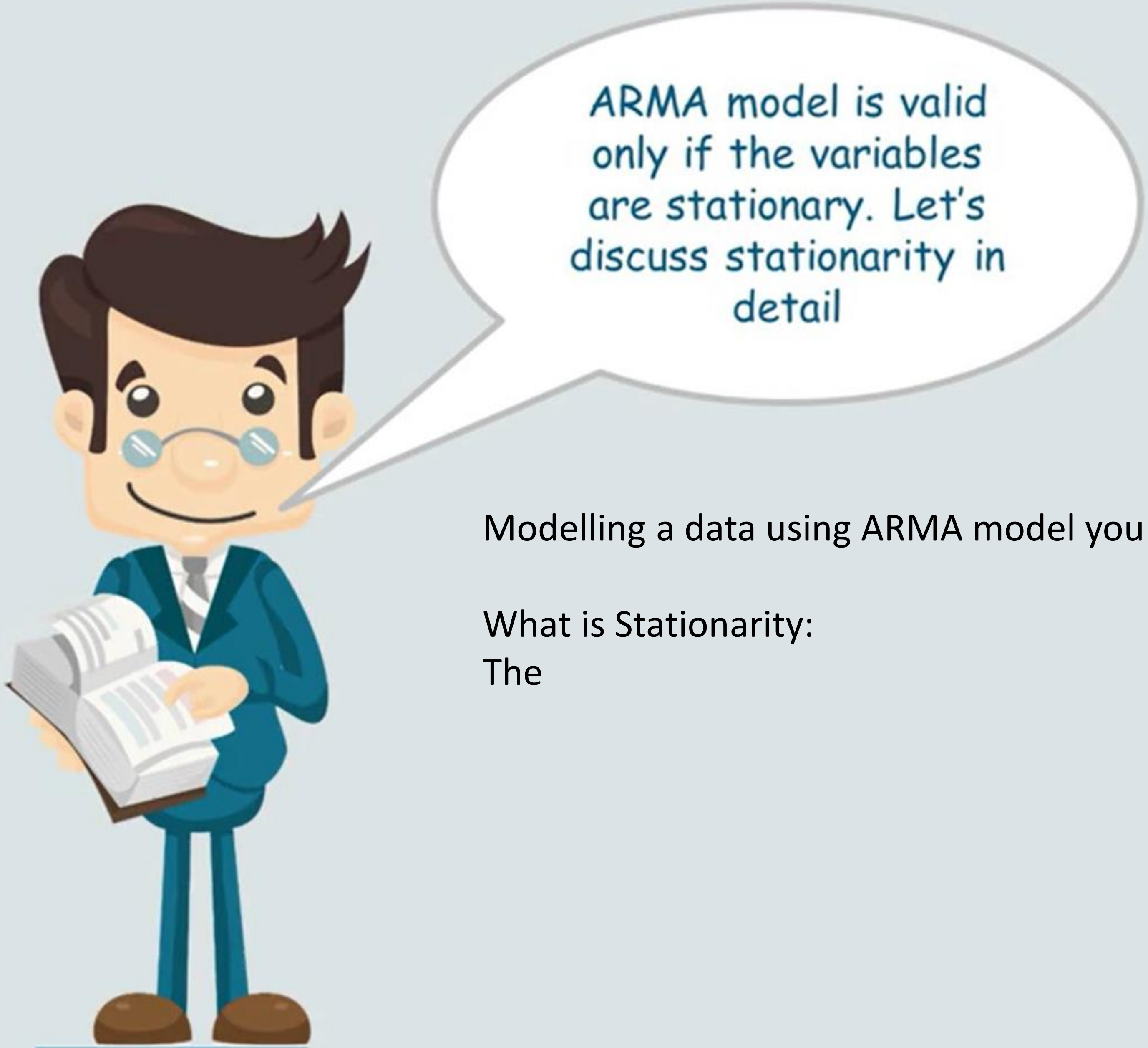
ARMA Model

Combination of both AR and MA models



Combine both p autoregressive terms and q moving average terms. Hence, called ARMA (p,q)

$$y_t = \mu + \underbrace{\sum_{i=1}^{q} \gamma_i \, y_{t-i}}_{\text{Auto Regressive part}} + \epsilon_t + \underbrace{\sum_{i=1}^{q} \theta_i \epsilon_{t-i}}_{\text{Moving Average part}}$$



Modelling a data using ARMA model you need data to have stationarity.

What is Stationarity:

The

Stationarity

Modelling an ARMA (p,q) process requires *stationarity*

STATIONARITY

- It is a process in which mean and variance remains unchanged or constant over the time . Such statistics are useful as descriptors of future behavior only if the **series is stationary**.

