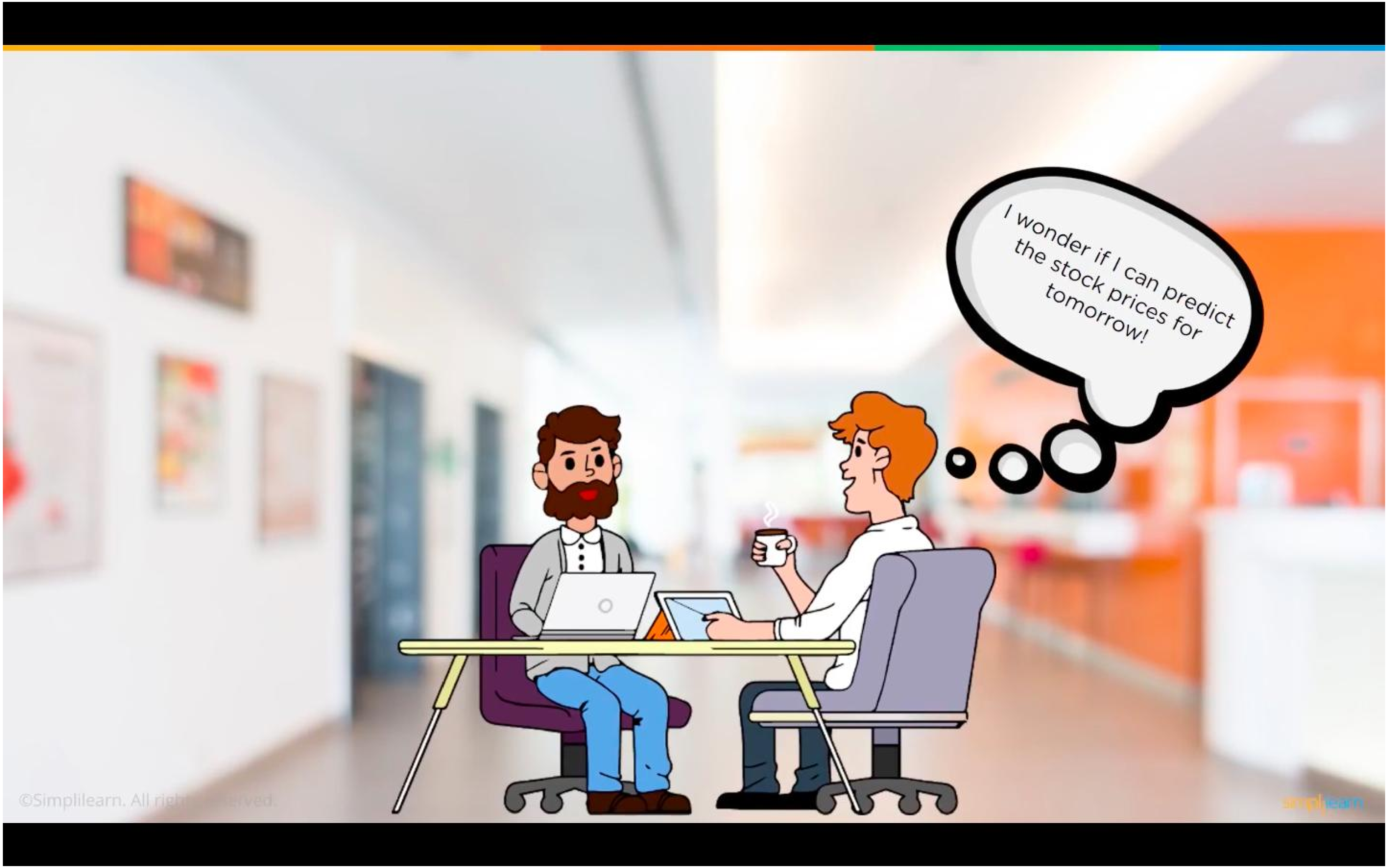
What's in it for you?

- Why Time Series?
- What is Time Series?
- Components of a Time Series
- When NOT to use Time Series?
- Why does a Time Series have to be stationary?
- How to make a Time Series stationary?
- Example: Forecast car sales for 5^{th} year


I wonder if I can predict the stock prices for tomorrow!
©Simplilearn. All rights reserved.
simplilearn


Well, you can actually do that with Time Series forecasting

Well, you can actually
that with Time Series
forecasting

We can predict:

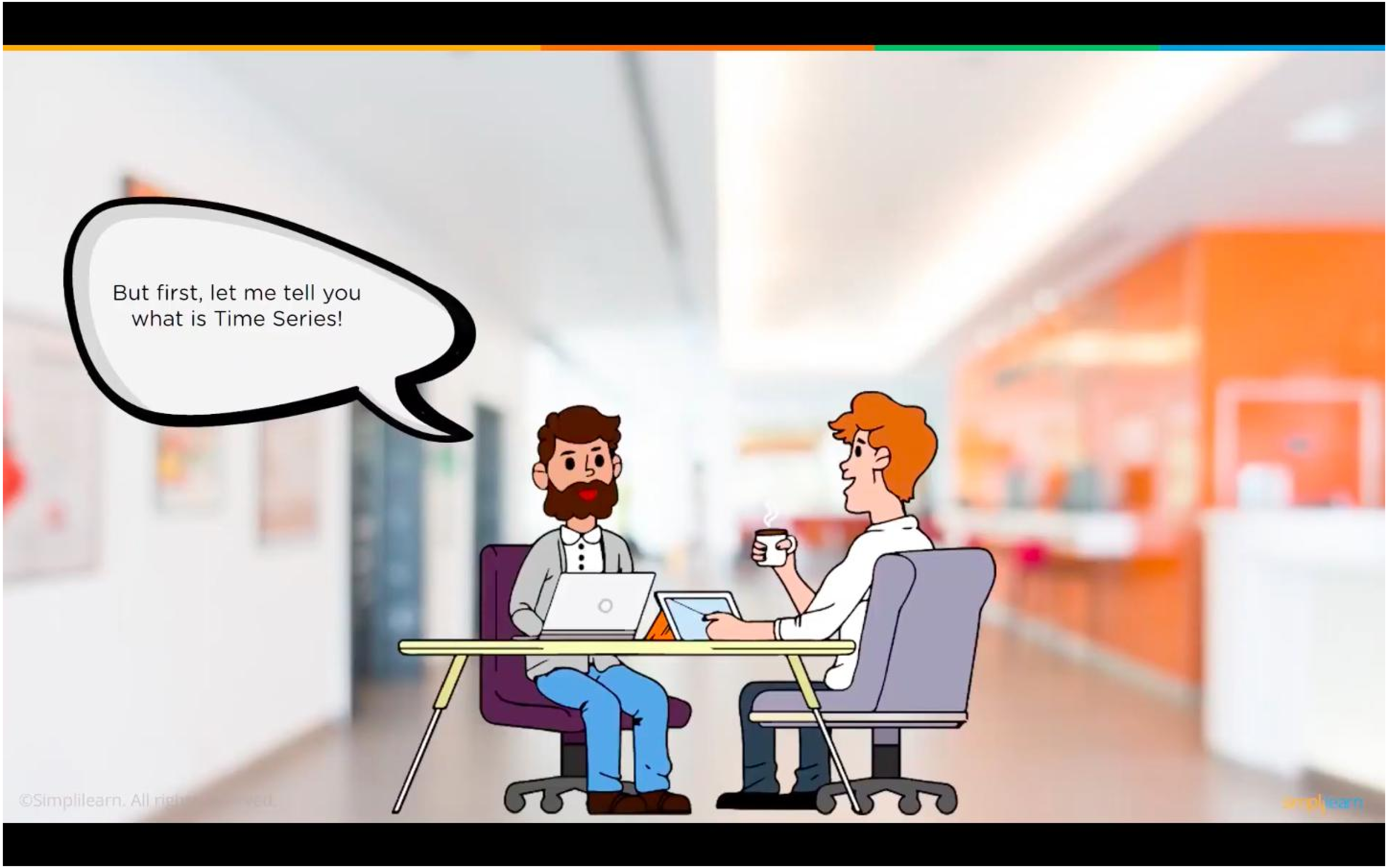
Daily Stock Price

Weekly interest rates

Sales figures

where the outcome (independent variable) is dependent on time

In such scenarios, we use Time Series forecasting


But first, let me tell you what is Time Series!
©Simplilearn. All rights reserved.

What is Time Series?

A Time Series data for stock price analysis may look like this:

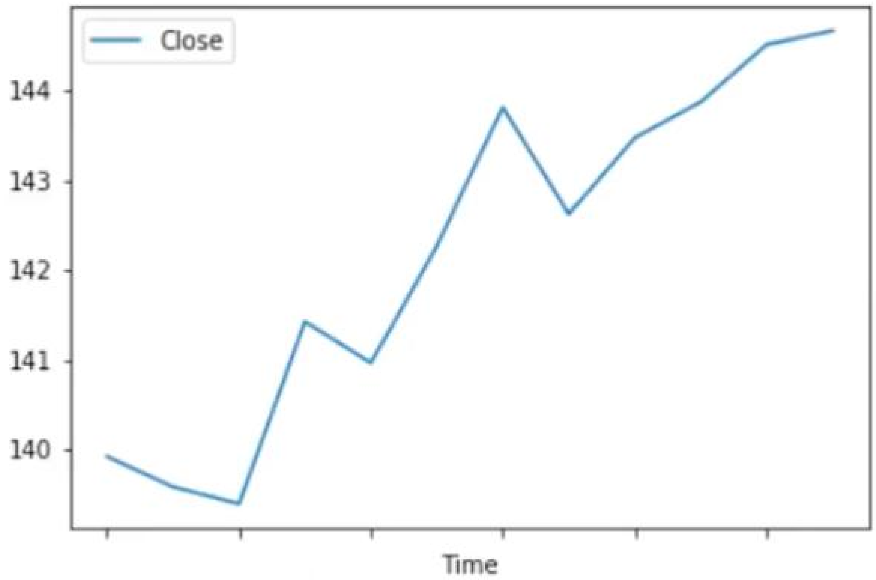
	A	B
1	Date	Close
2	1/4/2017	139.92
3	2/4/2017	139.58
4	3/4/2017	139.39
5	4/4/2017	141.42
6	5/4/2017	140.96
7	6/4/2017	142.27
8	7/4/2017	143.81
9	8/4/2017	142.62
10	9/4/2017	143.47
11	10/4/2017	143.87
12	11/4/2017	144.51
13	12/4/2017	144.66

The stock prices change everyday!

simplilearn

What is Time Series?

A simple plot shows that it is increasing with time!

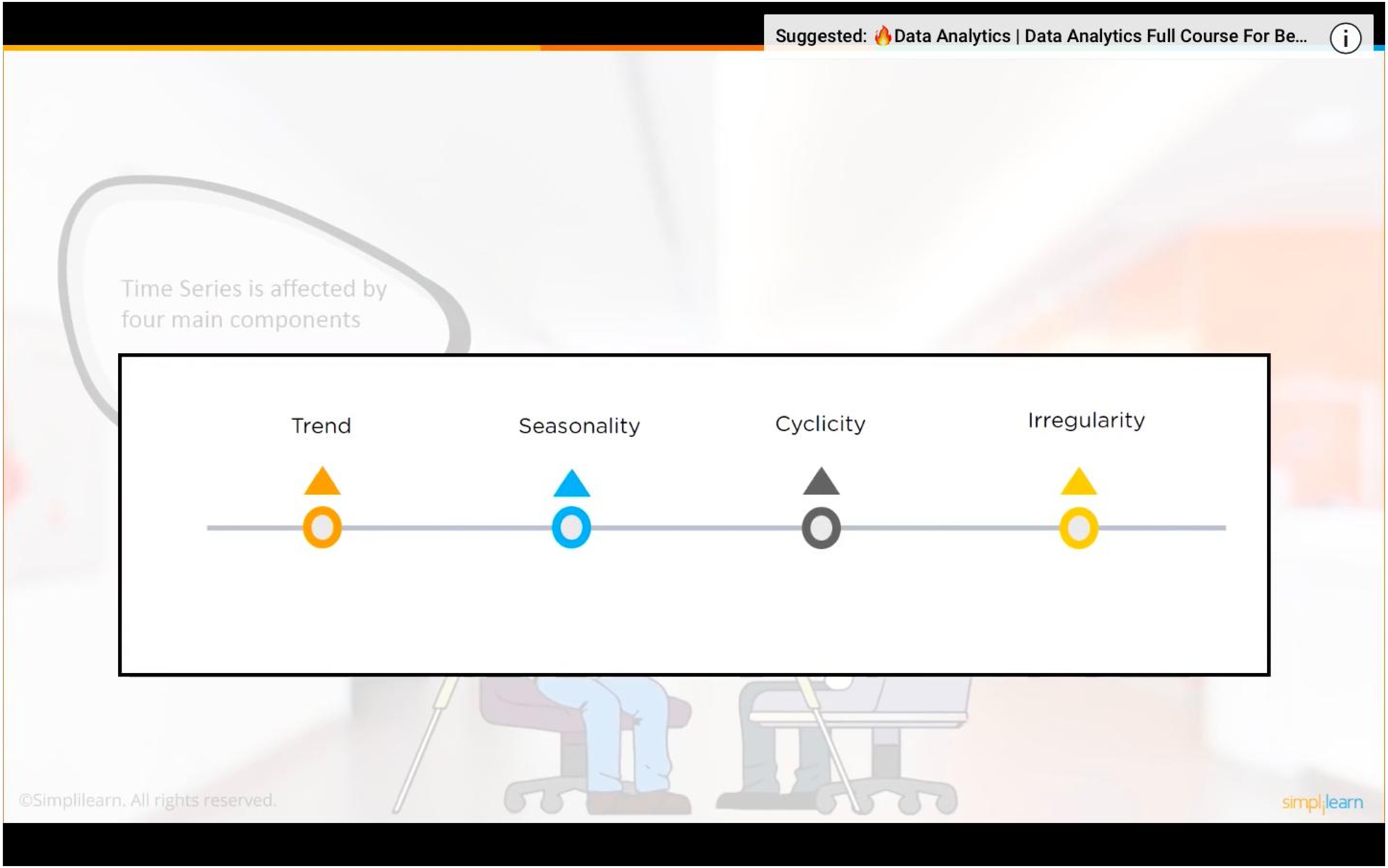


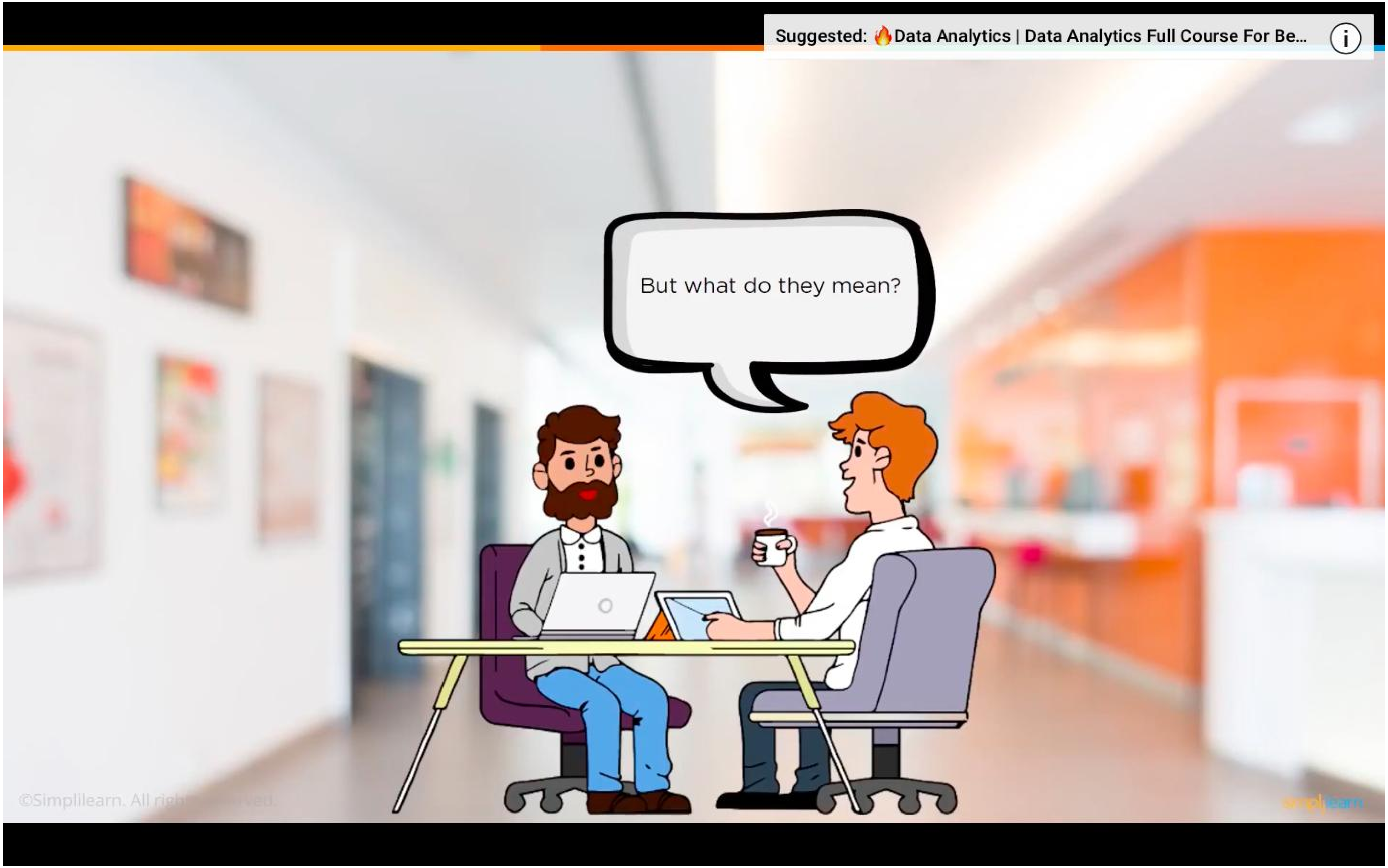
What is Time Series?

- A Time Series is a sequence of data being recorded at specific time intervals
- These data points (past values) are analyzed to forecast a future
- It is time-dependent

simplilearn


Time Series is affected by
four main components
©Simplilearn. All rights reserved.
simplilearn


Suggested: 🔥Data Analytics | Data Analytics Full Course For Be...
Time Series is affected by four main components
Trend
Seasonality
Cyclicity
Irregularity
©Simplilearn. All rights reserved.
simplilearn


Suggested: 🔥Data Analytics | Data Analytics Full Course For Be...
But what do they mean?

Time Series is affected by four main components

Trend

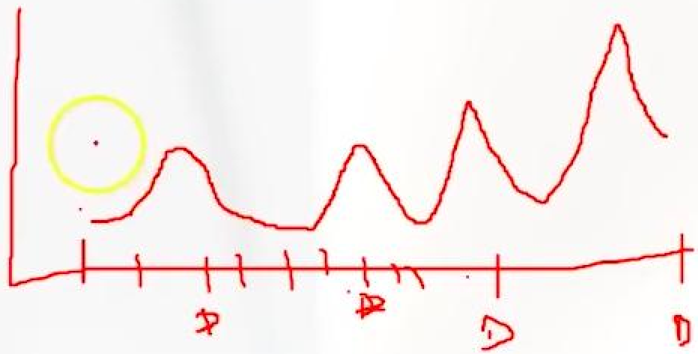
Trend is the increase or decrease in the series over a period of time, it persists over a long period of time

Example: Population growth over the years can be seen as an upward trend

Time Series is affected by four main components

Trend

- Trend is the increase or decrease in the series over a period of time, it persists over a long period of time
- Example: Population growth over the years can be seen as an upward trend



Time Series is affected by four main components

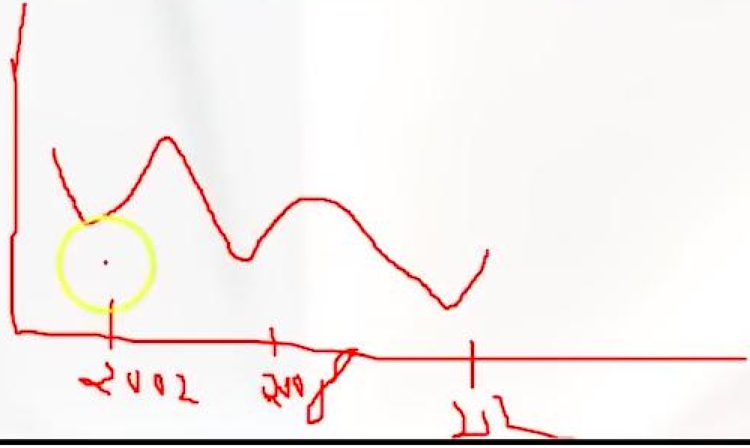
Seasonality

Regular pattern of up and down fluctuations
It is a short-term variation occurring due to seasonal factors

Example: Sales of ice-cream increases during summer season

season

Time Series is affected by four main components



Cyclicity

- It is a medium-term variation caused by circumstances, which repeat in irregular intervals
- Example: 5 years of economic growth, followed by 2 years of economic recession, followed be 7 years of economic growth followed by 1 year of economic recession

simplilearn

Time Series is affected by four main components

Irregularity

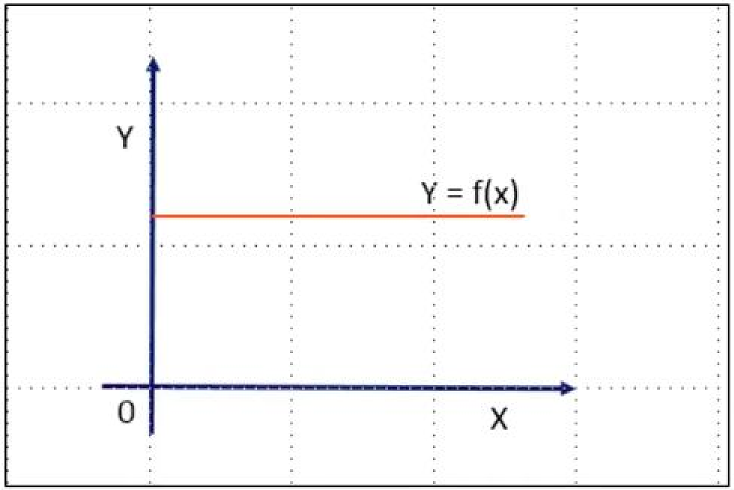
- It refers to variations which occur due to unpredictable factors and also do not repeat in particular patterns
- Example: Variations caused by incidents like earthquake, floods, war etc.


Are there conditions where we shouldn't use Time Series?

When NOT to use Time Series Analysis?

There are various conditions where you should not use Time Series:

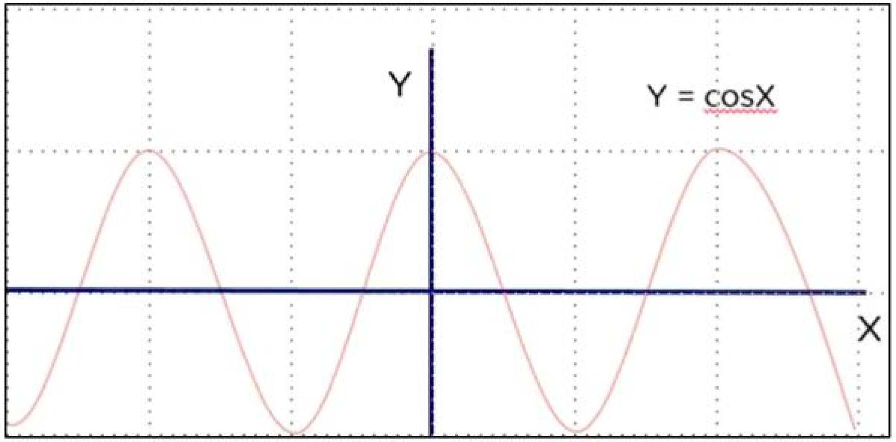
1. When the values are constant over a period of time:



When NOT to use Time Series Analysis?

There are various conditions where you should not use Time Series:

2 When values can be represented by known functions like cosx, sinx etc:



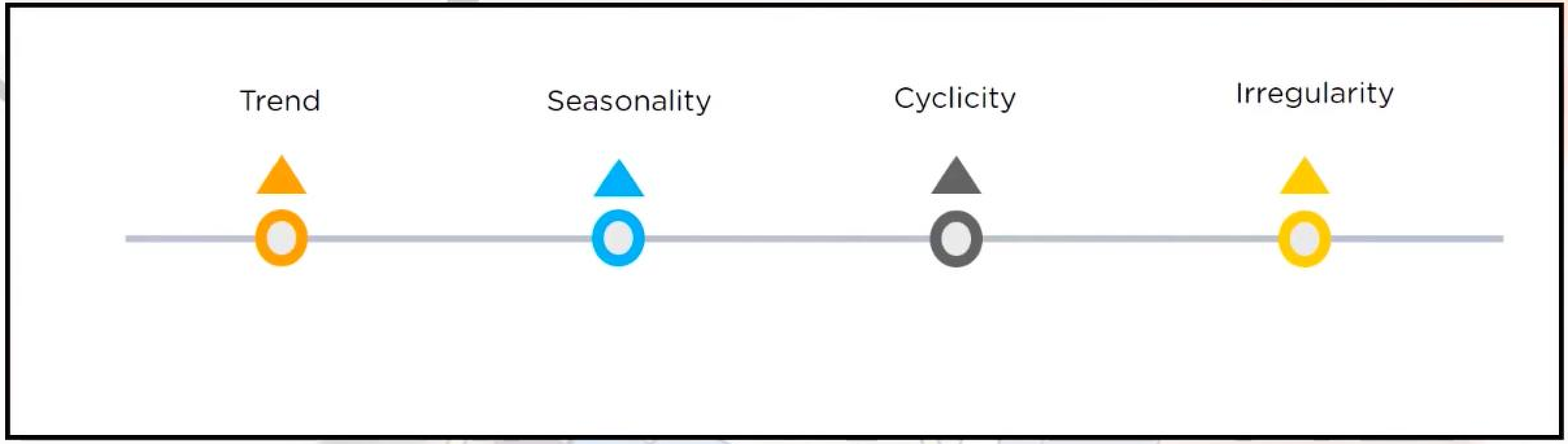

Before forecasting, you should make sure that Time Series is stationary


Why? And what do you mean by making time series stationary?


Okay, first let's understand what is Non-Stationary Time Series!


You remember the four components of Time Series?

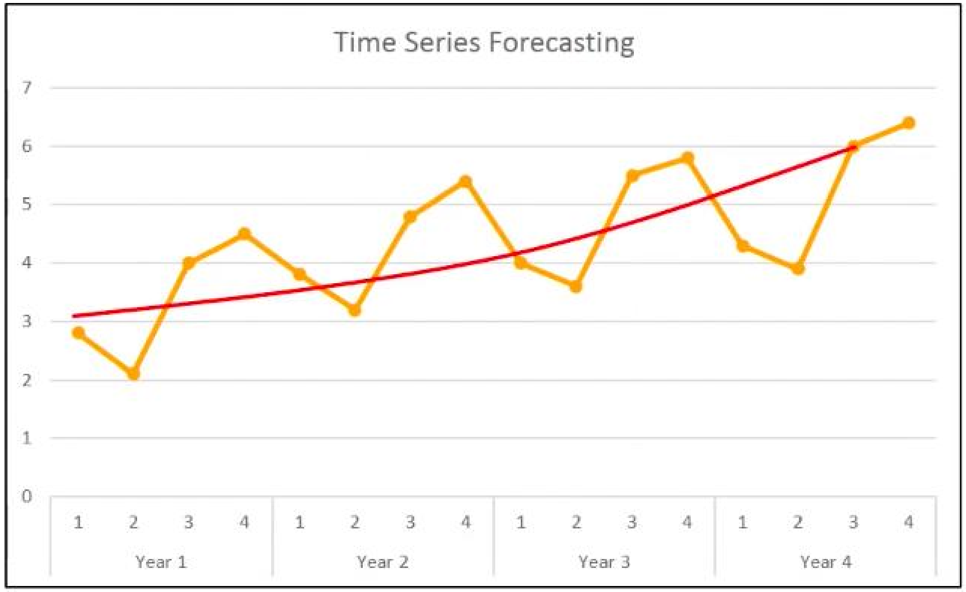
Time Series is affected by four main components

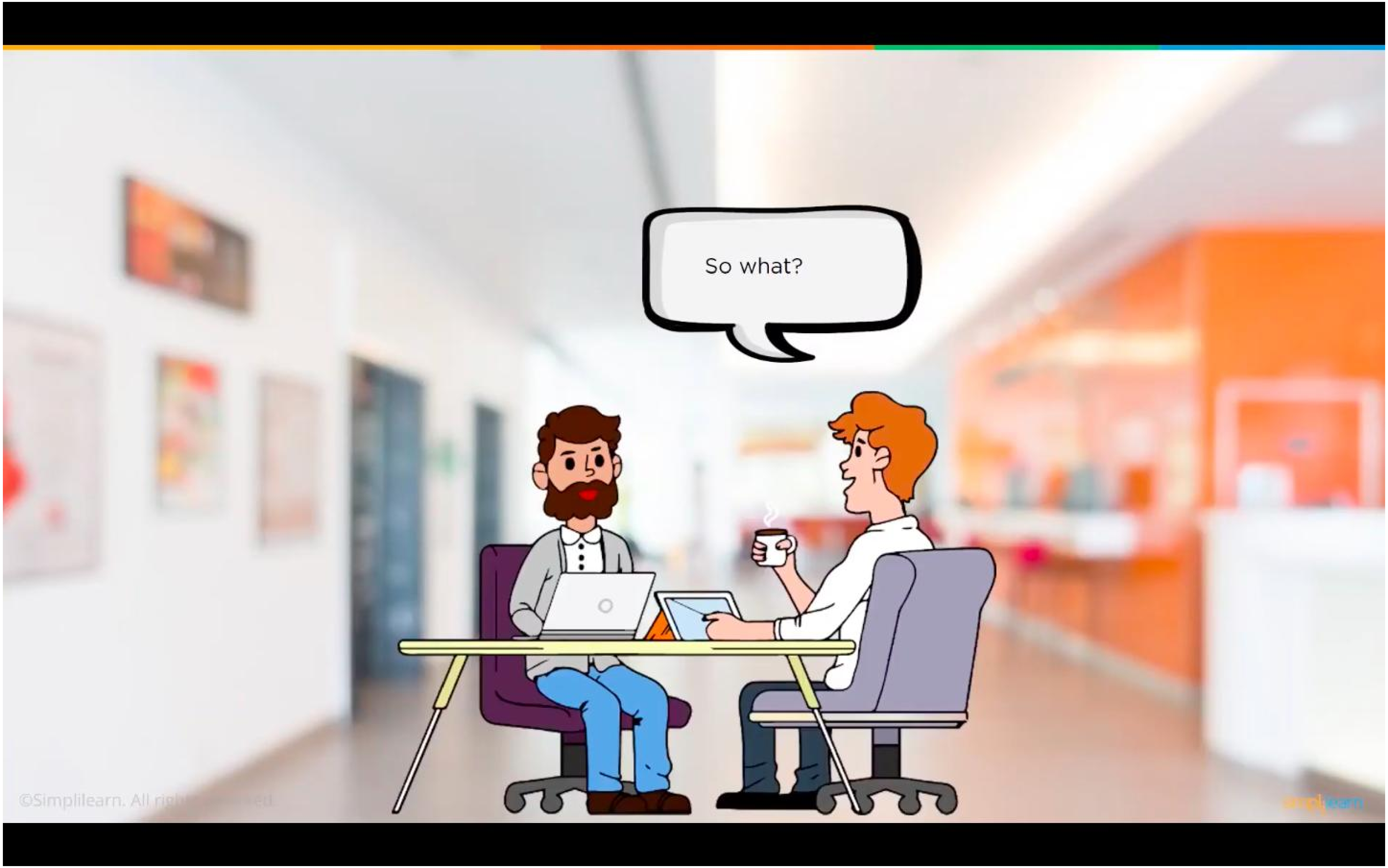



If these components are present in Time Series data, it is a Non-Stationary Time Series Data.

For example, look at this graph:

Here, the mean is non-constant and there is clearly an upward trend




So what?

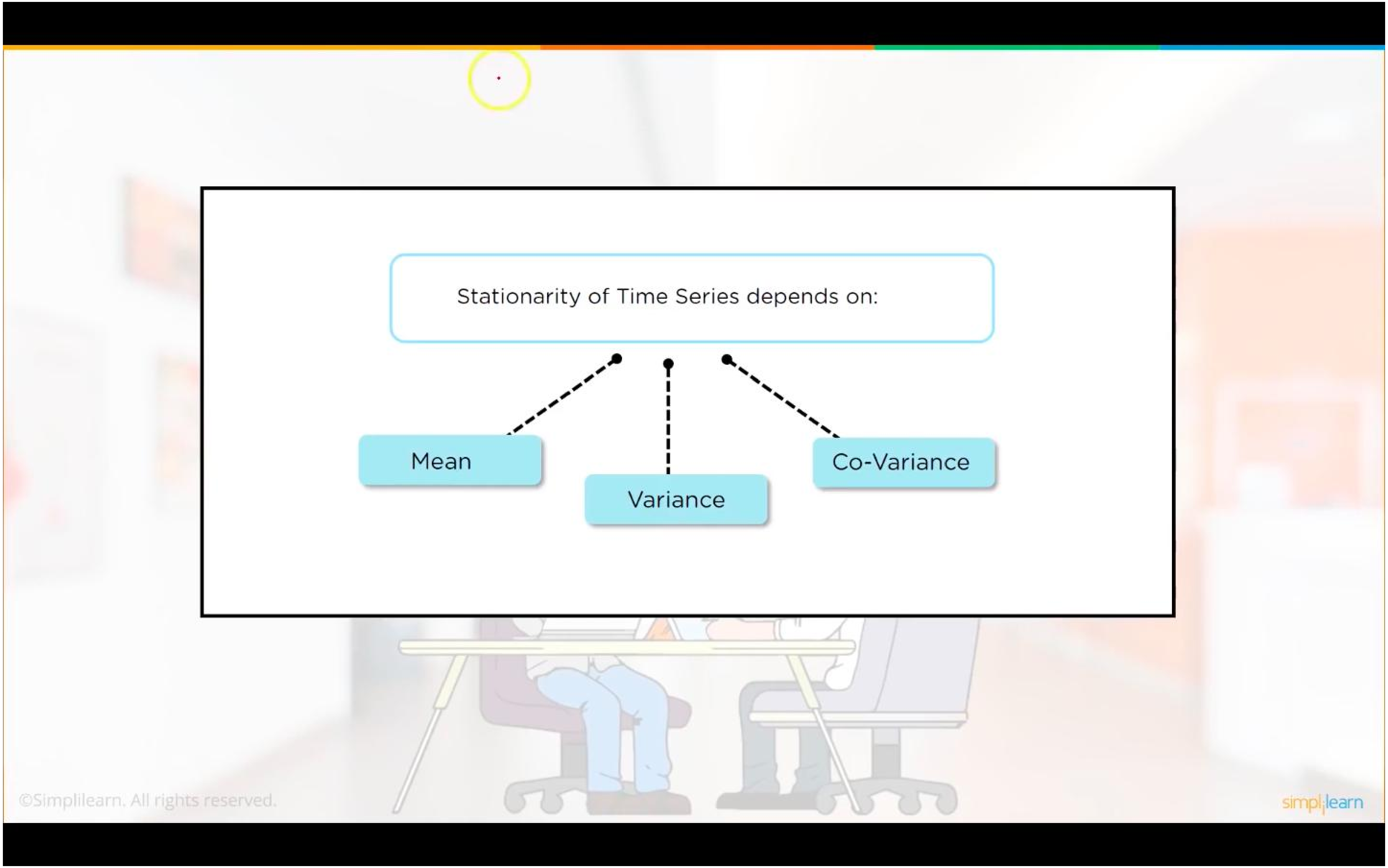
Stationarity of Time Series

A Non-stationary Time Series has trend and seasonality components, which will affect the forecasting of Time Series

When a Time Series is stationary, we can identify previously unnoticed components to strengthen their forecasting

simplilearn


How do you differentiate between a stationary and Non-Stationary time series?



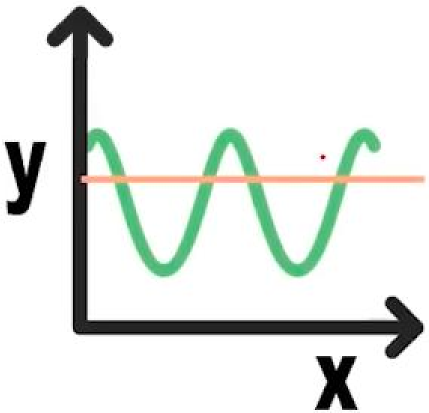
simplilearn

Stationarity of Time Series

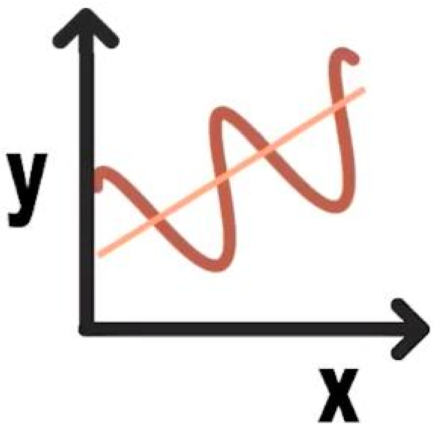
The mean of the series should not be a function of time rather should be a constant