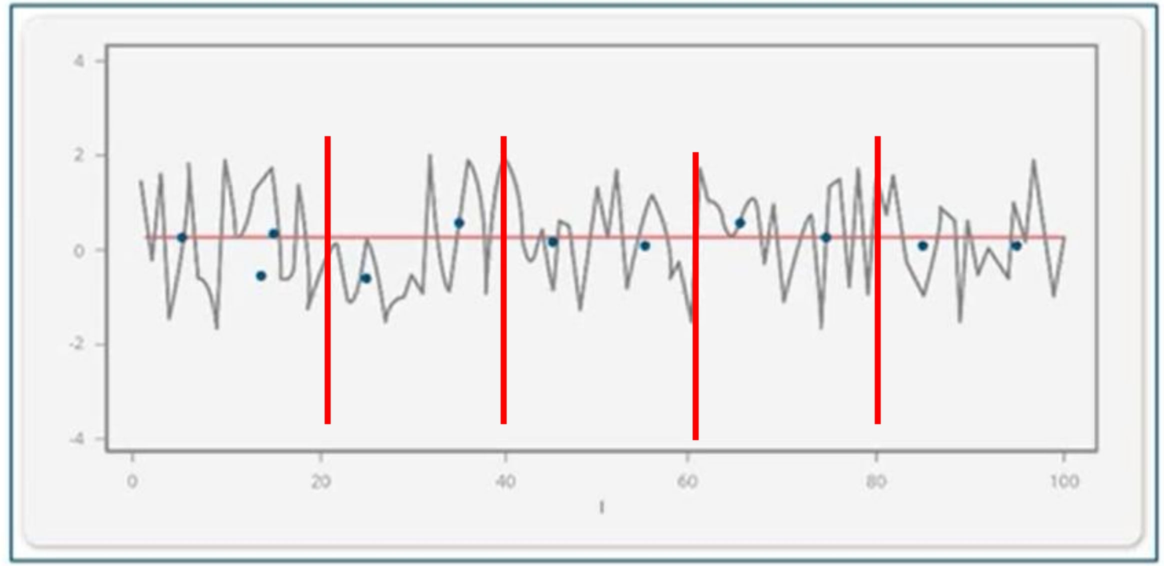
Stationarity

Description 1

Modelling an ARMA (p,q) process requires stationarity

Description 2

A process with mean and variance that do not change over time and do not have trends



A stationary plot

If the mean and variance of the times series at the various cuts are constant, then the data is stationary
Mean :5
Variance is 10
And remains same at various points
then it is stationary

Stationarity

What is ***p***

P is stationary

Modelling an ARMA (p,q) process requires stationarity

A process with mean and variance that do not change over time and do not have trends