Stationary

Here, mean is constant with time



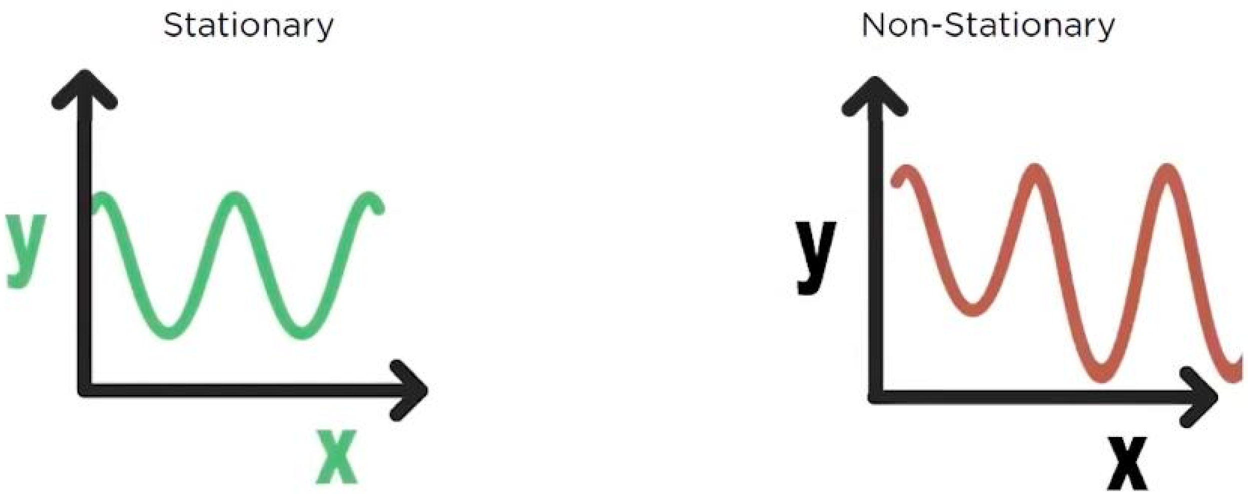
Non-Stationary



Here, mean is increasing with time

Stationarity of Time Series

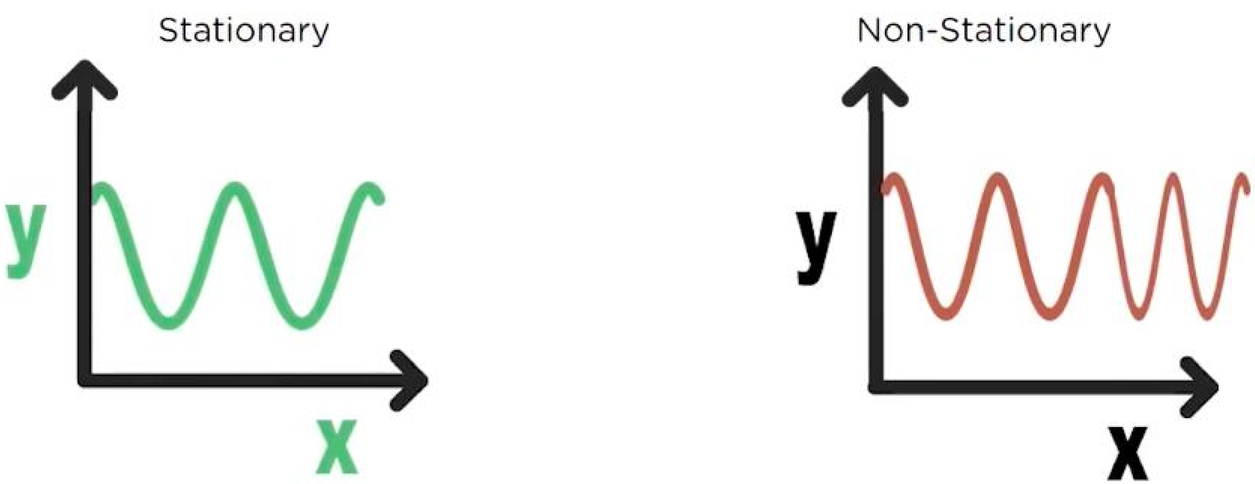
The variance of the series should not be a function of time



Notice, the varying spread of distribution in the red grapgh, this shows that variance is changing with time

Stationarity of Time Series

The covariance of the i th term and the (i + m) th term should not be a function of time



The spread becomes closer as the time increases. Hence, the covariance is not constant with time for the 'red series' and is Non-Stationary!

simplilearn

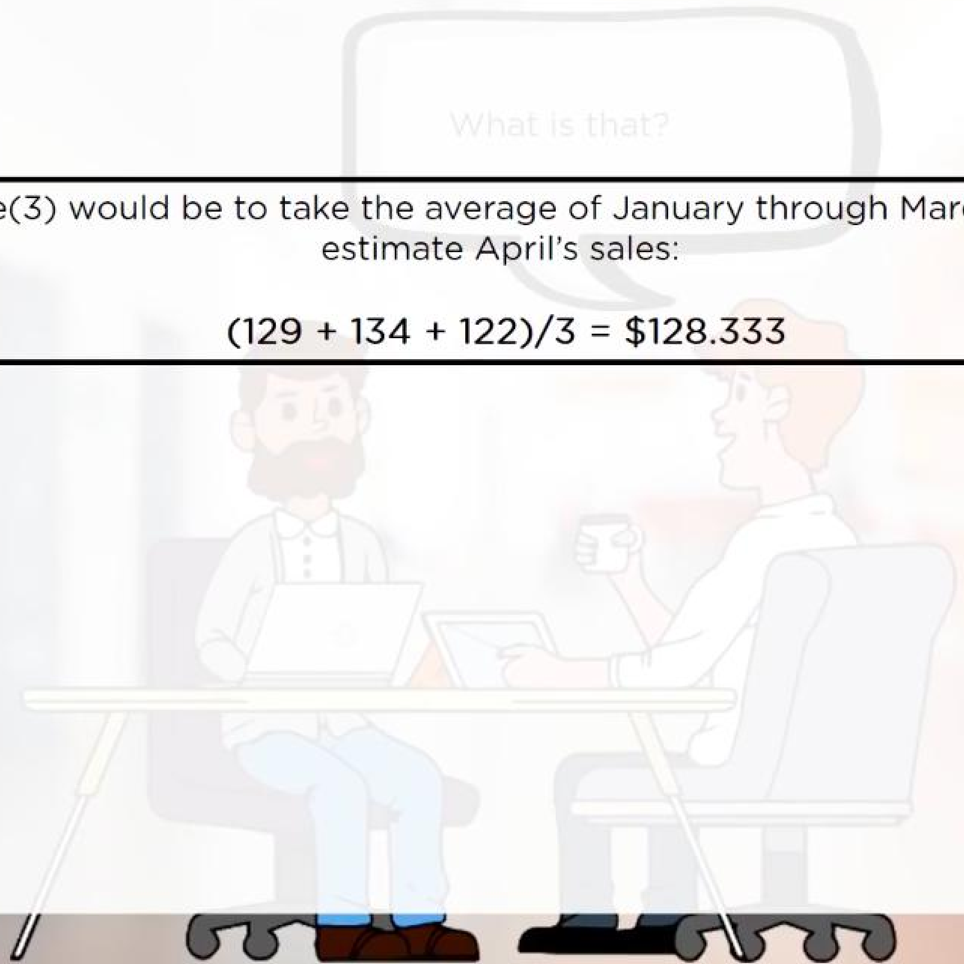

Before I tell you how to make it stationary and model it, let's look at Moving Average method
©Simplilearn. All rights reserved.
simplilearn


What is that?

A moving average is a technique to get an overall idea of the trends in a data set; it is an average of any subset of numbers

Let's say that you have been in business for three months, January through March, and wanted to forecast April's sales. Your sales for the last three months look like this:

Month	Sales ($000)
January	129
February	134
March	122

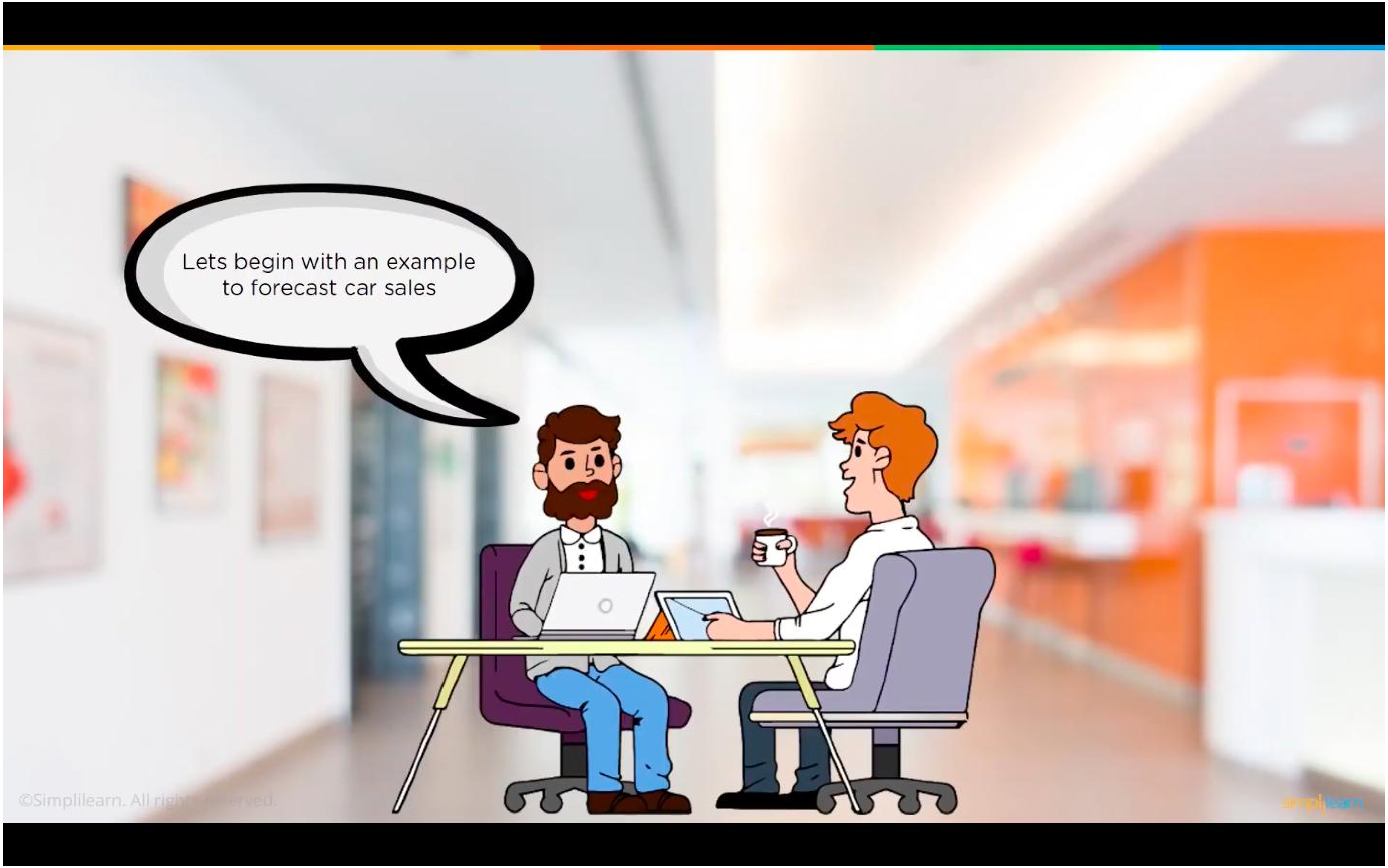


Moving Average(3) would be to take the average of January through March and use that to estimate April's sales:

$$(129 + 134 + 122)/3 = \$128.333$$

Hence, based on the sales of January through March, you predict that sales in April will be $128.333

Once April's actual sales come in, you would then compute the forecast for May, this time using February through April


Lets begin with an example
to forecast car sales
©Simplilearn. All rights reserved.
simplilearn

Here, we have Quarterly sales data for car sales.

Let's look at the data to see how it moves through time

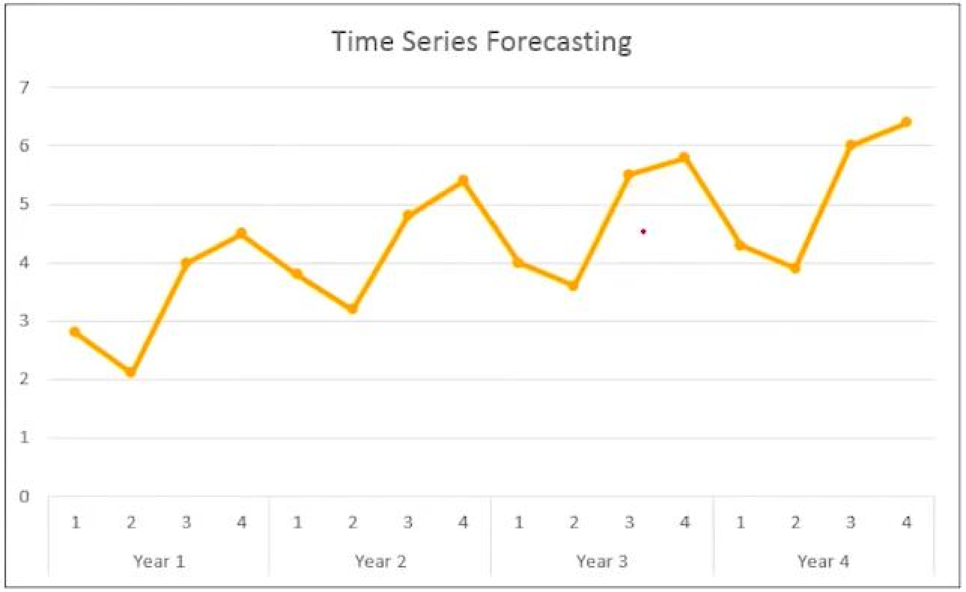
Year	Quarter	Sales(1000s)
Year 1	1	2.8
	2	2.1
	3	4
	4	4.5
Year 2	1	3.8
	2	3.2
	3	4.8
	4	5.4
Year 3	1	4
	2	3.6
	3	5.5
	4	5.8
Year 4	1	4.3
	2	3.9
	3	6
	4	6.4

We want to forecast the sales for the next four quarters (or the fifth year)

Year	Quarter	Sales(1000s)
Year 1	1	2.8
	2	2.1
	3	4
	4	4.5
Year 2	1	3.8
	2	3.2
	3	4.8
	4	5.4
Year 3	1	4
	2	3.6
	3	5.5
	4	5.8
Year 4	1	4.3
	2	3.9
	3	6
	4	6.4
Year 5	1	?
	2	?
	3	?
	4	?

Example to forecast Time Series

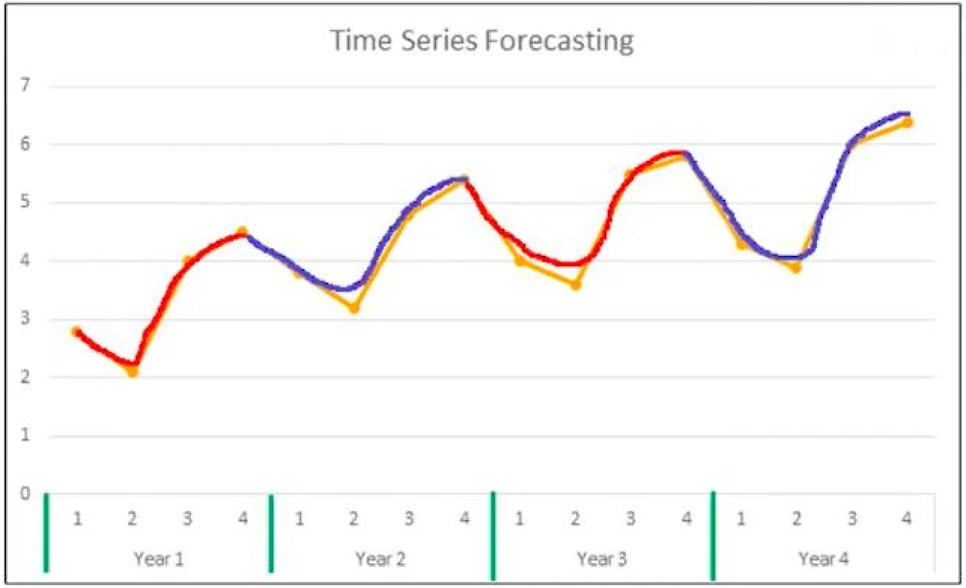
Plot of our Time Series data for the first four years looks like:



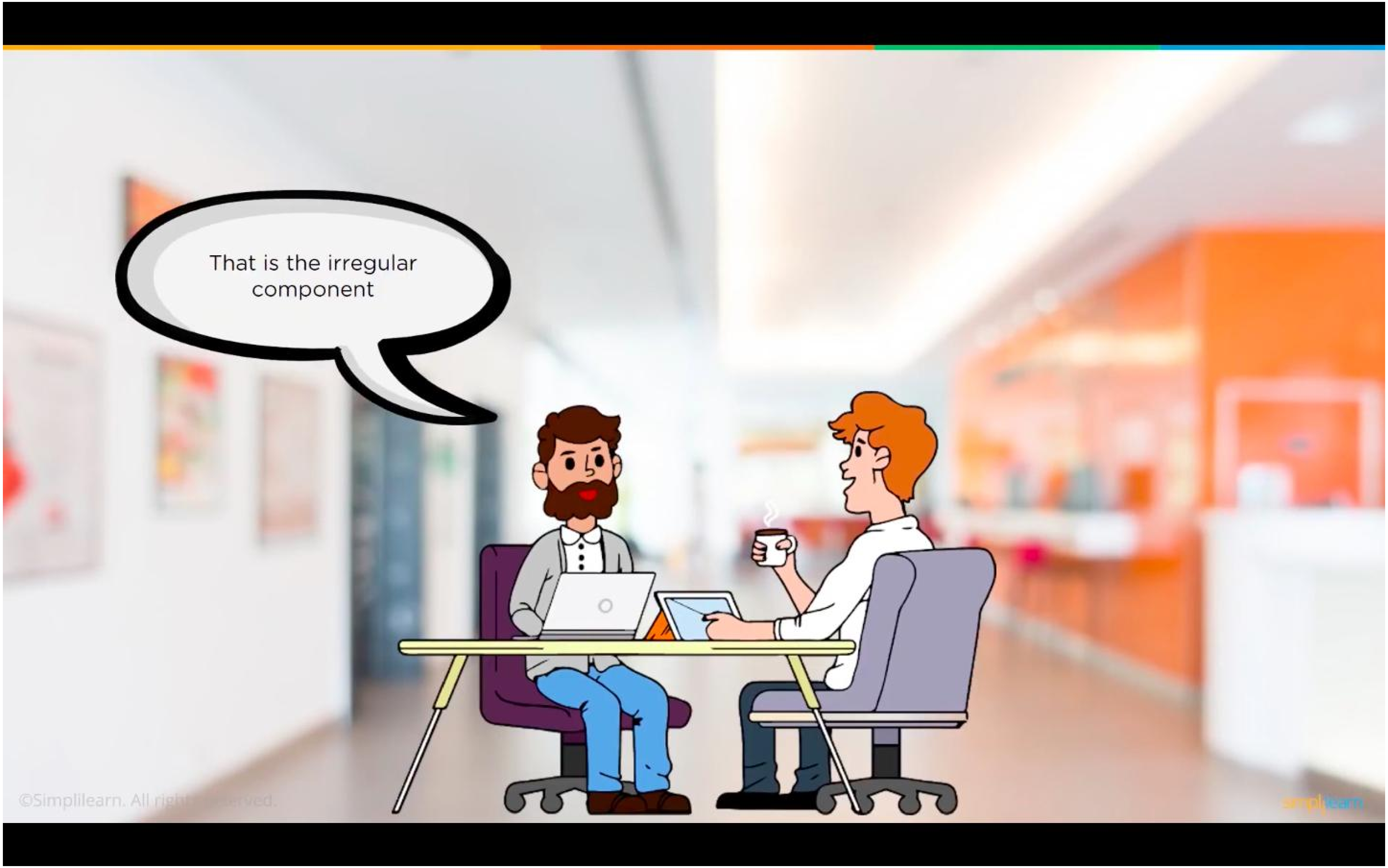
simplilearn

Example to forecast Time Series

Clearly, we see that the curve repeats in a pattern after every quarter (yearly) indicating seasonality:




And even though the pattern is clearly visible there is no data as perfect data, it will have some variability


That is the irregular component
©Simplilearn. All rights reserved.

We will give a time code(t) variable to each row indicating each time period, we will use this variable later:
this variable later:

t	Year	Quarter	Sales(1000s)
1	Year 1	1	2.8
2		2	2.1
3		3	4
4		4	4.5
5	Year 2	1	3.8
6		2	3.2
7		3	4.8
8		4	5.4
9	Year 3	1	4
10		2	3.6
11		3	5.5
12		4	5.8
13	Year 4	1	4.3
14		2	3.9
15		3	6
16		4	6.4

Here, we have calculate Moving Average(4) for the first four quarters:

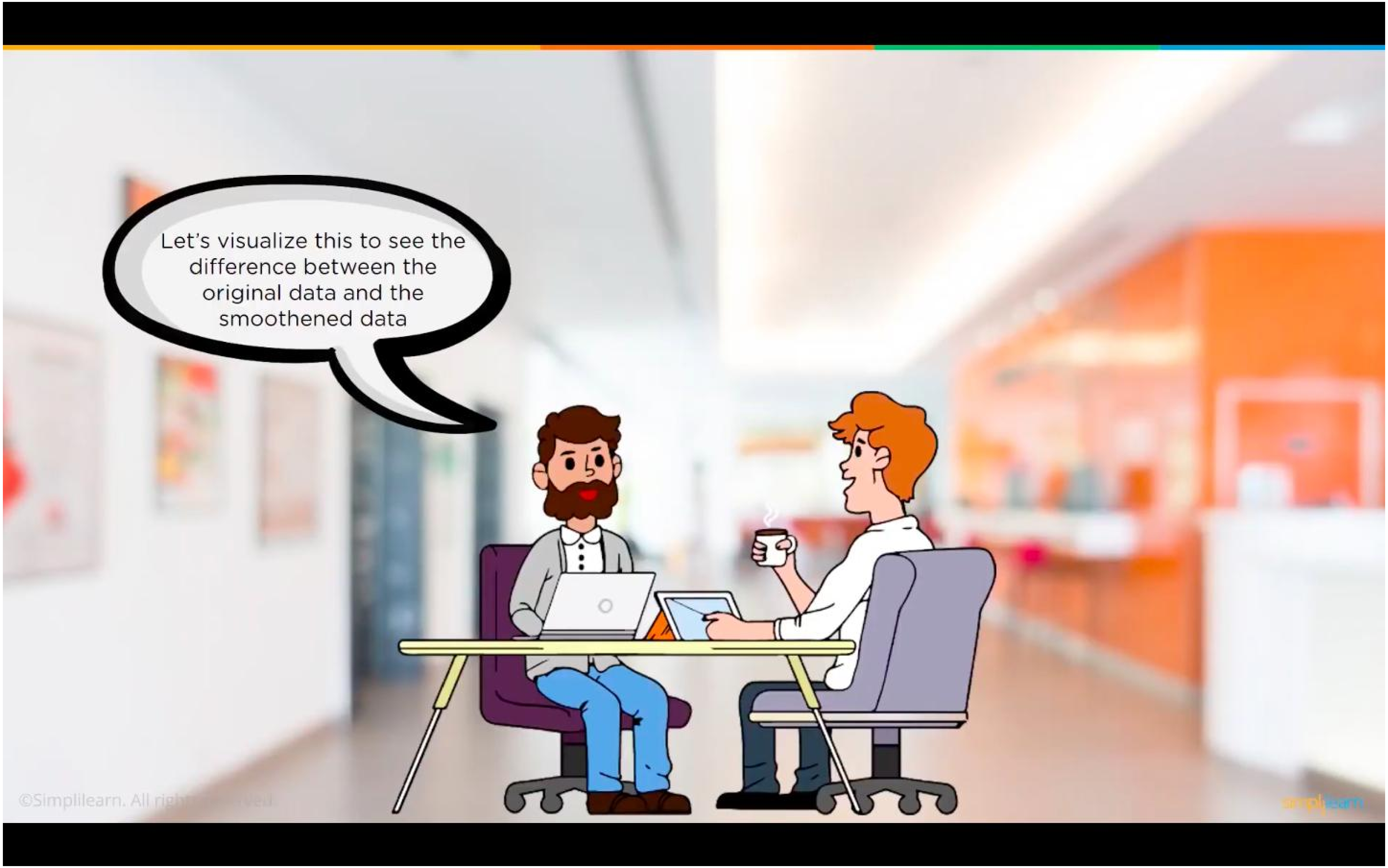
t	Year	Quarter	Sales(1000s)	MA(4)
1	Year 1	1	2.8	
2		2	2.1	
3		3	4	3.4
4		4	4.5	3.6
5	Year 2	1	3.8	3.9
6		2	3.2	4.1
7		3	4.8	4.3
8		4	5.4	4.4
9	Year 3	1	4	4.5
10		2	3.6	4.6
11		3	5.5	4.7
12		4	5.8	4.8
13	Year 4	1	4.3	4.9
14		2	3.9	5.0
15		3	6	5.2
16		4	6.4	

(2.8+2.1+4+4.5)/4

Now, since the moving average is not centered because of even number of data points, we will have to calculate "Centered Moving Average" as shown:

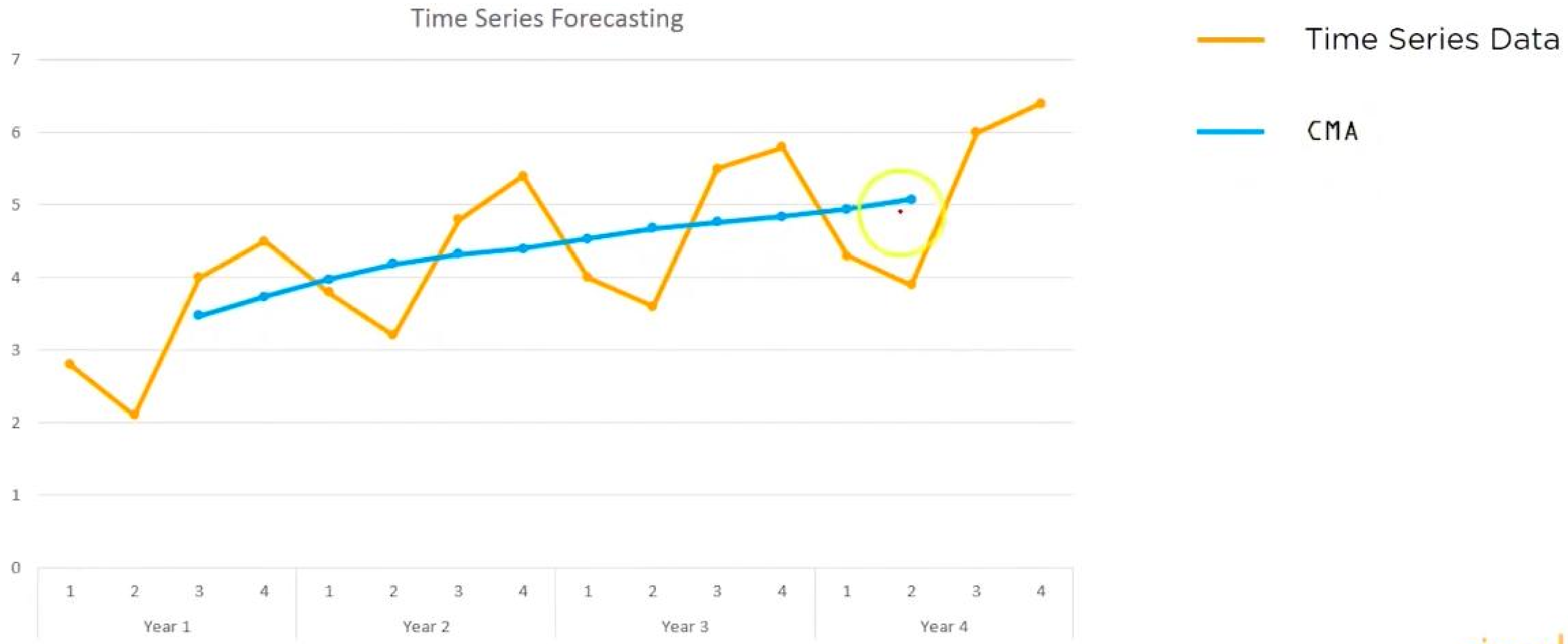
t	Year	Quarter	Sales(1000s)	MA(4)	CMA
1	Year 1	1	2.8		
2		2	2.1		
3		3	4	3.4	3.5
4		4	4.5	3.6	3.7
5	Year 2	1	3.8	3.9	4.0
6		2	3.2	4.1	4.2
7		3	4.8	4.3	4.3
8		4	5.4	4.4	4.4
9	Year 3	1	4	4.5	4.5
10		2	3.6	4.6	4.7
11		3	5.5	4.7	4.8
12		4	5.8	4.8	4.8
13	Year 4	1	4.3	4.9	4.9
14		2	3.9	5.0	5.1
15		3	6	5.2	
16		4	6.4		


How do you know that the data is smoothened?


Let's visualize this to see the difference between the original data and the smoothened data

Example to forecast Time Series

The blue line which represents "Centered Moving Average" is smoothened as compared to the original data, as we can see in the plot:



Example to forecast Time Series

In classical multiplicative model, Time Series value at time t = $S_t * I_t * T_t$

So, we calculate Seasonal and Irregular components by:

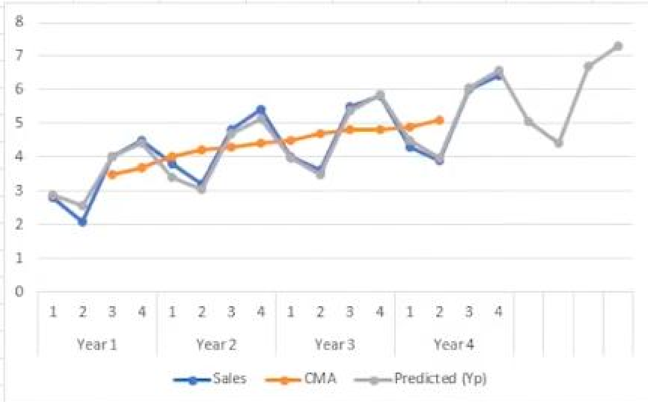
$S_t, I_t = Y_t / CMA$

		Y_t		baseline	Y_t/CMA
Year	Quarter	Sales(100 0s)	MA(4)	CMA	S_t, I_t
Year 1	1	2.8			
	2	2.1			
	3	4	3.4	3.5	1.15
	4	4.5	3.6	3.7	1.20
Year 2	1	3.8	3.9	4.0	0.96
	2	3.2	4.1	4.2	0.76
	3	4.8	4.3	4.3	1.11
	4	5.4	4.4	4.4	1.23
Year 3	1	4	4.5	4.5	0.88
	2	3.6	4.6	4.7	0.77
	3	5.5	4.7	4.8	1.15
	4	5.8	4.8	4.8	1.20
Year 4	1	4.3	4.9	4.9	0.87
	2	3.9	5.0	5.1	0.77
	3	6	5.2		
	4	6.4			

timeseries analysis wb - Excel

4R x 1C | f_x 5.05

	A	B	C	D	E	F	G	H	I	J	K
1				Yt		baseline	Yt/CMA	avg(respective qu	Yt/St	Intercept + slope * time code	
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasonalize	Trend	Predicted (Yp)
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89
4	2		2	2.1				0.77	2.73	3.36	2.59
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	4
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72	3.66	4.43
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22	3.81	3.43
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16	3.96	3.05
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	4.68
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26	5.15
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44	4.4	3.96
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68	4.55	3.51
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82	4.7	5.36
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79	4.85	5.87
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78	5	4.5
16	14		2	3.9	5	5.1	0.76	0.77	5.06	5.15	3.96
17	15		3	6	5.2			1.14	5.26	5.3	6.04
18	16		4	6.4				1.21	5.29	5.45	6.59
19	17	Year 5	1	?				0.9		5.59	5.05
20	18		2	?				0.77		5.74	4.41
21	19		3	?				1.14		5.89	6.71
22	20		4	?				1.21		6.04	[illegible].31
28	SUMMARY OUTPUT										
30	Regression Statistics										



	O	P
19	Quarter	St
20	1	0.91
21	2	0.76
22	3	1.13
23	4	1.22

timeseries analysis

READY AVERAGE: 5.87 COUNT: 4 SUM: 23.48 100%

simplilearn

Example to forecast Time Series

Now, let's get rid of the irregular component.
We can calculate only the seasonal component of every quarter by averaging each quarter per year

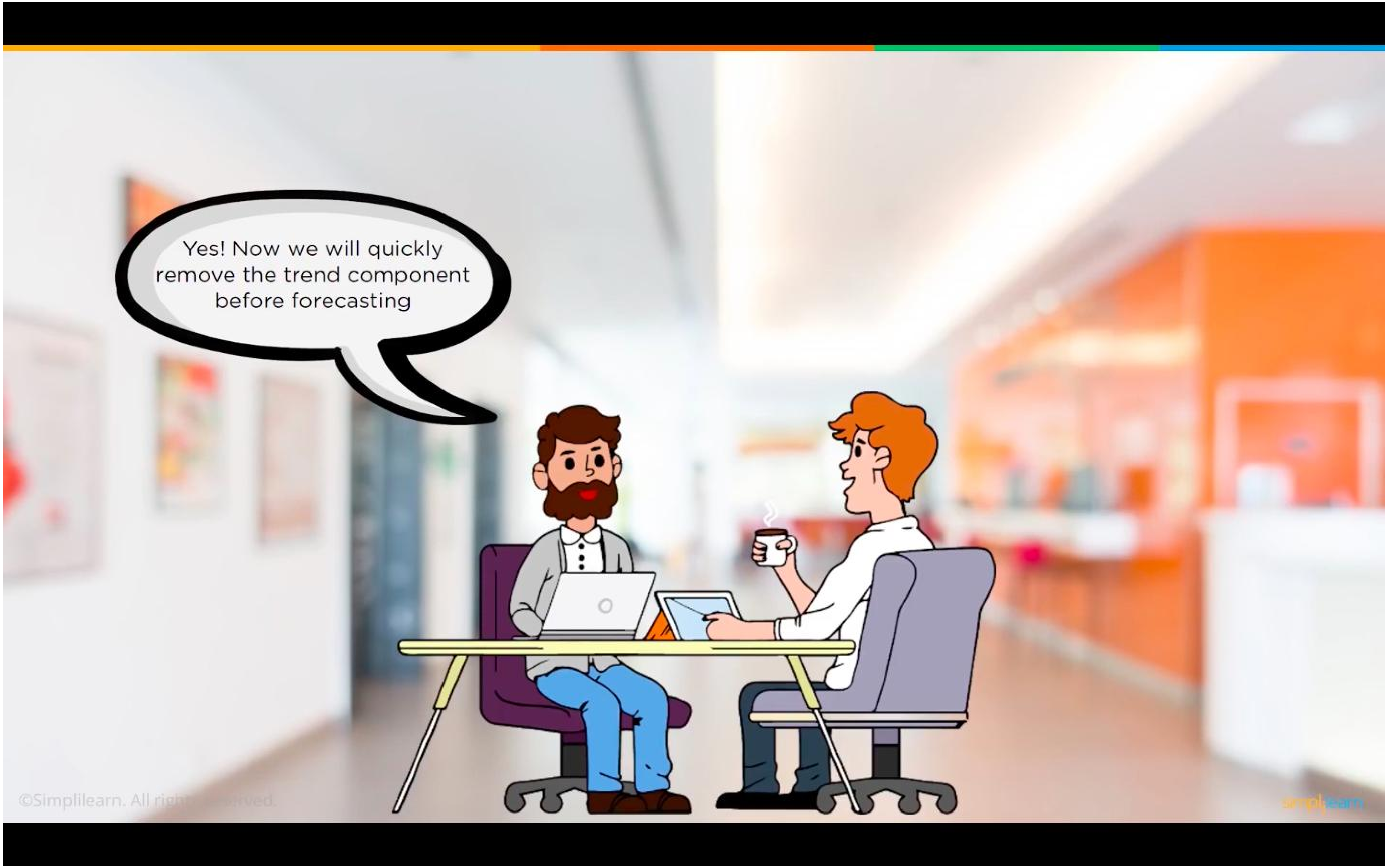
Year	Quarter	Sales(1000s)	MA(4)	CMA	S_t, I_t	S_t
Year 1	1	2.8				0.90
	2	2.1				0.77
	3	4	3.4	3.5	1.15	1.14
	4	4.5	3.6	3.7	1.20	1.21
Year 2	1	3.8	3.9	4.0	0.96	0.90
	2	3.2	4.1	4.2	0.76	0.77
	3	4.8	4.3	4.3	1.11	1.14
	4	5.4	4.4	4.4	1.23	1.21
Year 3	1	4	4.5	4.5	0.88	0.90
	2	3.6	4.6	4.7	0.77	0.77
	3	5.5	4.7	4.8	1.15	1.14
	4	5.8	4.8	4.8	1.20	1.21
Year 4	1	4.3	4.9	4.9	0.87	0.90
	2	3.9	5.0	5.1	0.77	0.77
	3	6	5.2			1.14
	4	6.4				1.21

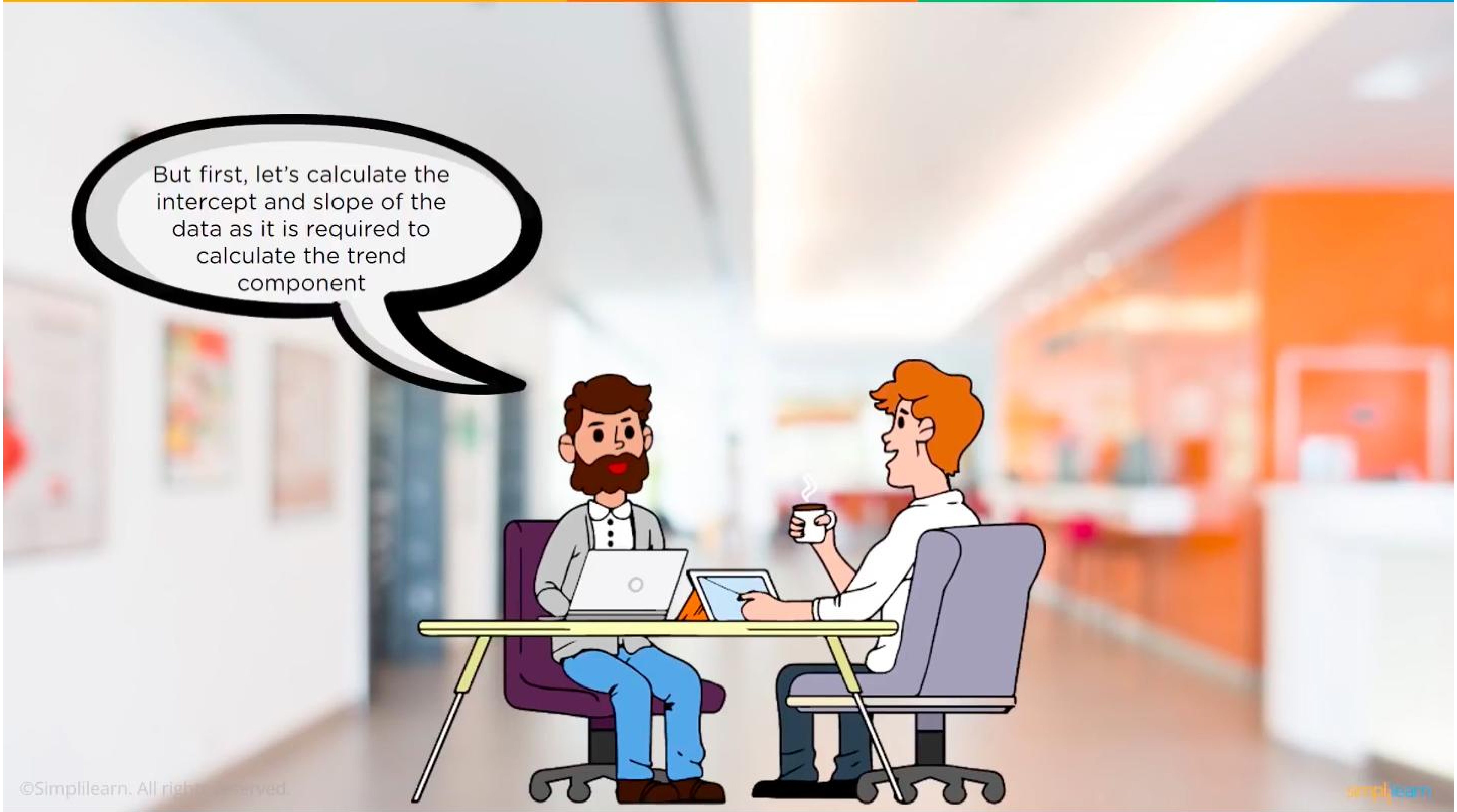
Example to forecast Time Series

Now, let's de-seasonalize the data:

		Y_t		baseline	Y_t/CMA		Yt/St
Year	Quarter	Sales(1000s)	MA(4)	CMA	S_t, I_t	S_t	De-seasonalize
Year 1	1	2.8				0.90	3.10
	2	2.1				0.77	2.74
	3	4	3.4	3.5	1.15	1.14	3.51
	4	4.5	3.6	3.7	1.20	1.21	3.72
Year 2	1	3.8	3.9	4.0	0.96	0.90	4.21
	2	3.2	4.1	4.2	0.76	0.77	4.17
	3	4.8	4.3	4.3	1.11	1.14	4.22
	4	5.4	4.4	4.4	1.23	1.21	4.46
Year 3	1	4	4.5	4.5	0.88	0.90	4.43
	2	3.6	4.6	4.7	0.77	0.77	4.69
	3	5.5	4.7	4.8	1.15	1.14	4.83
	4	5.8	4.8	4.8	1.20	1.21	4.79
Year 4	1	4.3	4.9	4.9	0.87	0.90	4.76
	2	3.9	5.0	5.1	0.77	0.77	5.08
	3	6	5.2			1.14	5.27
	4	6.4				1.21	5.29


So, we have finally got rid of the seasonal and the irregular components!
simplilearn


Yes! Now we will quickly remove the trend component before forecasting


But first, let's calculate the intercept and slope of the data as it is required to calculate the trend component
simplilearn

Example to forecast Time Series

Using simple regression, we will calculate the regression statistics using 'Data Analysis' feature in excel!