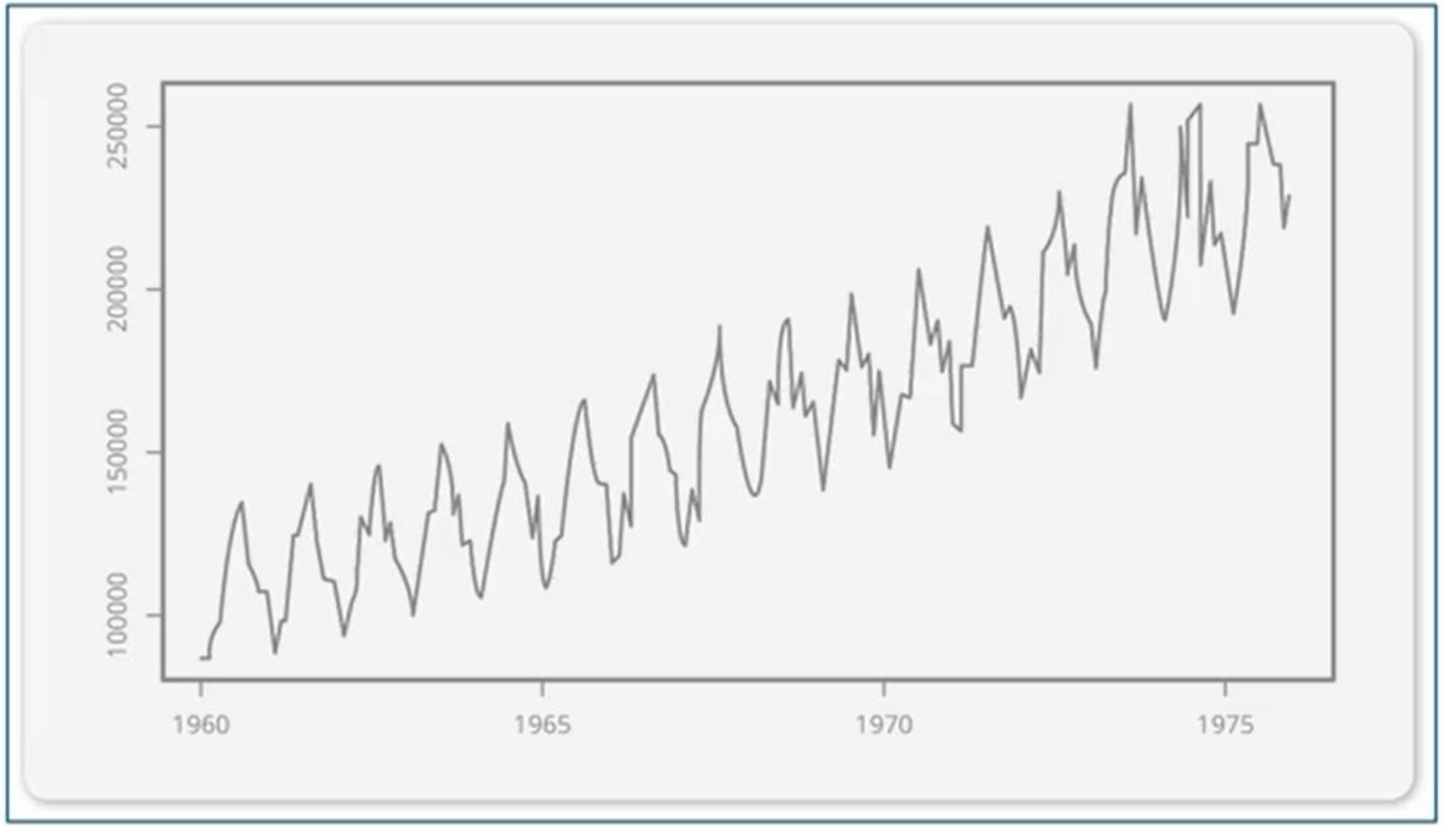
An AR(1) disturbance process:

$$\mu_t = \rho\, \mu_{t-1} + \epsilon_t$$

Stationary if $|\rho| < 1$ and ϵ_t is white noise

Stationarity – An Example

Consider an example of a time series variable:

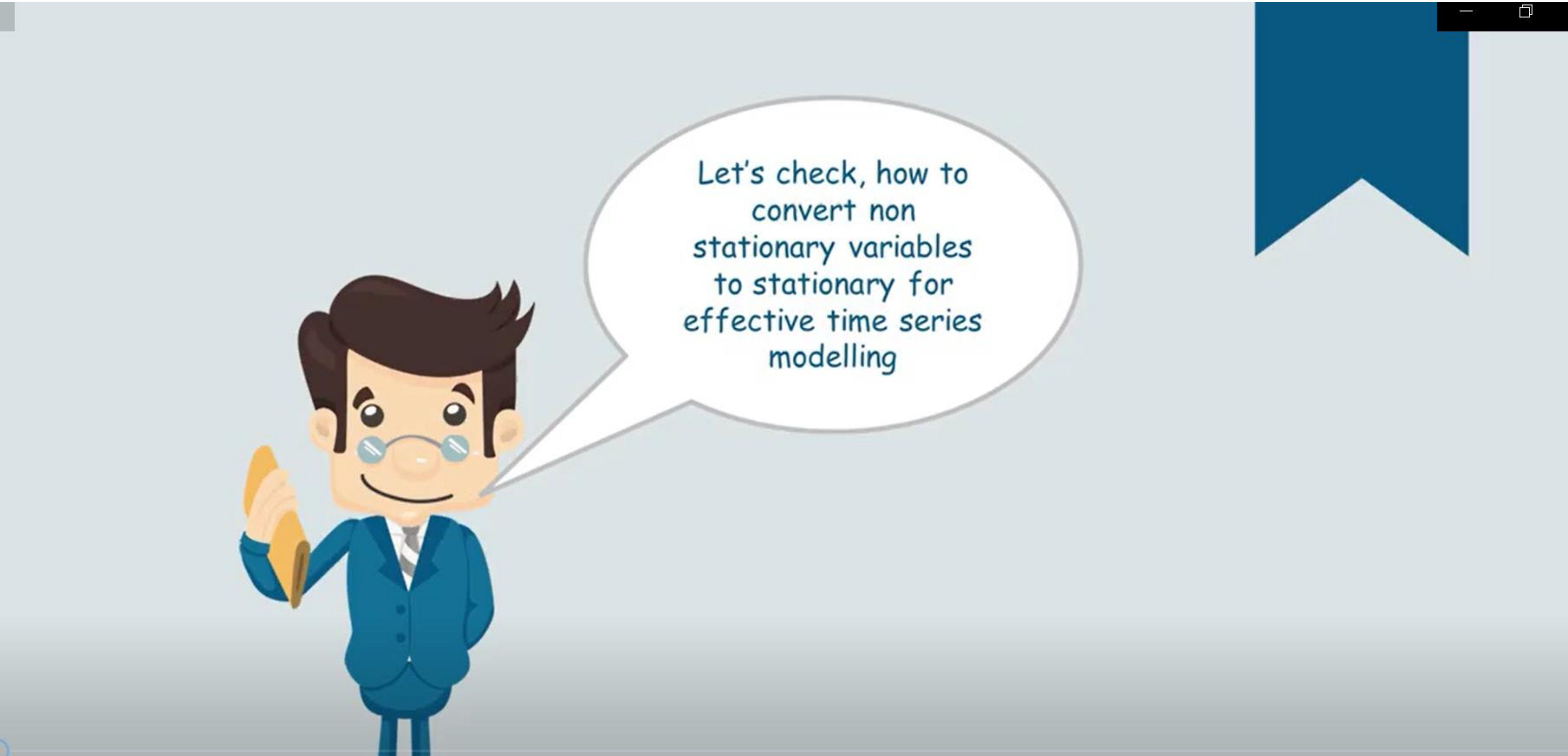


Here, the time series variable is not stationary as there is an increasing trend and it's oscillating over time

Has upward trend
Has seasonality
But there is no stationarity
To apply ARMA model convert non stationary into Stationary

HOW CONVERT NON STATIONARITY TO STATIONARITY

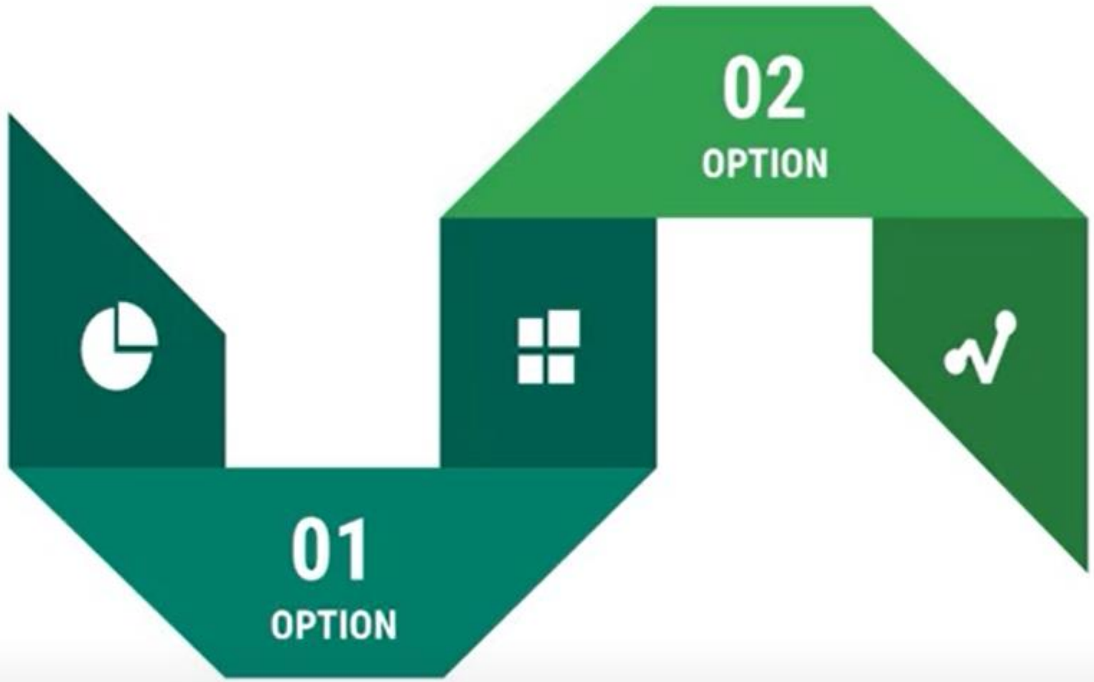
- It is a process in which mean and variance remains unchanged or constant over the time . Such statistics are useful as descriptors of future behavior only if the **series is stationary**.


Let's check, how to convert non stationary variables to stationary for effective time series modelling

Approaches to Remove Non-Stationarity

Maths:
Differentiation
Integration

Detrending



Differencing

We have to Use Differentiation to make a Time Series Stationary

De Trending
A Variable can be Detrend by Regressing the variable on a Time Trend
Fit a line through the time series
Find the residuals
We plot all the residuals against the time

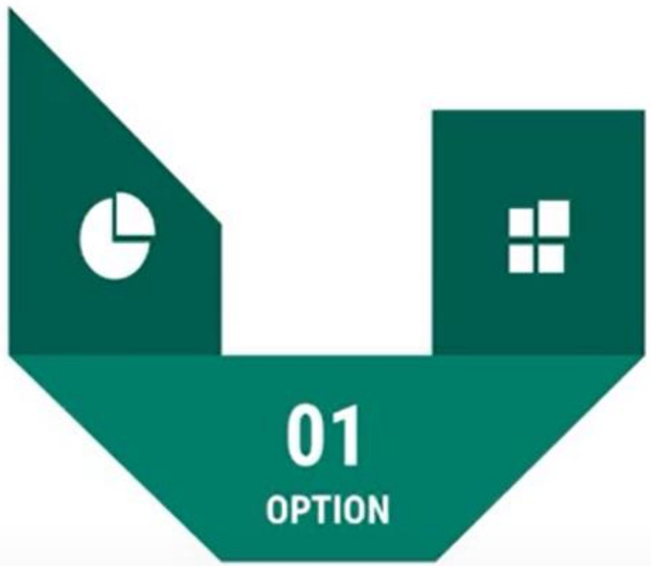
Detrending

Detrending

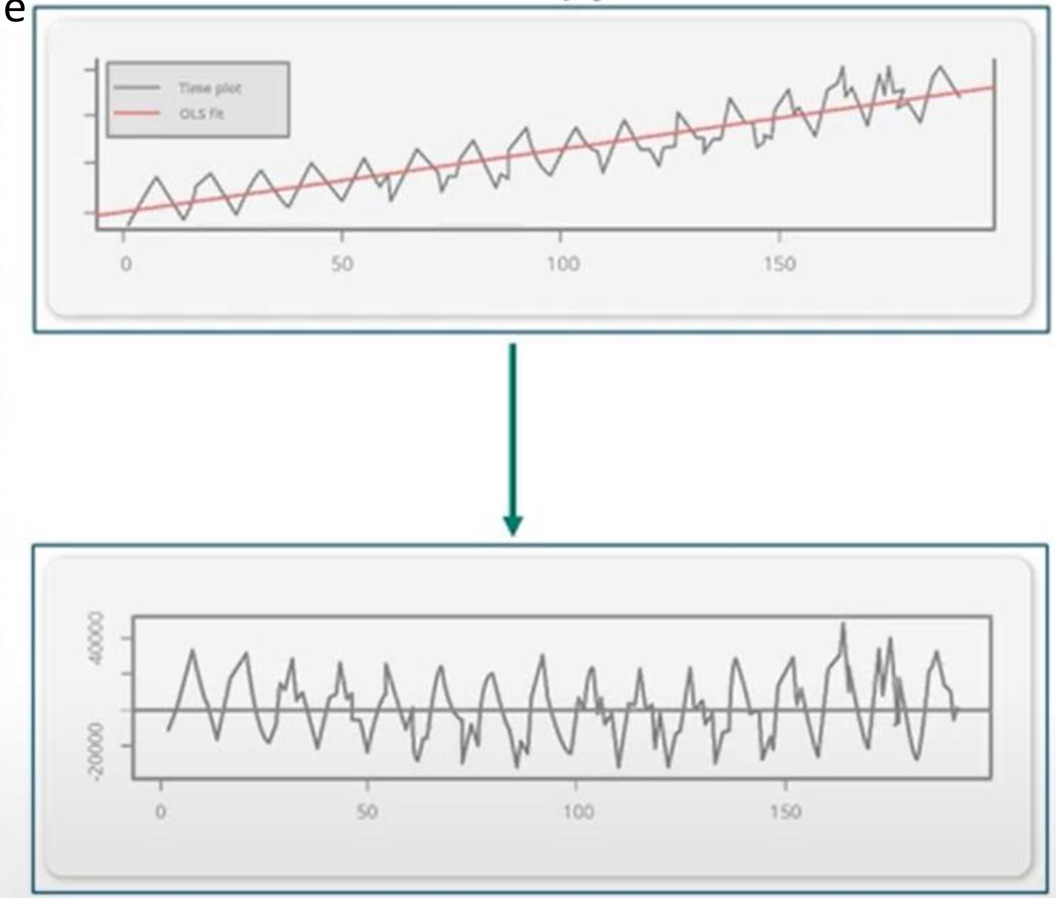
A variable can be detrended by regressing the variable on a time trend and obtaining the residuals:

$$y_t = \mu + \beta t + \epsilon_t$$

This equation is like y= mx+c



Variable y_t



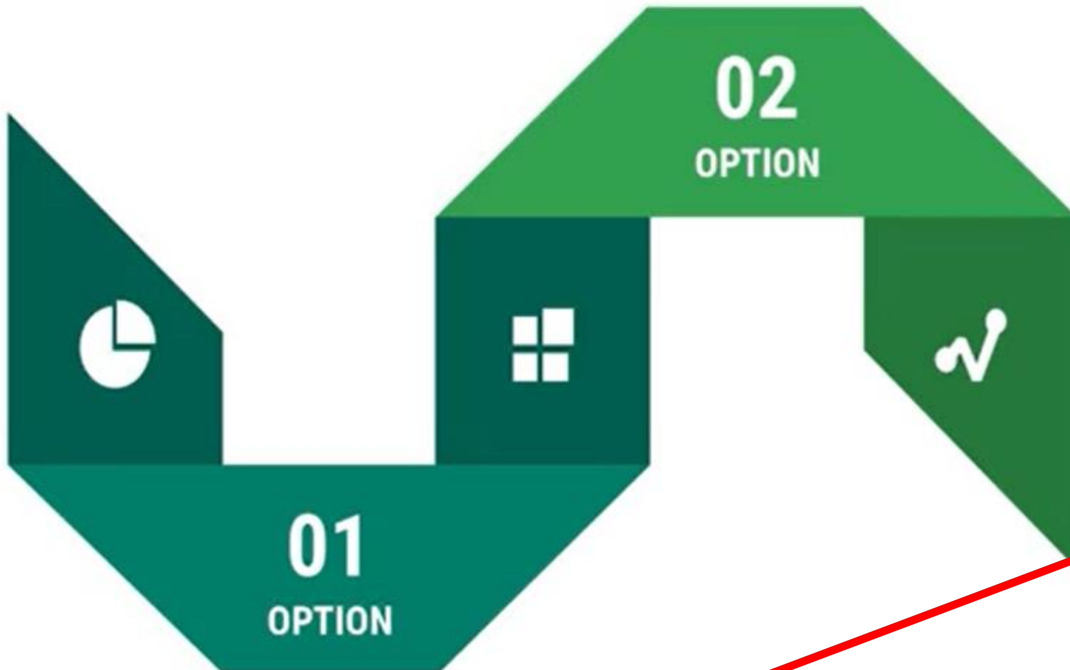
Detrended variable $\hat{\varepsilon}_t = y_t - \hat{\mu} + \beta t$

Differencing

Detrending

A variable can be detrended by regressing the variable on a time trend and obtaining the residuals:

$$y_t = \mu + \beta t + \epsilon_t$$



Differentiation

Differencing

Uses the concept of differenced variable:

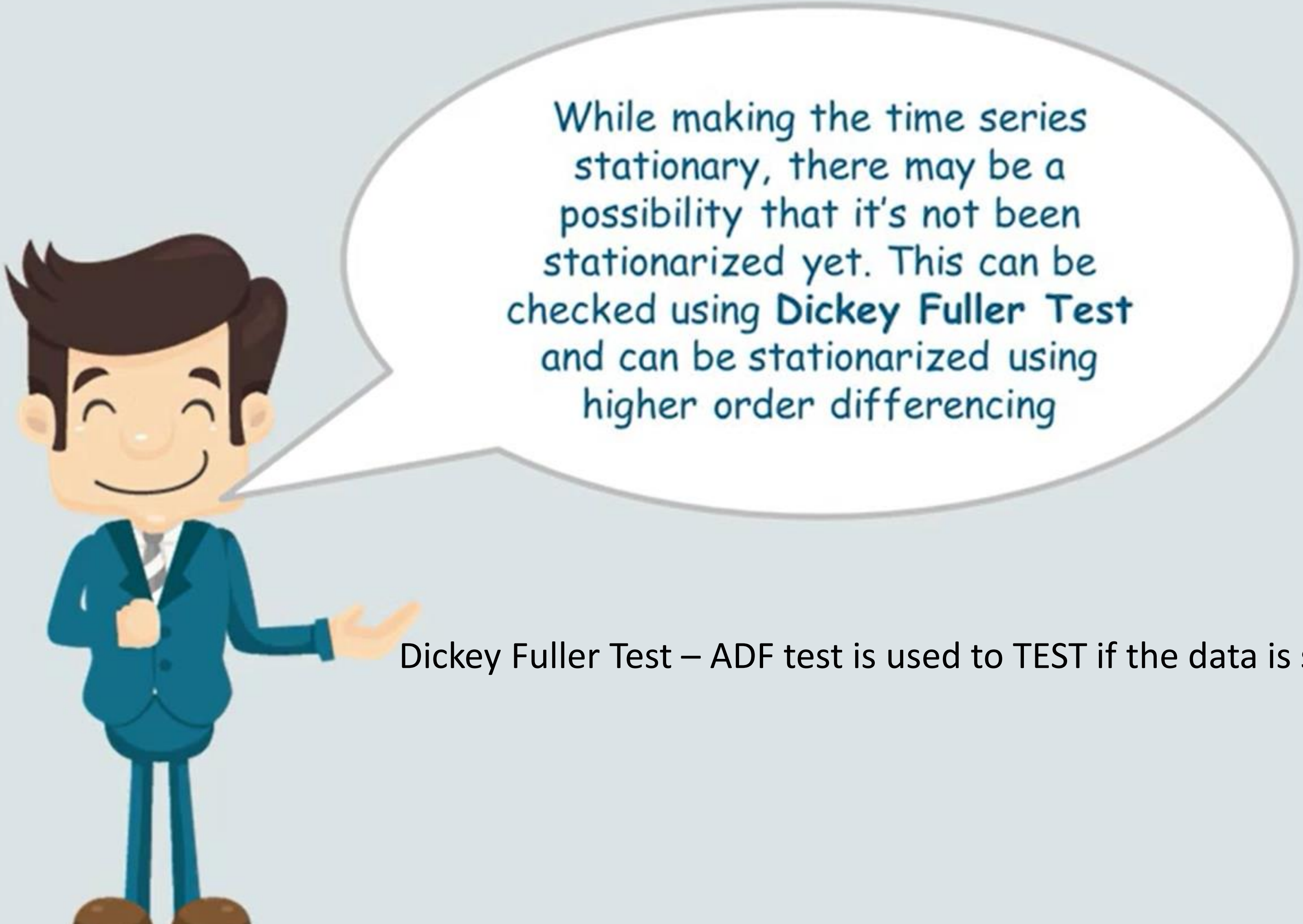
$\Delta y_t = y_t - y_{t-1}$, for first order differences

The variable y_t is integrated of order one, denoted I(1), if taking a first difference, producing a stationary process

This is differencing.
We keep Differencing the time series till it becomes stationary

First we started with Time Series. We applied AR and MA model
Then we combined the 2 models ARMA
We have to remove the stationarity
So we have
AR Integrated MA (ARIMA)

ARIMA (p,d,q) denotes an ARMA model with p autoregressive lags, q moving average lags and difference in the order of d



Dickey Fuller Test – ADF test is used to TEST if the data is stationary or not stationary

Dickey Fuller Test for Stationarity

Assume an AR(1) model.