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.951127049
R Square	0.904642664
Adjusted R Square	0.897831426
Standard Error	0.238070383
Observations	16

ANOVA

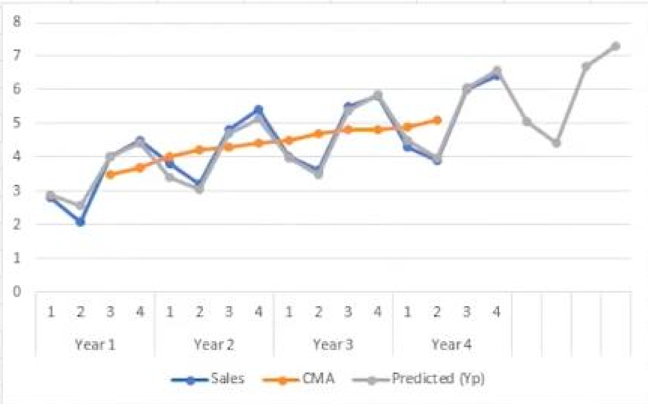
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.52769011	7.52769011	132.8161818	1.56872E-08
Residual	14	0.793485102	0.056677507		
Total	15	8.321175212			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.064766682	0.124845162	24.54854182	6.57968E-13	2.79700044	3.332532924	2.79700044	3.332532924
t	0.148796052	0.012911179	11.52459031	1.56872E-08	0.121104326	0.176487777	0.121104326	0.176487777

timeseries analysis wb - Excel

J3 = 3.21

	A	B	C	D	E	F	G	H	I	J	K
1				Yt		baseline	Yt/CMA	avg(respe	Yt/St	Intercept + slope * time code	
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasoı	Trend	Predicted (Yp)
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89
4	2		2	2.1				0.77	2.73	3.36	2.59
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	4
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72	3.66	4.43
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22	3.81	3.43
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16	3.96	3.05
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	4.68
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26	5.15
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44	4.4	3.96
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68	4.55	3.51
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82	4.7	5.36
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79	4.85	5.87
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78	5	4.5
16	14		2	3.9	5	5.1	0.76	0.77	5.06	5.15	3.96
17	15		3	6	5.2			1.14	5.26	5.3	6.04
18	16		4	6.4				1.21	5.29	5.45	6.59
19	17	Year 5	1	?				0.9		5.59	5.05
20	18		2	?				0.77		5.74	4.41
21	19		3	?				1.14		5.89	6.71
22	20		4	?				1.21		6.04	7.31
28	SUMMARY OUTPUT										
30	Regression Statistics										

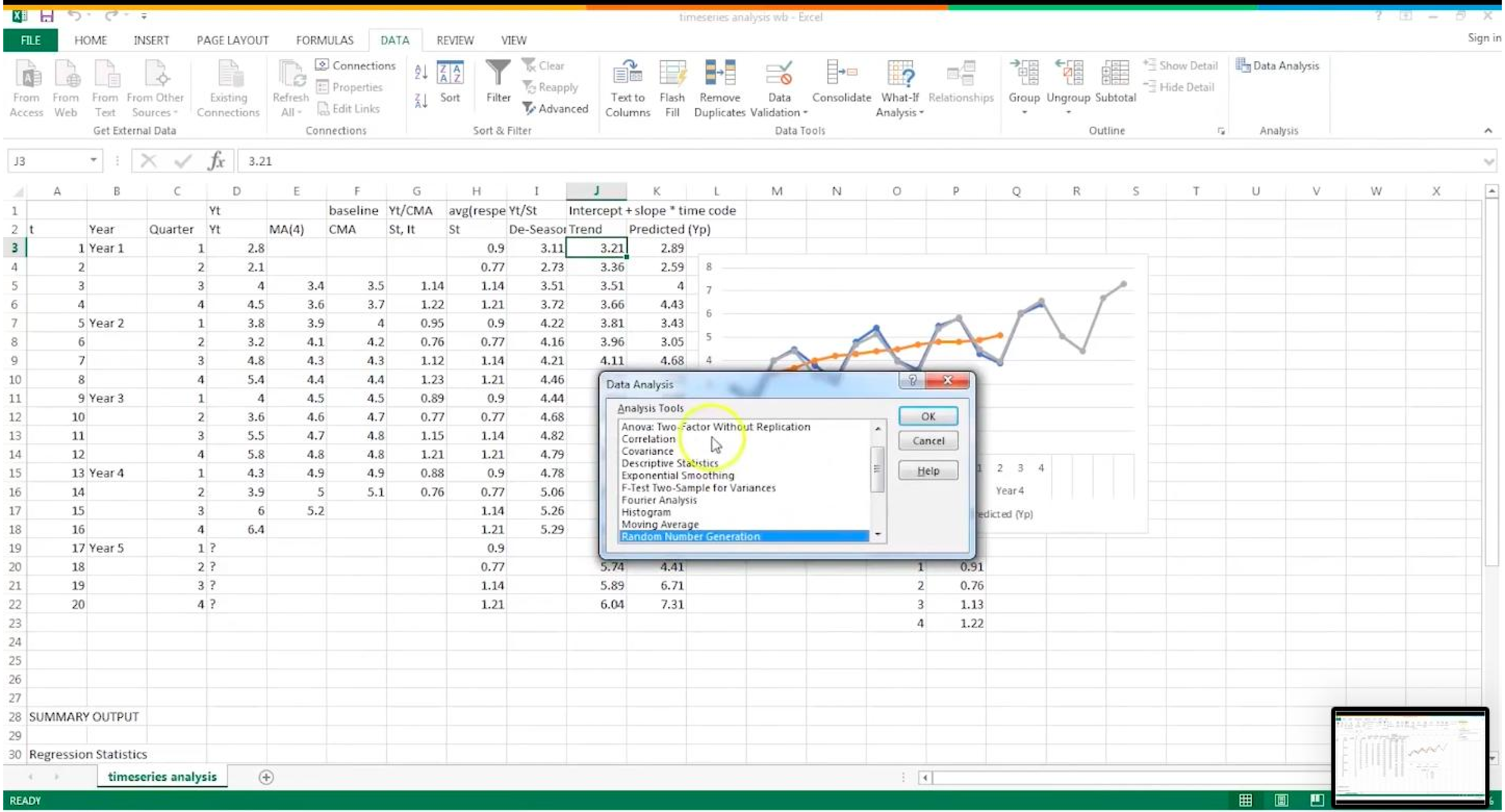


Quarter	St
1	0.91
2	0.76
3	1.13
4	1.22

Data Analysis Tools

Tools for financial and scientific data analysis.

timeseries analysis



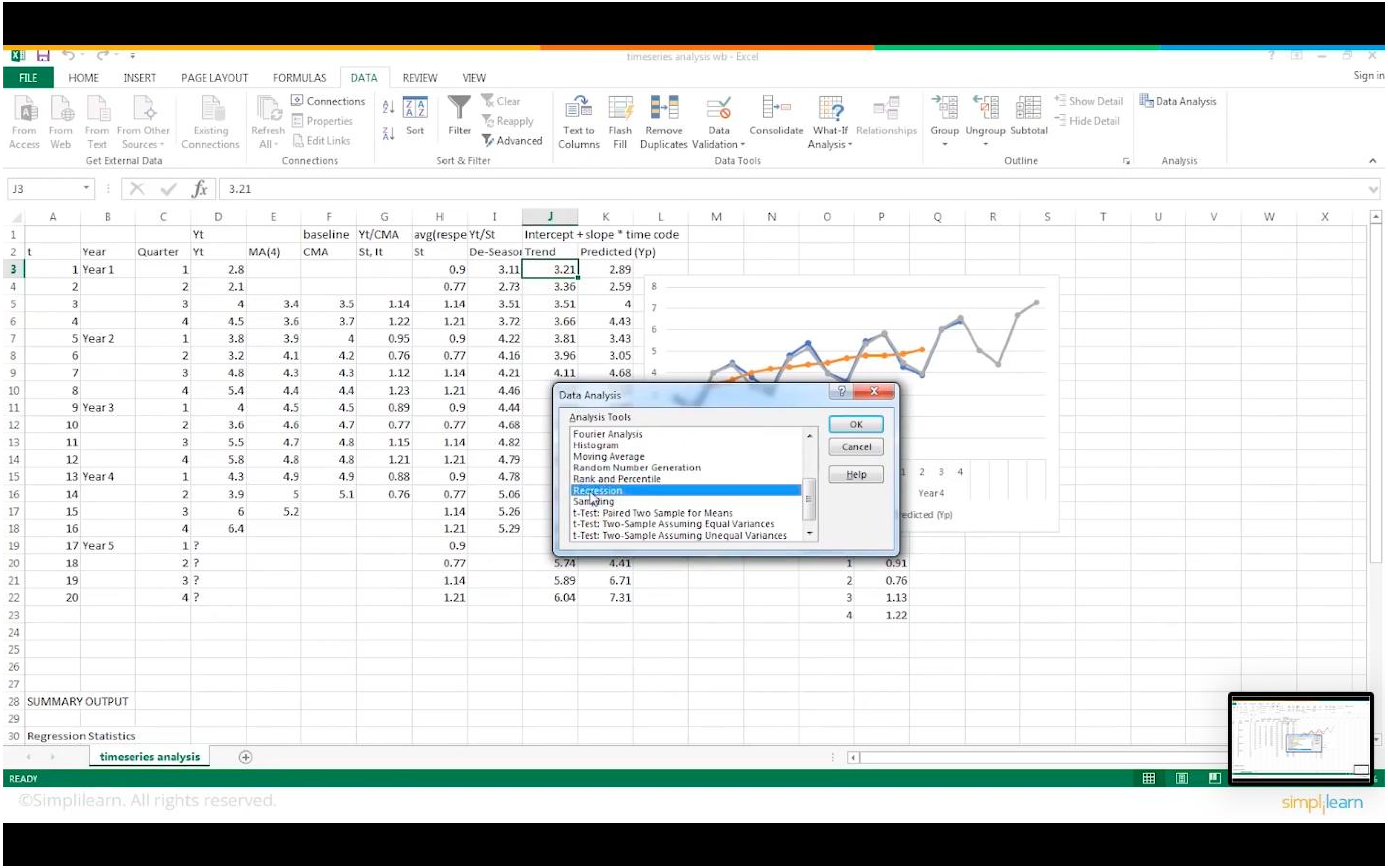
	A	B	C	D	E	F	G	H	I	J	K
1				Yt		baseline	Yt/CMA	avg(respe	Yt/St	Intercept + slope * time code	
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasoı	Trend	Predicted (Yp)
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89
4	2		2	2.1				0.77	2.73	3.36	2.59
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	4
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72	3.66	4.43
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22	3.81	3.43
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16	3.96	3.05
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	4.68
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46		
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44		
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68		
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82		
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79		
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78		
16	14		2	3.9	5	5.1	0.76	0.77	5.06		
17	15		3	6	5.2			1.14	5.26		
18	16		4	6.4				1.21	5.29		
19	17	Year 5	1	?				0.9			
20	18		2	?				0.77		5.74	4.41
21	19		3	?				1.14		5.89	6.71
22	20		4	?				1.21		6.04	7.31

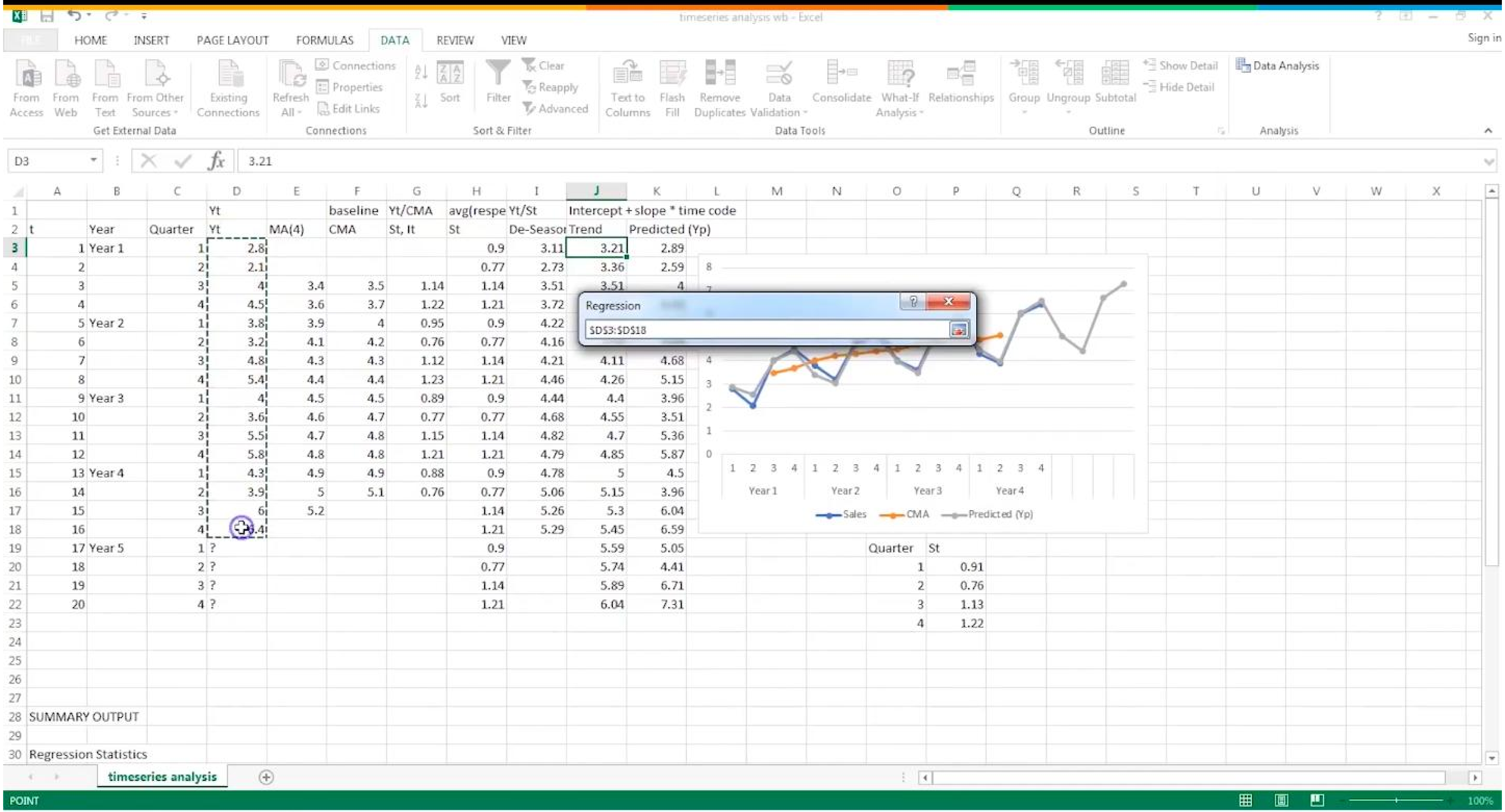
	O	P
20	1	0.91
21	2	0.76
22	3	1.13
23	4	1.22

28 SUMMARY OUTPUT

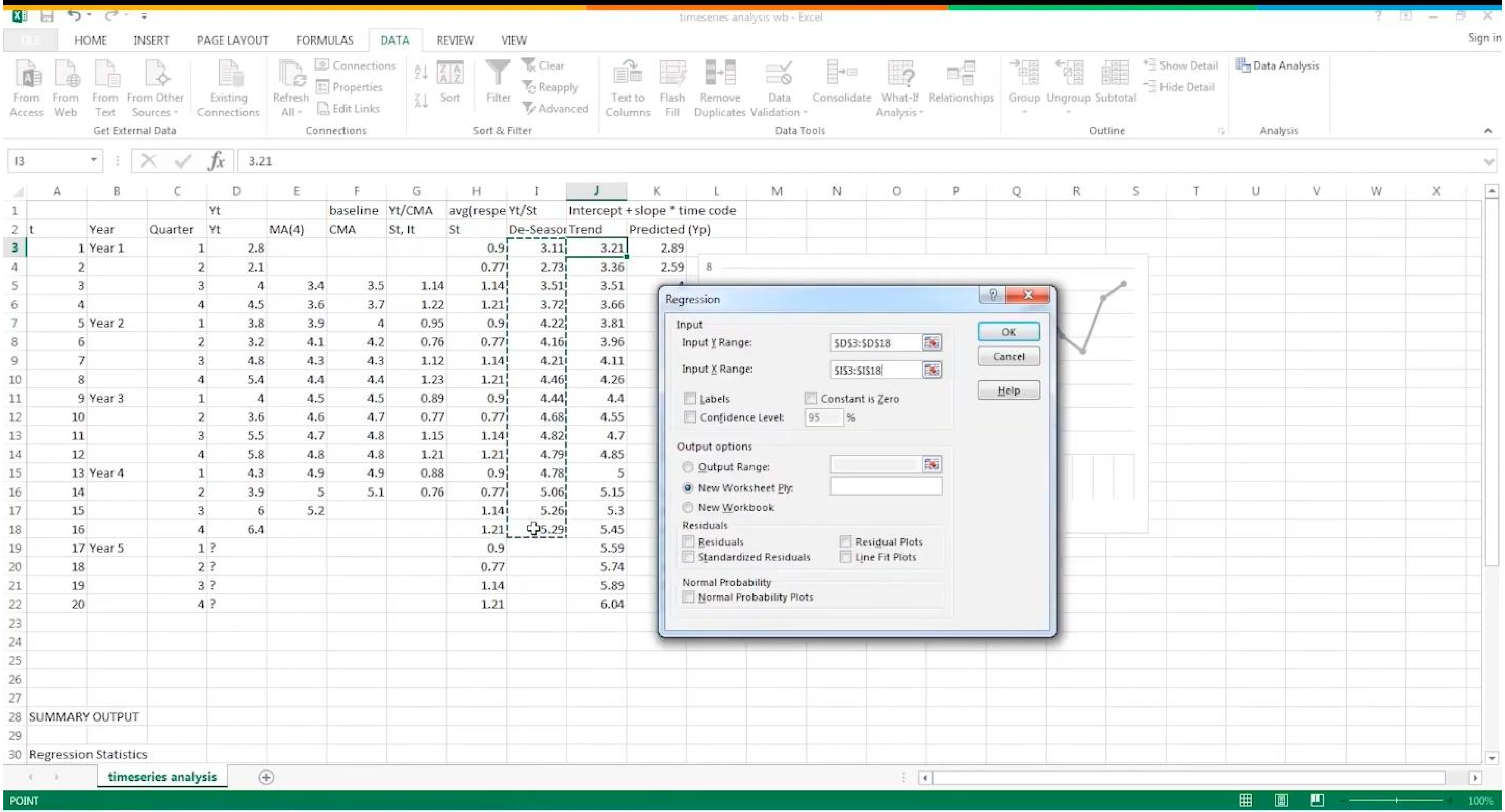
30 Regression Statistics

simplilearn


timeseries analysis wb - Excel
FILE
HOME
INSERT
PAGE LAYOUT
FORMULAS
DATA
REVIEW
VIEW
Sign in
Get External Data
Connections
Sort & Filter
Data Tools
Outline
Analysis
Data Analysis
J3
3.21
Intercept + slope * time code
SUMMARY OUTPUT
Regression Statistics
Data Analysis
Analysis Tools
Fourier Analysis
Histogram
Moving Average
Random Number Generation
Rank and Percentile
Regression
Sampling
t-Test: Paired Two Sample for Means
t-Test: Two-Sample Assuming Equal Variances
t-Test: Two-Sample Assuming Unequal Variances
OK
Cancel
Help
timeseries analysis
READY
©Simplilearn. All rights reserved.
simplilearn

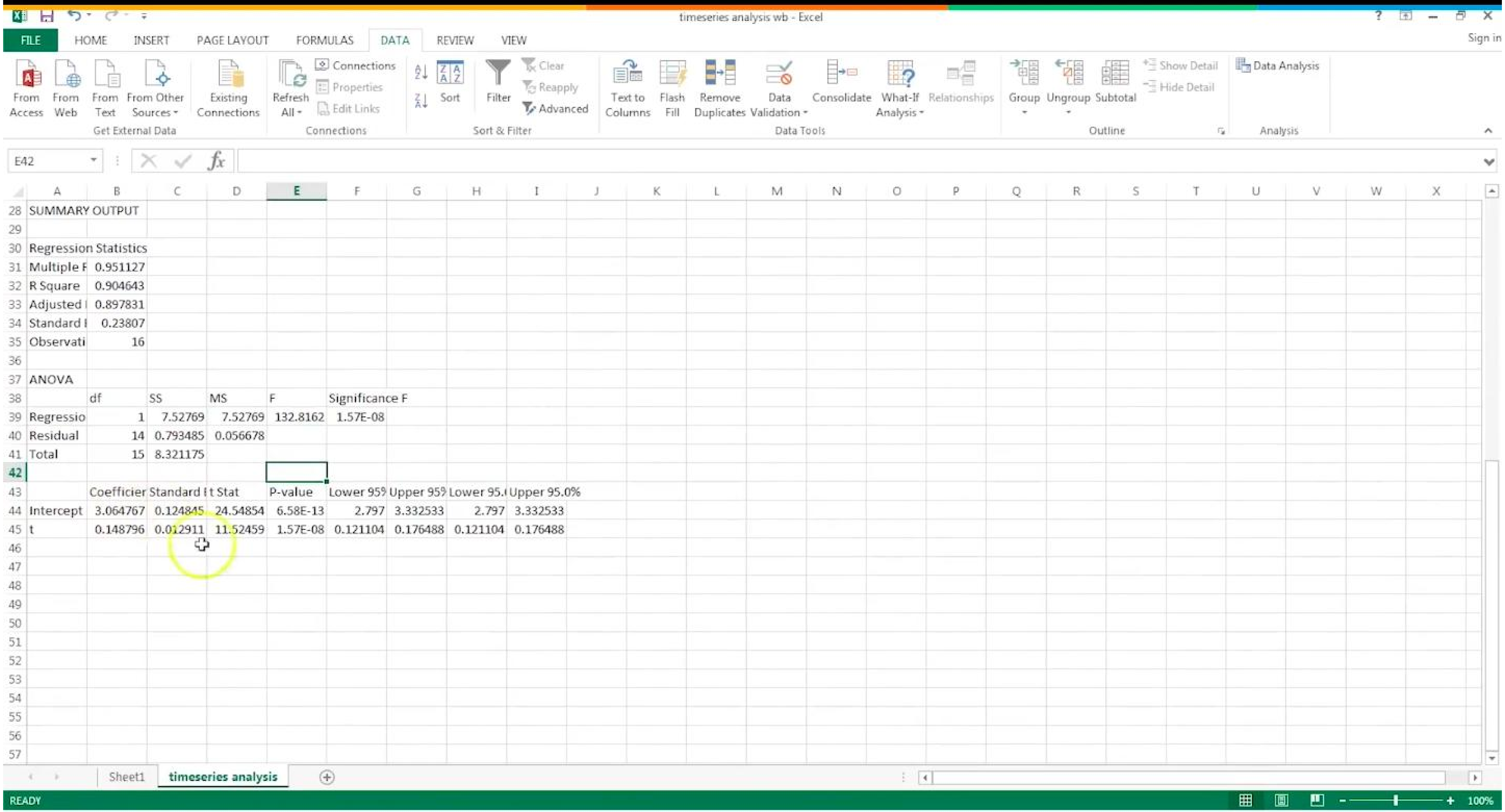


	A	B	C	D	E	F	G	H	I	J	K	O	P
1				Yt		baseline	Yt/CMA	avg(respe	Yt/St	Intercept + slope * time code			
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasoı	Trend	Predicted (Yp)		
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89		
4	2		2	2.1				0.77	2.73	3.36	2.59		
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	4		
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72				
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22				
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16				
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	4.68		
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26	5.15		
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44	4.4	3.96		
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68	4.55	3.51		
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82	4.7	5.36		
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79	4.85	5.87		
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78	5	4.5		
16	14		2	3.9	5	5.1	0.76	0.77	5.06	5.15	3.96		
17	15		3	6	5.2			1.14	5.26	5.3	6.04		
18	16		4	[illegible].4				1.21	5.29	5.45	6.59		
19	17	Year 5	1	?				0.9		5.59	5.05	Quarter	St
20	18		2	?				0.77		5.74	4.41	1	0.91
21	19		3	?				1.14		5.89	6.71	2	0.76
22	20		4	?				1.21		6.04	7.31	3	1.13
23												4	1.22
28	SUMMARY OUTPUT												
30	Regression Statistics												



	A	B	C	D	E	F	G	H	I	J	K
1				Yt		baseline	Yt/CMA	avg(respe	Yt/St	Intercept + slope * time code	
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasor	Trend	Predicted (Yp)
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89
4	2		2	2.1				0.77	2.73	3.36	2.59
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72	3.66	
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22	3.81	
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16	3.96	
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26	
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44	4.4	
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68	4.55	
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82	4.7	
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79	4.85	
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78	5	
16	14		2	3.9	5	5.1	0.76	0.77	5.06	5.15	
17	15		3	6	5.2			1.14	5.26	5.3	
18	16		4	6.4				1.21	5.29	5.45	
19	17	Year 5	1	?				0.9		5.59	
20	18		2	?				0.77		5.74	
21	19		3	?				1.14		5.89	
22	20		4	?				1.21		6.04	
28	SUMMARY OUTPUT										
30	Regression Statistics										

simplilearn



	A	B	C	D	E	F	G	H	I
28	SUMMARY OUTPUT								
29									
30	Regression Statistics								
31	Multiple F	0.951127							
32	R Square	0.904643							
33	Adjusted I	0.897831							
34	Standard I	0.23807							
35	Observati	16							
36									
37	ANOVA								
38		df	SS	MS	F	Significance F			
39	Regressio	1	7.52769	7.52769	132.8162	1.57E-08			
40	Residual	14	0.793485	0.056678					
41	Total	15	8.321175						
42									
43		Coefficier	Standard E	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0	Upper 95.0%
44	Intercept	3.064767	0.124845	24.54854	6.58E-13	2.797	3.332533	2.797	3.332533
45	t	0.148796	0.012911	11.52459	1.57E-08	0.121104	0.176488	0.121104	0.176488

timeseries analysis wb - Excel

	A	B	C	D	E	F	G	H	I	J	K
1				Yt		baseline	Yt/CMA	avg(respe	Yt/St	Intercept + slope * time code	
2	t	Year	Quarter	Yt	MA(4)	CMA	St, It	St	De-Seasor	Trend	Predicted (Yp)
3	1	Year 1	1	2.8				0.9	3.11	3.21	2.89
4	2		2	2.1				0.77	2.73	3.36	2.59
5	3		3	4	3.4	3.5	1.14	1.14	3.51	3.51	4
6	4		4	4.5	3.6	3.7	1.22	1.21	3.72	3.66	4.43
7	5	Year 2	1	3.8	3.9	4	0.95	0.9	4.22	3.81	3.43
8	6		2	3.2	4.1	4.2	0.76	0.77	4.16	3.96	3.05
9	7		3	4.8	4.3	4.3	1.12	1.14	4.21	4.11	4.68
10	8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26	5.15
11	9	Year 3	1	4	4.5	4.5	0.89	0.9	4.44	4.4	3.96
12	10		2	3.6	4.6	4.7	0.77	0.77	4.68	4.55	3.51
13	11		3	5.5	4.7	4.8	1.15	1.14	4.82	4.7	5.36
14	12		4	5.8	4.8	4.8	1.21	1.21	4.79	4.85	5.87
15	13	Year 4	1	4.3	4.9	4.9	0.88	0.9	4.78	5	4.5
16	14		2	3.9	5	5.1	0.76	0.77	5.06	5.15	3.96
17	15		3	6	5.2			1.14	5.26	5.3	6.04
18	16		4	6.4				1.21	5.29	5.45	6.59
19	17	Year 5	1	?				0.9		5.59	5.05
20	18		2	?				0.77		5.74	4.41
21	19		3	?				1.14		5.89	6.71
22	20		4	?				1.21		6.04	7.31
28	SUMMARY OUTPUT										
30	Regression Statistics										

Quarter	St
1	0.91
2	0.76
3	1.13
4	1.22

Sheet1 | timeseries analysis

Example to forecast Time Series

Using simple regression, we will calculate the regression statistics using 'Data Analysis' feature in excel!

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.951127049
R Square	0.904642664
Adjusted R Square	0.897831426
Standard Error	0.238070383
Observations	16

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	7.52769011	7.52769011	132.8161818	1.56872E-08
Residual	14	0.793485102	0.056677507		
Total	15	8.321175212			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	3.064766682	0.124845162	24.54854182	6.57968E-13	2.79700044	3.332532924	2.79700044	3.332532924
t	0.148796052	0.012911179	11.52459031	1.56872E-08	0.121104326	0.176487777	0.121104326	0.176487777

Intercept

Slope


So, to calculate trend, we will need intercept and slope as seen in previous slide
©Simplilearn. All rights reserved.
simplilearn

Example to forecast Time Series

```
Trend = Intercept + Slope * Time code
```

t	Year	Quarter	Sales(1000s)	MA(4)	CMA	S_t, I_t	S_t	Seasonalize	T_t
1	Year 1	1	2.8				0.90	3.10	3.21
2		2	2.1				0.77	2.74	3.36
3		3	4	3.4	3.5	1.15	1.14	3.51	3.51
4		4	4.5	3.6	3.7	1.20	1.21	3.72	3.66
5	Year 2	1	3.8	3.9	4.0	0.96	0.90	4.21	3.81
6		2	3.2	4.1	4.2	0.76	0.77	4.17	3.96
7		3	4.8	4.3	4.3	1.11	1.14	4.22	4.11
8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26
9	Year 3	1	4	4.5	4.5	0.88	0.90	4.43	4.40
10		2	3.6	4.6	4.7	0.77	0.77	4.69	4.55
11		3	5.5	4.7	4.8	1.15	1.14	4.83	4.70
12		4	5.8	4.8	4.8	1.20	1.21	4.79	4.85
13	Year 4	1	4.3	4.9	4.9	0.87	0.90	4.76	5.00
14		2	3.9	5.0	5.1	0.77	0.77	5.08	5.15
15		3	6	5.2			1.14	5.27	5.30
16		4	6.4				1.21	5.29	5.45

Example to forecast Time Series

Trend = Intercept + Slope * Time code

t	Year	Quarter	Sales(1000s)	MA(4)	CMA	S_t, I_t	S_t	Seasonalize	T_t
1	Year 1	1	2.8				0.90	3.10	3.21
2		2	2.1				0.77	2.74	3.36
3		3	4	3.4	3.5	1.15	1.14	3.51	3.51
4		4	4.5	3.6	3.7	1.20	1.21	3.72	3.66
5	Year 2	1	3.8	3.9	4.0	0.96	0.90	4.21	3.81
6		2	3.2	4.1	4.2	0.76	0.77	4.17	3.96
7		3	4.8	4.3	4.3	1.11	1.14	4.22	4.11
8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26
9	Year 3	1	4	4.5	4.5	0.88	0.90	4.43	4.40
10		2	3.6	4.6	4.7	0.77	0.77	4.69	4.55
11		3	5.5	4.7	4.8	1.15	1.14	4.83	4.70
12		4	5.8	4.8	4.8	1.20	1.21	4.79	4.85
13	Year 4	1	4.3	4.9	4.9	0.87	0.90	4.76	5.00
14		2	3.9	5.0	5.1	0.77	0.77	5.08	5.15
15		3	6	5.2			1.14	5.27	5.30
16		4	6.4				1.21	5.29	5.45

Trend Component = Intercept + Slope * Time Code

simplilearn

So, using the multiplicative model, we can make the predictions.
To do that, we have to combine all the components that we separated

Predicted (Y_p) = Seasonal component (S_t) * Trend Component (T_t)

Example to forecast Time Series

We see that the predicted values(Yp) are similar to the actual values(Yt)
Hence, our calculations are correct

t	Year	Quarter	Y_t	S_t	De-Seasonalize	T_t	Predicted (Y_p)
1	Year 1	1	2.8	0.90	3.10	3.21	2.90
2		2	2.1	0.77	2.74	3.36	2.58
3		3	4	1.14	3.51	3.51	4.00
4		4	4.5	1.21	3.72	3.66	4.43
5	Year 2	1	3.8	0.90	4.21	3.81	3.44
6		2	3.2	0.77	4.17	3.96	3.04
7		3	4.8	1.14	4.22	4.11	4.68
8		4	5.4	1.21	4.46	4.26	5.15
9	Year 3	1	4	0.90	4.43	4.40	3.98
10		2	3.6	0.77	4.69	4.55	3.49
11		3	5.5	1.14	4.83	4.70	5.35
12		4	5.8	1.21	4.79	4.85	5.87
13	Year 4	1	4.3	0.90	4.76	5.00	4.51
14		2	3.9	0.77	5.08	5.15	3.95
15		3	6	1.14	5.27	5.30	6.03
16		4	6.4	1.21	5.29	5.45	6.59


Now, let's forecast the values for the four quarters of the 5th year


Now, let's forecast the values for the four quarters of the 5th year


Now, let's forecast the values for the four quarters of the 5th year


As you know the seasonal component will remain the same for the Time Series Data
©Simplilearn. All rights reserved.
simplilearn

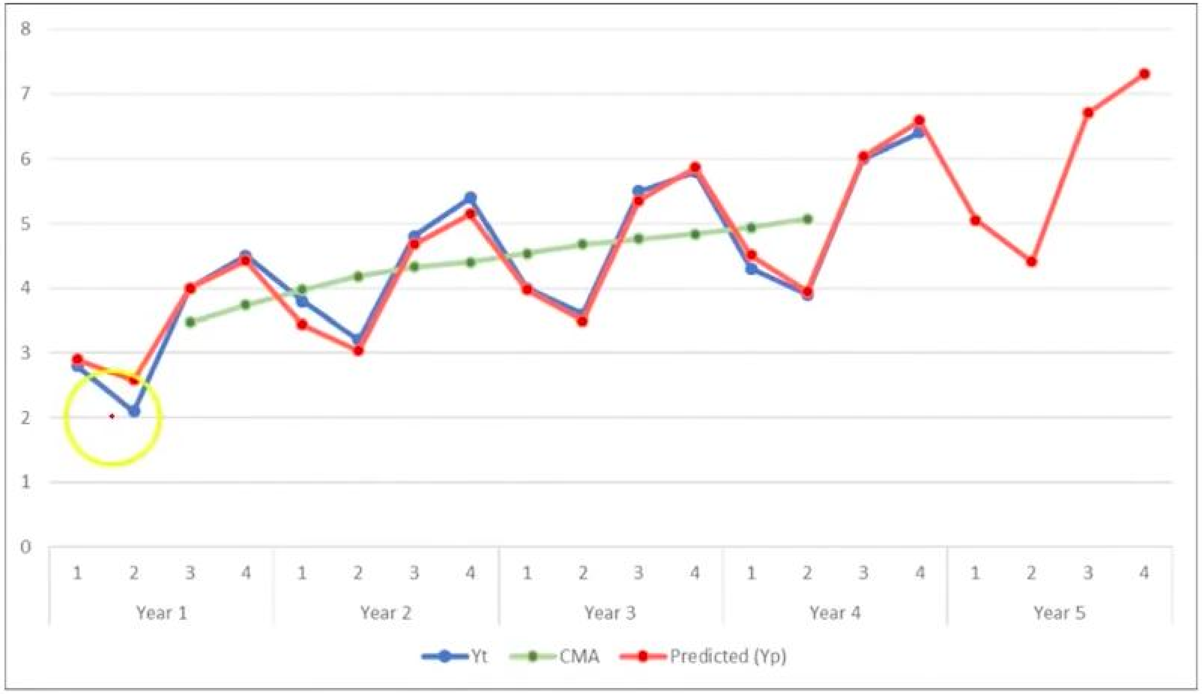

And using that seasonal component, intercept and slope, we can calculate the trend component easily
©Simplilearn. All rights reserved.
simplilearn

t	Year	Quarter	Y_t	MA(4)	CMA	S_t, I_t	S_t	De-Seasonalize	T_t
1	Year 1	1	2.8				0.90	3.10	3.21
2		2	2.1				0.77	2.74	3.36
3		3	4	3.4	3.5	1.15	1.14	3.51	3.51
4		4	4.5	3.6	3.7	1.20	1.21	3.72	3.66
5	Year 2	1	3.8	3.9	4.0	0.96	0.90	4.21	3.81
6		2	3.2	4.1	4.2	0.76	0.77	4.17	3.96
7		3	4.8	4.3	4.3	1.11	1.14	4.22	4.11
8		4	5.4	4.4	4.4	1.23	1.21	4.46	4.26
9	Year 3	1	4	4.5	4.5	0.88	0.90	4.43	4.40
10		2	3.6	4.6	4.7	0.77	0.77	4.69	4.55
11		3	5.5	4.7	4.8	1.15	1.14	4.83	4.70
12		4	5.8	4.8	4.8	1.20	1.21	4.79	4.85
13	Year 4	1	4.3	4.9	4.9	0.87	0.90	4.76	5.00
14		2	3.9	5.0	5.1	0.77	0.77	5.08	5.15
15		3	6	5.2			1.14	5.27	5.30
16		4	6.4				1.21	5.29	5.45
17	Year 5	1	?				0.90		5.59
18		2	?				0.77		5.74
19		3	?				1.14		5.89
20		4	?				1.21		6.04


Because to forecast, we only need these two component. So, let's not do the other unnecessary calculations!
©Simplilearn. All rights reserved.
simplilearn