The model is non-stationary or a unit root is present if $|\rho| = 1$

$$y_t = \rho y_{t-1} + \epsilon_t$$

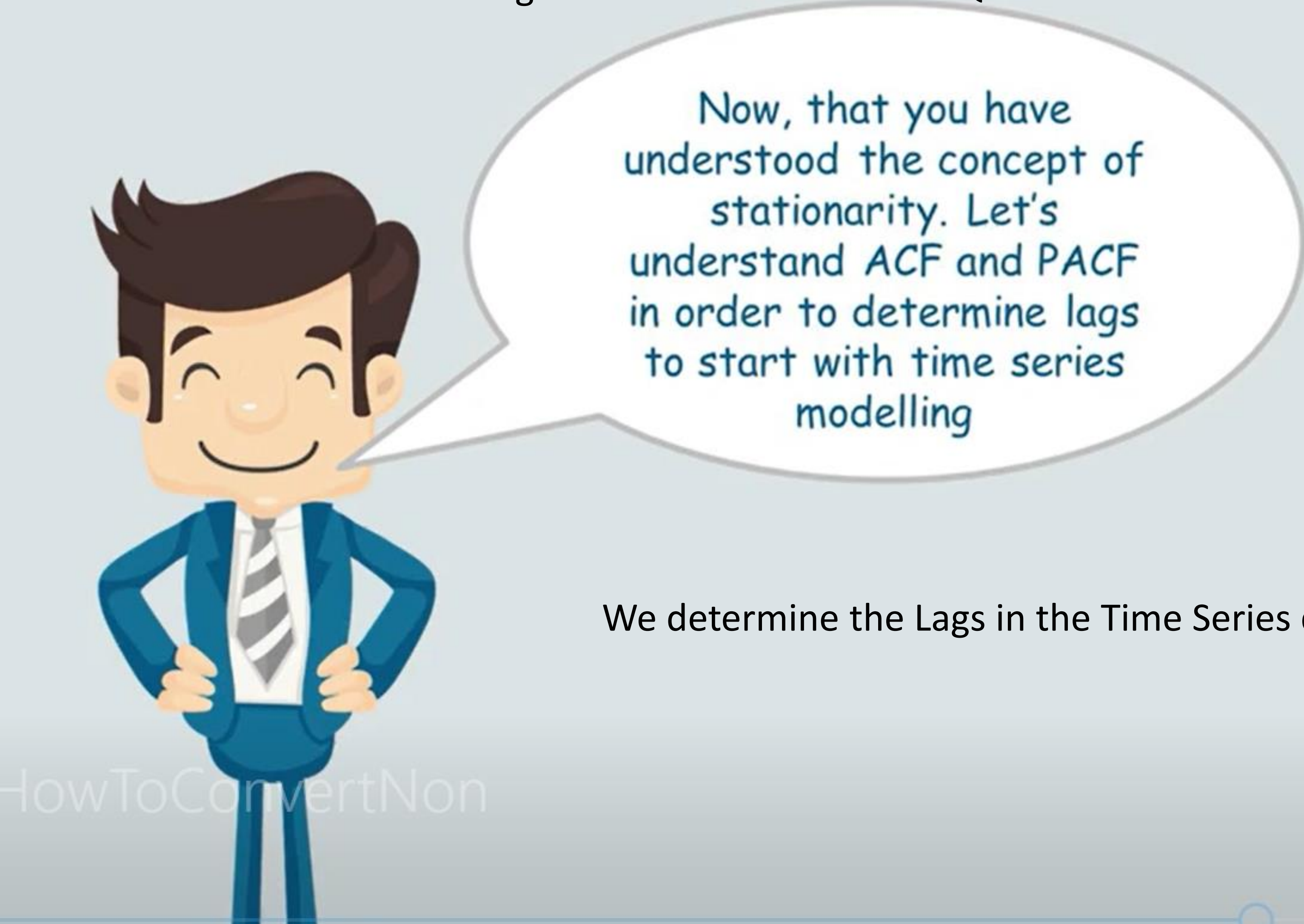
$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \epsilon_t$$

$$\Delta y_t = (\rho - 1) y_{t-1} + \epsilon_t = \gamma y_{t-1} + \epsilon_t$$

You can estimate the above model for stationarity by testing the significance of the γ coefficient:

- If the null hypothesis is not rejected, $\gamma^* = 0$, then y_t is not stationary
- Difference the variable and repeat the test to see if the differenced variable is stationary
- If the null hypothesis is rejected, $\gamma^* > 0$, then y_t is stationary

Let us determine the LAGS in the time series modelling using the ACF and PCF
We need to know the Lags to determine the P and Q in time series modelling



We determine the Lags in the Time Series data using ACF and PCF

ACF (Auto Correlation Function)

1. ACF is the proportion of the covariance of y_t and y_{t-k} to the variance of a dependent variable y_t :

$$\text{ACF(k)} = \rho_k = \frac{\text{Cov}(y_t, y_{t-k})}{\text{Var}(y_t)}$$

2. Gives the gross correlation between y_t and y_{t-k}

3. For an AR(1) model, the ACF(k) = $\rho_k = \gamma^k$

PACF (Partial Auto Correlation Function)

01

Simple correlation between y_t and y_{t-k} minus the part explained by the intervening lags

$\rho_k^* = \text{Corr}|\ y_t - E^*(y_t|\ y_{t-1},\dots, y_{t-k+1}), y_{t-k})|$

Where $E^*(y_t|\ y_{t-1},\dots, y_{t-k+1})$ is the minimum mean squared predictor of y_t by $y_{t-1},\dots, y_{t-k+1}$

02

For an AR(1) model, the PACF is γ for the first lag

P value is given by PACF

Do It Skills
do it. enjoy it.

THANK YOU

ARUNKG99@GMAIL.COM

WWW.DOITSKILLS.COM