Example to forecast Time Series

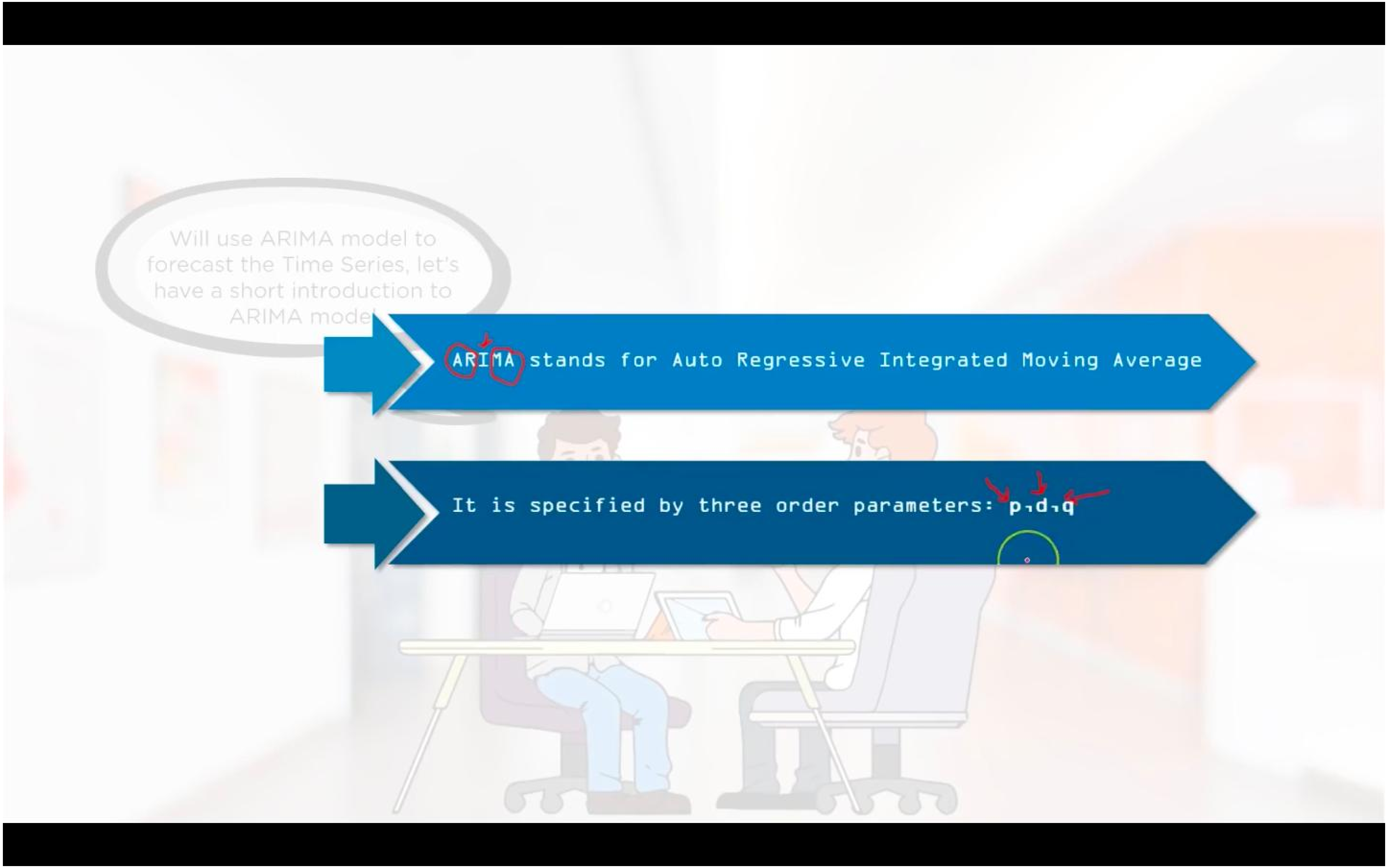
Here we can see that the predicted values overlap with the actual values and continues till year 5 which shows the accuracy of our forecasted values




Now, let's move on with our implementation of Time Series using R


Will use ARIMA model to forecast the Time Series, let's have a short introduction to ARIMA model


Will use ARIMA model to forecast the Time Series, let's have a short introduction to ARIMA model
ARIMA stands for Auto Regressive Integrated Moving Average
It is specified by three order parameters: p,d,q


Will use ARIMA model to
forecast the Time Series, let's
have a short introduction to
ARIMA mode
ARIMA stands for Auto Regressive Integrated Moving Average
It is specified by three order parameters: p,d,q

ARIMA models are classified by three factors:

p = number of autoregressive terms (AR),

d = how many non-seasonal differences are needed to achieve stationarity (I),

q = number of lagged forecast errors in the prediction equation (MA)

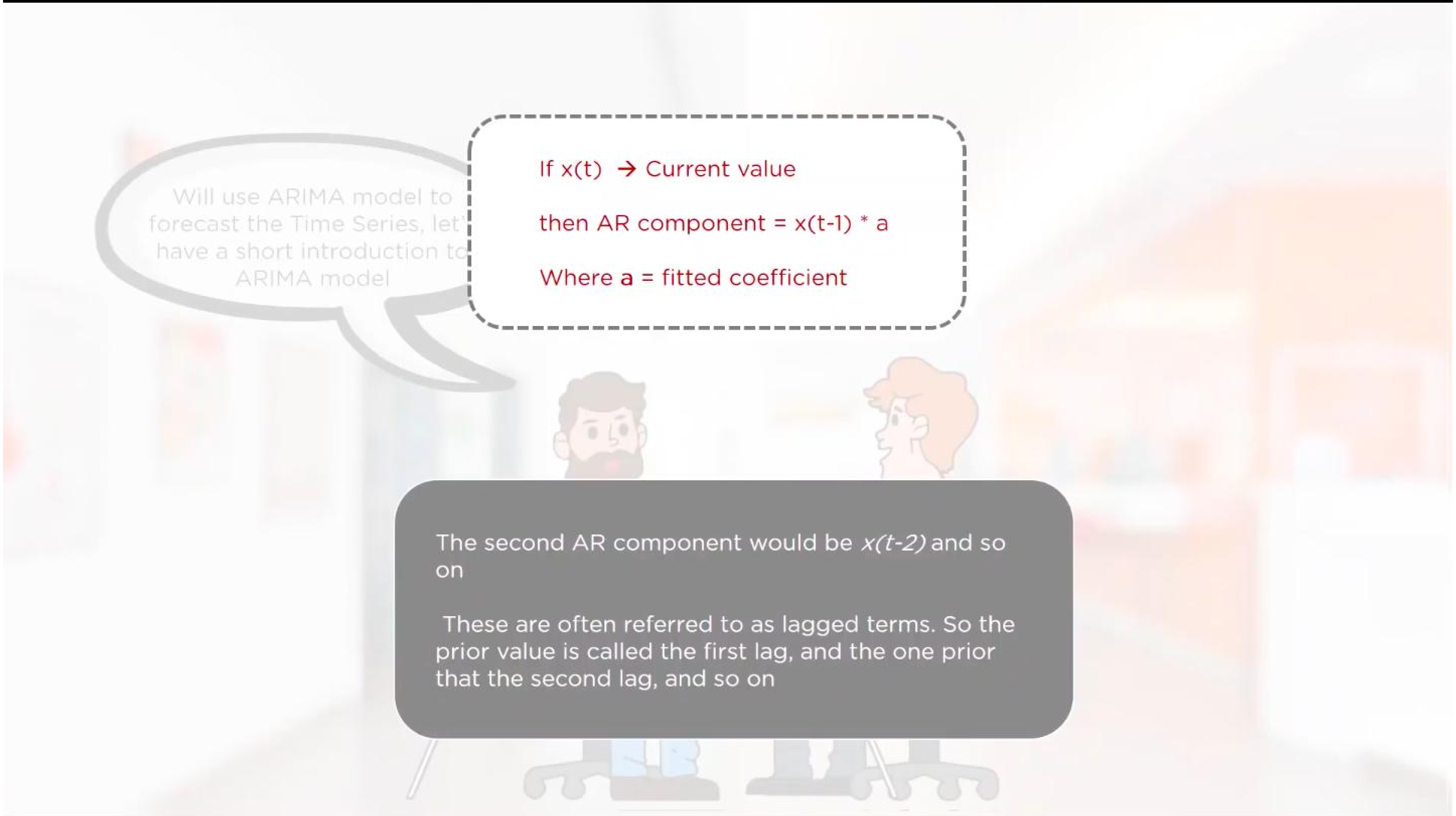
Auto-Regressive Parameter(AR), p	Degree of Differencing(I), d	Moving Average(MA), q

AR(p): number of autoregressive terms (AR)

Example: ARIMA(2,0,0) has a value of p as 2


But what are 'AR terms'?


In terms of a regression model, autoregressive components refer to prior values of the current value


Will use ARIMA model to forecast the Time Series, let's have a short introduction to ARIMA model
If x(t) → Current value
then AR component = x(t-1) * a
Where a = fitted coefficient
The second AR component would be *x(t-2)* and so on
These are often referred to as lagged terms. So the prior value is called the first lag, and the one prior that the second lag, and so on

Auto-Regressive Parameter(AR), p	Degree of Differencing(I), d	Moving Average(MA), q

It is equal to the number of non-seasonal differences needed to achieve stationarity

1 level of differencing would mean you take the current value and subtract the prior value from it

Differencing: Subtracting prior values from the current values:

Values	1st Order Differencing	Result
5	NA	NA
4	4-5	-1
6	6-4	2
7	7-6	1
9	9-7	2
12	12-9	3
12	12-12	0

If this series still shows a trend then you can do another level of differencing with the first level differenced series

Auto-Regressive Parameter(AR), p	Degree of Differencing(I), d	Moving Average(MA), q

It represents the error of the model as a combination of previous error terms e_t


But, ARIMA models work on the assumption of stationarity

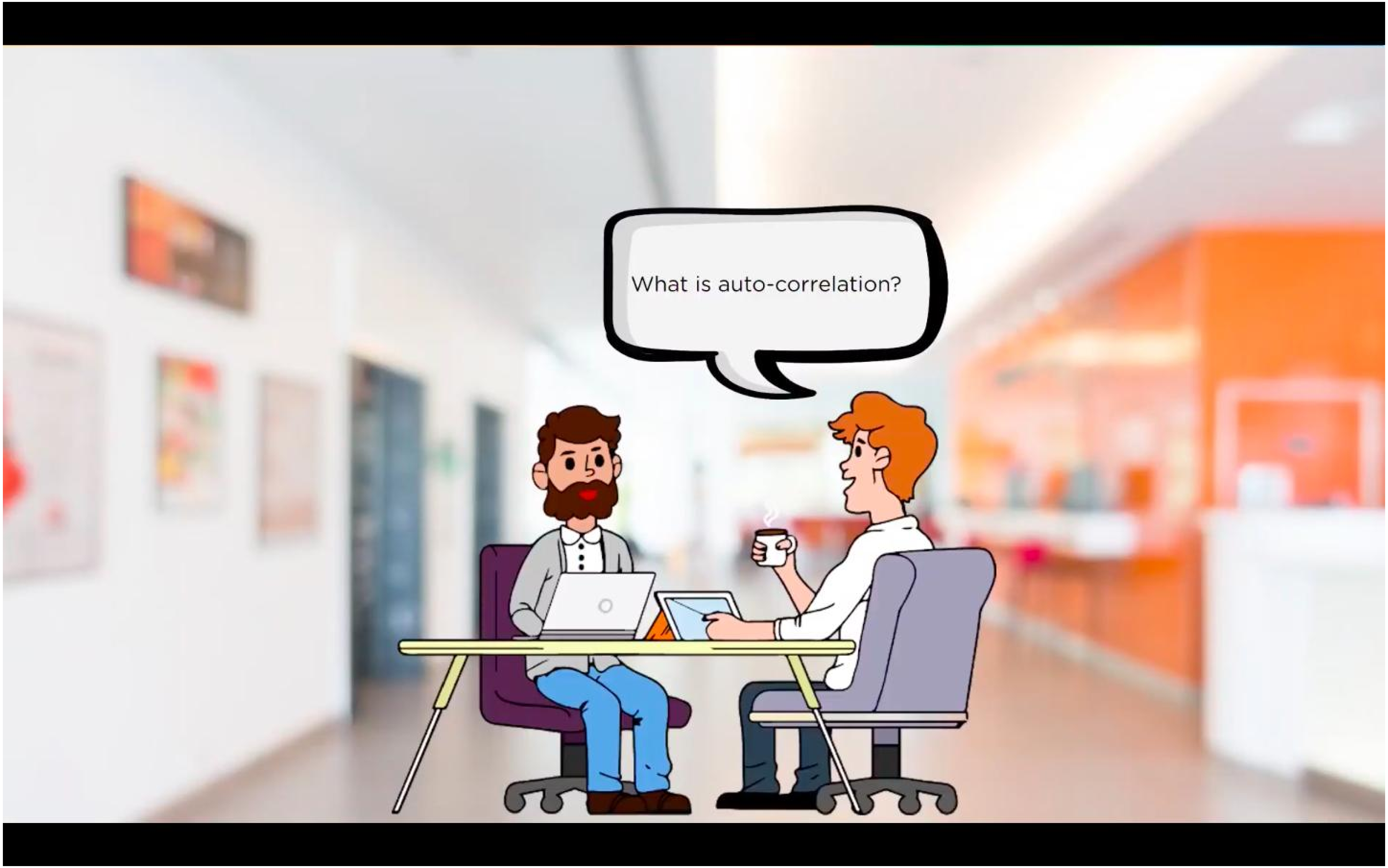

Oh you mean, removing trend and seasonality to make the data stationary, right?


Yes, like we saw in the previous part

ACF/PACF

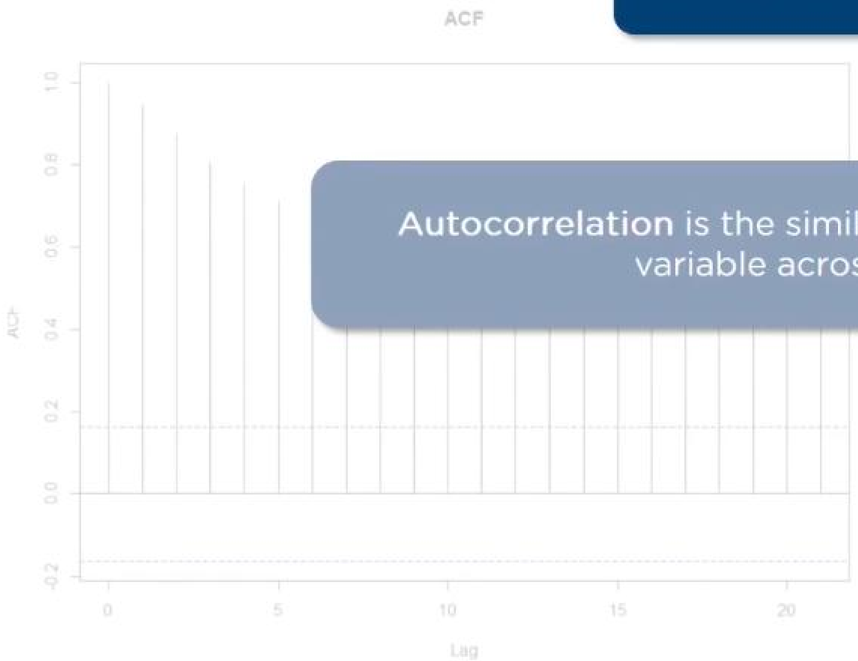
In order to test whether or not the series and their error term is auto correlated, we usually use:

- Auto-correlation function (ACF)
- Partial Auto-correlation function (PACF)


What is auto-correlation?

ACF

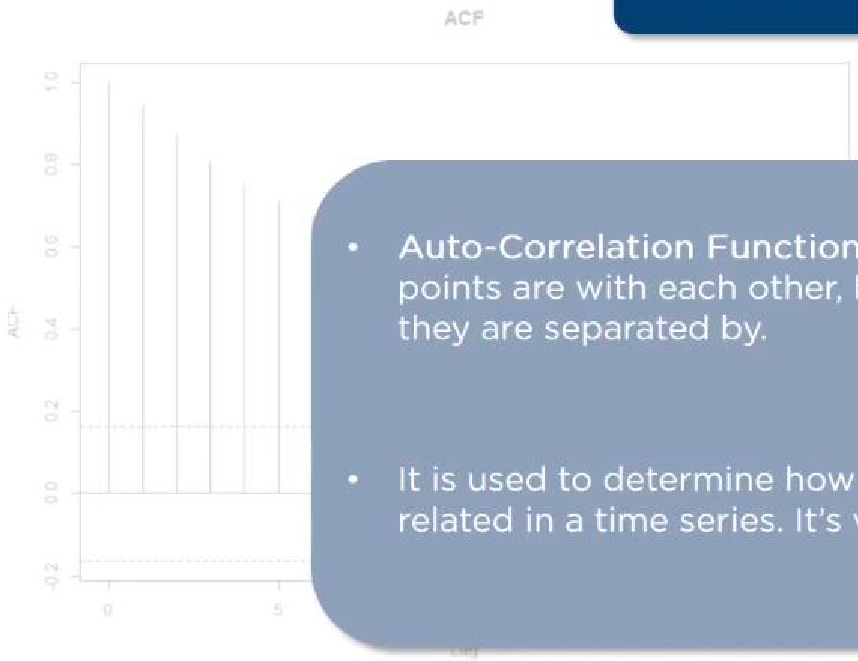
What is Auto-correlation?



Autocorrelation is the similarity between values of a same variable across observations

ACF

What is Auto-correlation?



- **Auto-Correlation Function(ACF)** tells you how correlated points are with each other, based on how many time steps they are separated by.

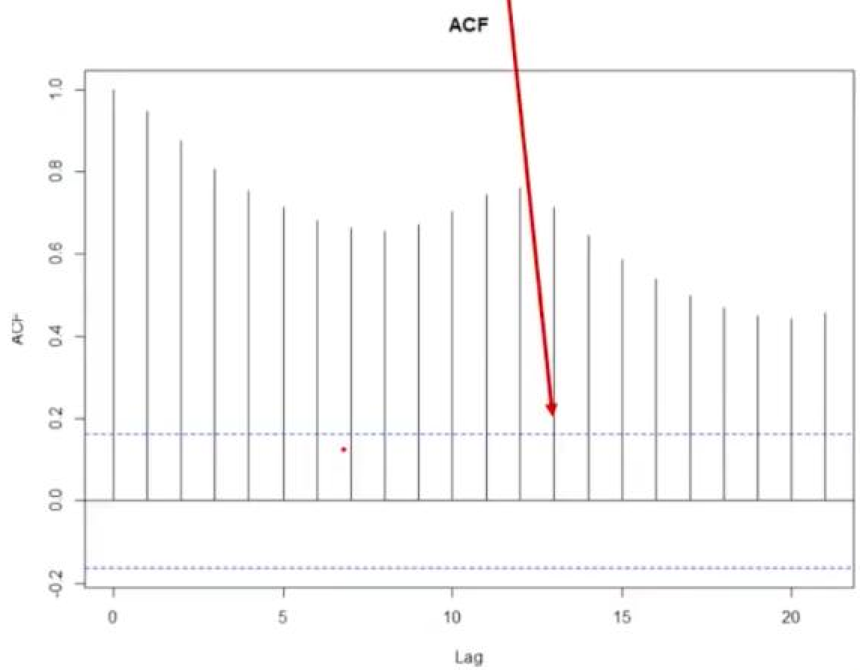
- It is used to determine how past and future data points are related in a time series. It's value can range from -1 to 1

ACF

When we plot ACF for our dataset, it crosses the blue dashed line which indicates that the values are correlated. Hence, non-stationary.

```
> acf(ts(AP),main='ACF')
```

R plots 95% significance boundaries as blue dotted lines



PACF

What is Partial Auto-correlation?

The PACF function shows a definite pattern, which does not repeat, we can conclude that the data does not show any seasonality

Partial autocorrelation is the degree of association between two variables while adjusting the effect of one or more additional variables

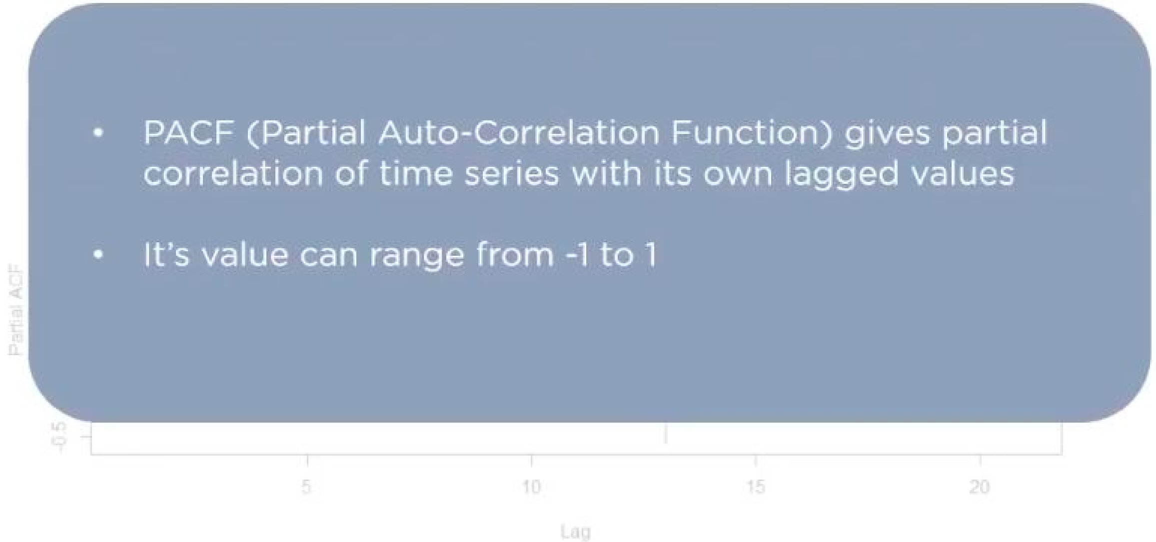
simplilearn

PACF

What is Partial Auto-correlation?

The PACF function shows a definite pattern, which does not repeat, we can conclude that the data does not show any seasonality

- PACF (Partial Auto-Correlation Function) gives partial correlation of time series with its own lagged values
- It's value can range from -1 to 1

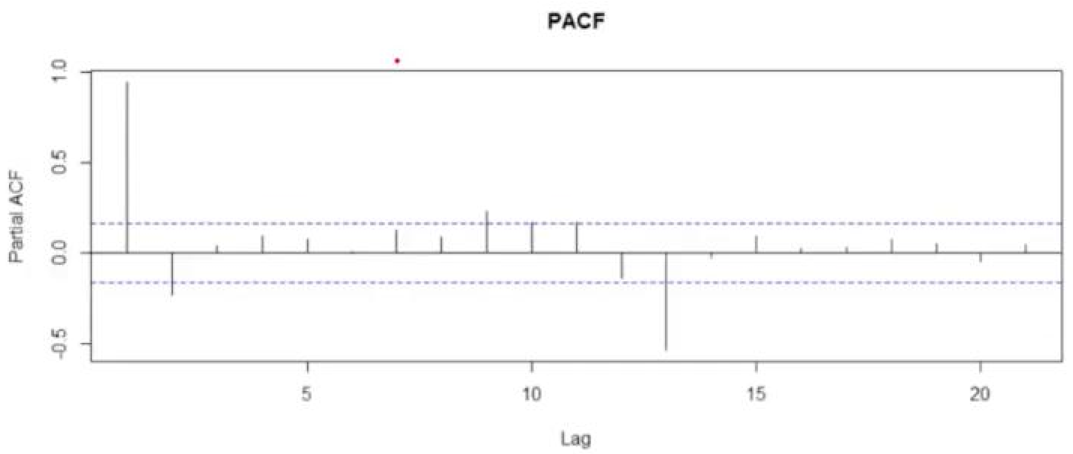


simplilearn

PACF

```
> pacf(ts(AP),main='PACF')
```

The PACF function shows a definite pattern, which does not repeat, we can conclude that the data does not show any seasonality



simplilearn

Use Case: Time Series Forecasting

Objective: To predict the airline tickets' sales of 1961 using Time Series Analysis

Data Description: 10 year air-ticket sales data of airline industry from 1949-1960

simplilearn

Use Case: Time Series Forecasting

Goal of Time Series

- **Time Series Behavior**: Identification of the Time Series components like trend , seasonality to describe the behavior
- **Time Series Forecasting**: Forecasting the values of the Time Series, depending on its actual and past values

simplilearn

Use Case: Exploratory Data Analysis

Load the data

Look at the data

```
> library(forecast)
>
> data(AirPassengers)
> class(AirPassengers)
[1] "ts"
```

It is a Time Series dataset

```
> AP
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
>
```

simplilearn

Use Case: Exploratory Data Analysis

Load the data

Look at the data

```
> library(forecast)
>
> data(AirPassengers)
> class(AirPassengers)
[1] "ts"
```

It is a Time Series dataset

```
> AirPassengers
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
>
```

Use Case: Exploratory Data Analysis

Starting point

```
> start(AirPassengers)
[1] 1949    1
>
> end(AirPassengers)
[1] 1960   12
>
> frequency(AirPassengers)
[1] 12
>
```

simplilearn

Use Case: Exploratory Data Analysis

End point

```
> start(AirPassengers)
[1] 1949    1
>
> end(AirPassengers)
[1] 1960   12
>
> frequency(AirPassengers)
[1] 12
>
```

Use Case: Exploratory Data Analysis

Frequency of the Data

```
> start(AirPassengers)
[1] 1949    1
>
> end(AirPassengers)
[1] 1960   12
>
> frequency(AirPassengers)
[1] 12
>
```

Use Case: Exploratory Data Analysis

And view the summary:

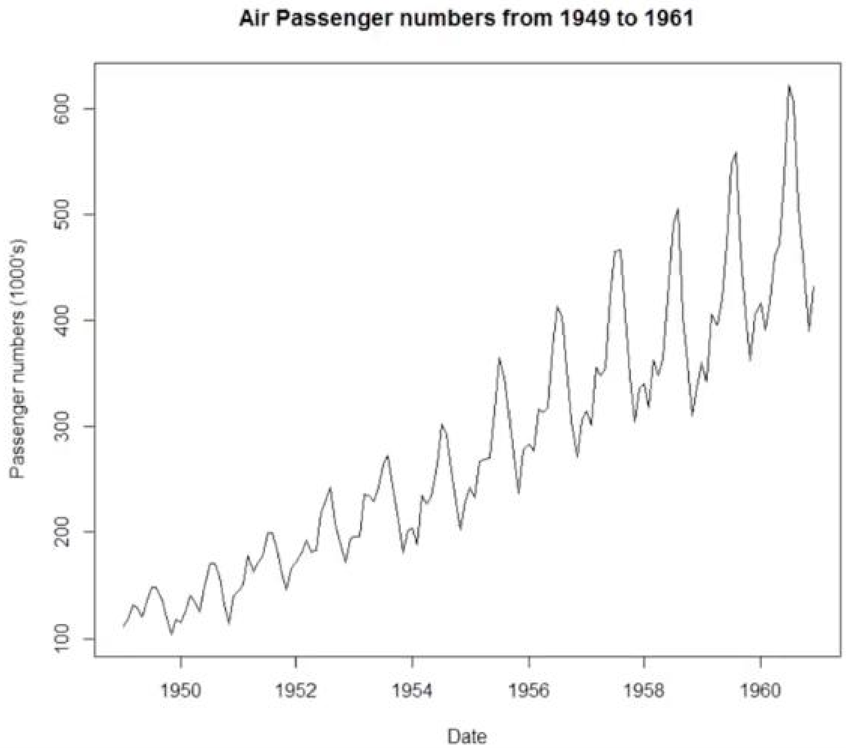
```
> sum(is.na(tsdata))
[1] 0
>
> summary(AirPassengers)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  104.0   180.0   265.5   280.3   360.5   622.0 
> 
```

simplilearn

Use Case: Exploratory Data Analysis

Let's plot the raw data using 'plot' function:

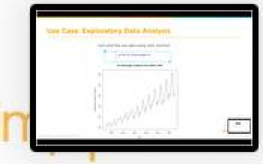
```
> plot(AirPassengers)
>
```



Use Case: Exploratory Data Analysis

Cycle of the data is:

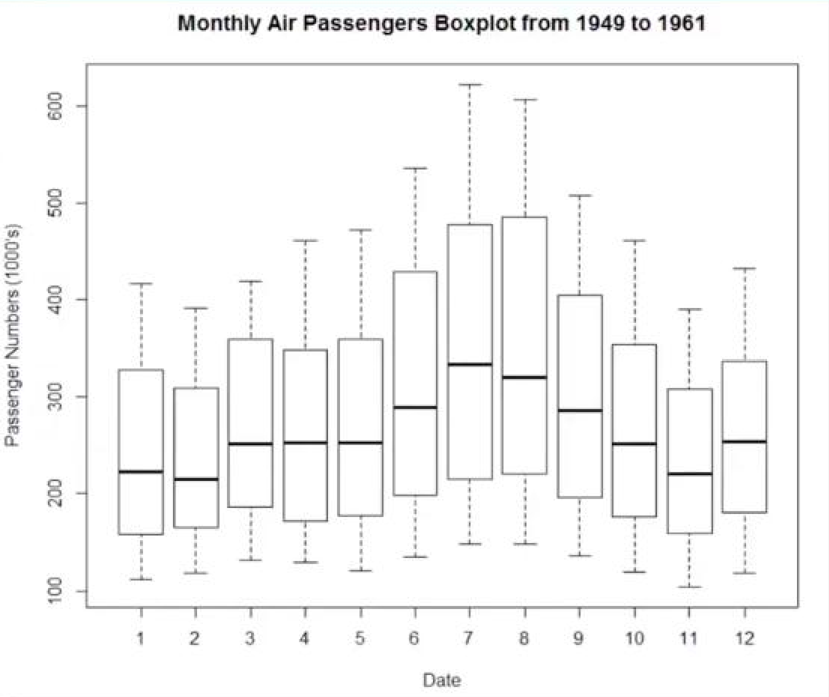
```
> cycle(AirPassengers)
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949   1   2   3   4   5   6   7   8   9  10  11  12
1950   1   2   3   4   5   6   7   8   9  10  11  12
1951   1   2   3   4   5   6   7   8   9  10  11  12
1952   1   2   3   4   5   6   7   8   9  10  11  12
1953   1   2   3   4   5   6   7   8   9  10  11  12
1954   1   2   3   4   5   6   7   8   9  10  11  12
1955   1   2   3   4   5   6   7   8   9  10  11  12
1956   1   2   3   4   5   6   7   8   9  10  11  12
1957   1   2   3   4   5   6   7   8   9  10  11  12
1958   1   2   3   4   5   6   7   8   9  10  11  12
1959   1   2   3   4   5   6   7   8   9  10  11  12
1960   1   2   3   4   5   6   7   8   9  10  11  12
> 
```



Use Case: Exploratory Data Analysis

Let's use the boxplot function to see any seasonal effects:

```
>
> #get a boxplot by cycle
> boxplot(AirPassengers~cycle(AirPassengers))
>
```



Use Case: Exploratory Data Analysis

Thus, we can make some initial inferences:

TREND

The passenger numbers increase over time indicating an increasing linear trend

SEASONALITY

In the boxplot there are more passengers travelling in months 6 to 9, indicating seasonality with a apparent cycle of 12 months

Use Case: Time Series Decomposition

We will decompose the Time Series

Decomposing means separating the original Time Series into its components(trend, seasonality, irregularity)

```
> tsdata<-ts(AirPassengers,frequency = 12)
>
> ddata<-decompose(tsdata, "multiplicative")
> |
```

Using decompose function in R

Applying multiplicative model


The data must have a constant variance and mean, right?


Let's fit the model!

Use Case: Fit A Time Series Model

```
> mymodel<-auto.arima(AirPassengers)
> mymodel
Series: AirPassengers
ARIMA(2,1,1)(0,1,0)[12]

Coefficients:
         ar1     ar2      ma1
      0.5960  0.2143  -0.9819
s.e.  0.0888  0.0880   0.0292

sigma^2 estimated as 132.3:  log likelihood=-504.92
AIC=1017.85   AICc=1018.17   BIC=1029.35
>
```

ARIMA Model

simplilearn

Use Case: Fit A Time Series Model

```
> auto.arima(AirPassengers,ic="aic" ,trace = TRUE)

 ARIMA(2,1,2)(1,1,1)[12]                    : Inf
 ARIMA(0,1,0)(0,1,0)[12]                    : 1031.508
 ARIMA(1,1,0)(1,1,0)[12]                    : 1020.393
 ARIMA(0,1,1)(0,1,1)[12]                    : 1021.003
 ARIMA(1,1,0)(0,1,0)[12]                    : 1020.394
 ARIMA(1,1,0)(2,1,0)[12]                    : 1019.24
 ARIMA(1,1,0)(2,1,1)[12]                    : Inf
 ARIMA(0,1,0)(2,1,0)[12]                    : 1032.12
 ARIMA(2,1,0)(2,1,0)[12]                    : 1021.12
 ARIMA(1,1,1)(2,1,0)[12]                    : 1021.033
 ARIMA(2,1,1)(2,1,0)[12]                    : 1017.658
 ARIMA(2,1,1)(1,1,0)[12]                    : 1017.915
 ARIMA(2,1,1)(2,1,1)[12]                    : Inf
 ARIMA(3,1,1)(2,1,0)[12]                    : 1019.459
 ARIMA(2,1,2)(2,1,0)[12]                    : 1019.504
 ARIMA(3,1,2)(2,1,0)[12]                    : 1021.454

 Best model: ARIMA(2,1,1)(2,1,0)[12]

Series: AirPassengers
ARIMA(2,1,1)(2,1,0)[12]

Coefficients:
         ar1     ar2      ma1     sar1    sar2
      0.5741  0.2547  -0.9822  -0.1136  0.1641
s.e.  0.0899  0.0933   0.0286   0.0963  0.1073

sigma^2 estimated as 129.5:  log likelihood=-502.83
AIC=1017.66   AICc=1018.34   BIC=1034.91
```

ARIMA Model

simplilearn

FIT A TIME SERIES MODEL

The ARIMA fitted model is:

Y^=0.5960Yt−2+0.2143Yt−12−0.9819et−1+EY^

=0.5960Yt−2+0.2143Yt−12−0.9819et−1+E

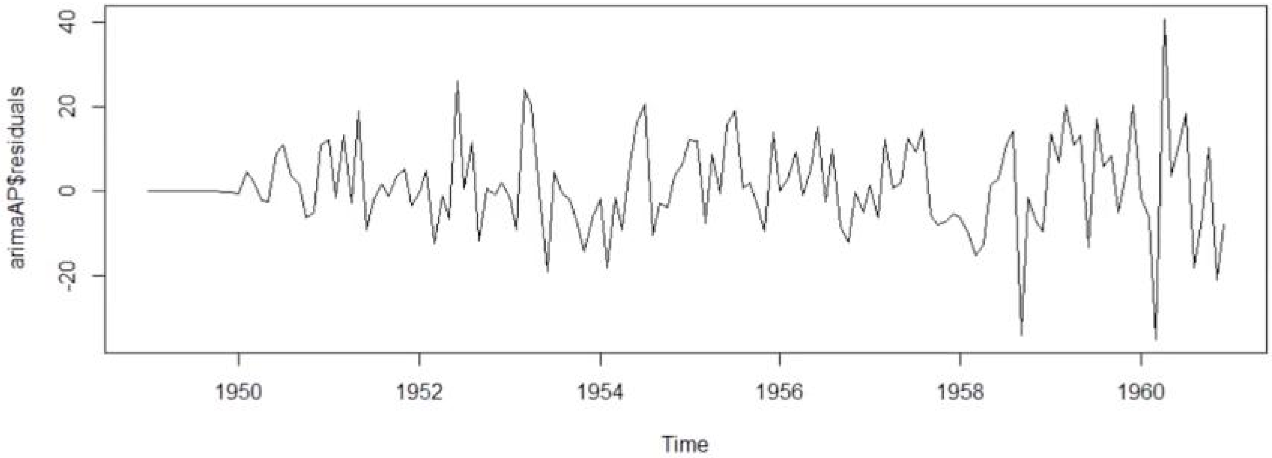
where E is error

ARIMA Model

Use Case: Diagnostics

Check the plot for the residuals which shows Stationarity:

```
> plot.ts(mymodel$residuals)
>
```



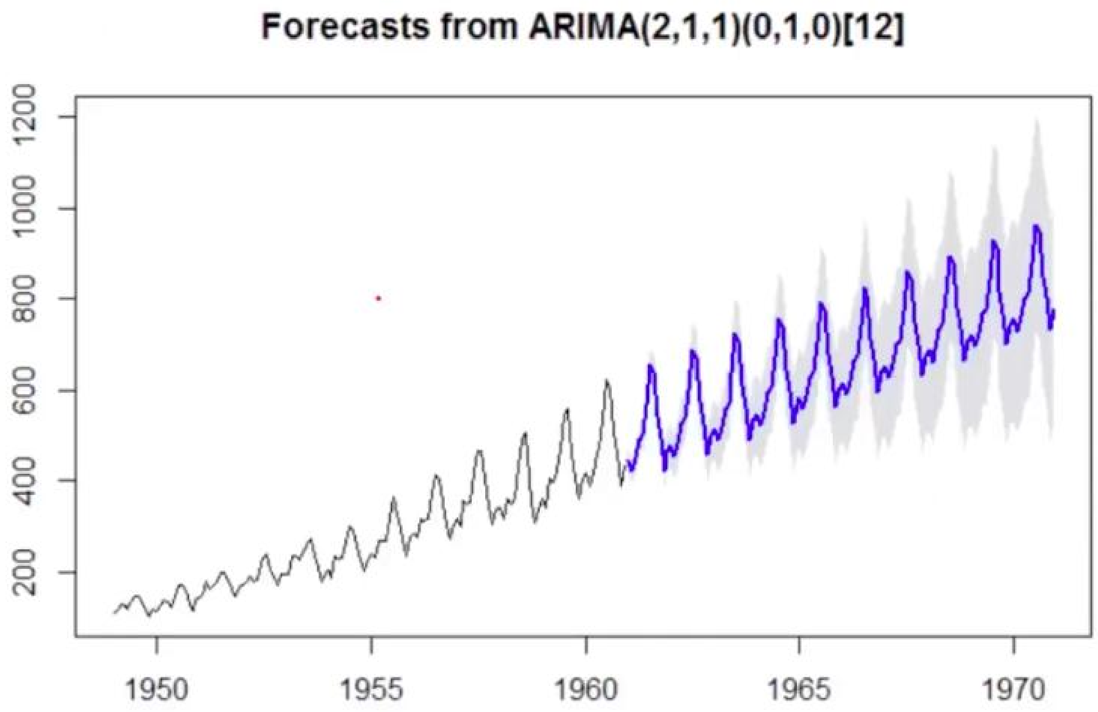
Use Case: Calculate Forecast

You can plot a forecast of the Time Series using the forecast function, with a 95% confidence interval where h is the forecast horizon periods in months

```
> myforecast<-forecast(mymodel,level=c(95),h=10*12)
> plot(myforecast)
>
```

simplilearn

Use Case: Calculate Forecast




Can we validate this model?

Validation

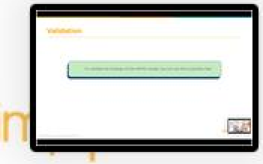
To validate the findings of the ARIMA model, we can use the Ljung-Box test

Use Case: Validation

```
> Box.test(mymodel$resid, lag=5, type="Ljung-Box")

        Box-Ljung test

data:  mymodel$resid
X-squared = 2.9244, df = 5, p-value = 0.7116

>   Box.test(mymodel$resid, lag=10, type="Ljung-Box")

        Box-Ljung test

data:  mymodel$resid
X-squared = 8.6878, df = 10, p-value = 0.562

>     Box.test(mymodel$resid, lag=15, type="Ljung-Box")

        Box-Ljung test

data:  mymodel$resid
X-squared = 11.582, df = 15, p-value = 0.7104
```

Conclusion

We can arbitrarily select the lag value. In this case, we take the lag values as 5, 10, 15




Right, it indicates that our model is free of auto-correlation

Use Case: Validation

Conclusion

We can conclude from the ARIMA output, that our model using parameters (2, 1, 1) has been shown to adequately fit the data