Big Data and Analytics

Notemybook Services Pvt Ltd

Chapter 1

Learn Analysis Using MS Excel

Microsoft Excel is one of the most used software applications of all time. Hundreds of millions of people around the world use Microsoft Excel. You can use Excel to enter all sorts of data and perform financial, mathematical or statistical calculations.

Basics

1 **Ribbon:** Excel selects the ribbon's Home tab when you open it. Learn how to minimize and customize the ribbon.

2 **Workbook:** A workbook is another word for your Excel file. Excel automatically creates a blank workbook when you open it.

3 **Worksheets**: A worksheet is a collection of cells where you keep and manipulate the data. By default, each Excel workbook contains three worksheets.

4 **Format Cells:** When we format cells in Excel, we change the appearance of a number without changing the number itself.

5 **Find & Select:** Learn how to use Excel's Find, Replace and Go To Special feature.

6 **Templates:** Instead of creating an Excel workbook from scratch, you can create a workbook based on a template. There are many free templates available, waiting to be used.

7 **Data Validation:** Use data validation in Excel to make sure that users enter certain values into a cell.

8 **Keyboard Shortcuts:** Keyboard shortcuts allow you to do things with your keyboard instead of your mouse to increase your speed.

9 **Print:** This chapter teaches you how to print a worksheet and how to change some important print settings in Excel.

10 **Share:** Learn how to share Excel data with Word documents and other files.

11 **Protect:** Encrypt an Excel file with a password so that it requires a password to open it.

Functions

1 Count and Sum: The most used functions in Excel are the functions that count and sum. You can count and sum based on one criteria or multiple criteria.

2 Logical: Learn how to use Excel's logical functions such as the IF, AND and OR function.

3 Cell References: Cell references in Excel are very important. Understand the difference between relative, absolute and mixed reference, and you are on your way to success.

4 Date & Time: To enter a date in Excel, use the "/" or "-" characters. To enter a time, use the ":" (colon). You can also enter a date and a time in one cell.

5 Text: Excel has many functions to offer when it comes to manipulating text strings.

Functions

6 Lookup & Reference: Learn all about Excel's lookup & reference functions such as the VLOOKUP, HLOOKUP, MATCH, INDEX and CHOOSE function.

7 Financial: This chapter illustrates Excel's most popular financial functions.

8 Statistical: An overview of some very useful statistical functions in Excel.

9 Round: This chapter illustrates three functions to round numbers in Excel. The ROUND, ROUNDUP and ROUNDDOWN function.

10 Formula Errors: This chapter teaches you how to deal with some common formula errors in Excel.

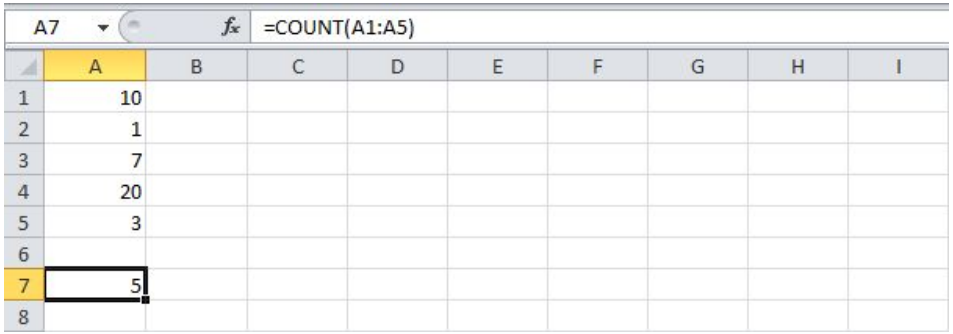
11 Array Formulas: This chapter helps you understand array formulas in Excel. Single cell array formulas perform multiple calculations in one cell.

1. Count and Sum Functions

The most used functions in Excel are the functions that count and sum. You can count and sum based on one criteria or multiple criteria.

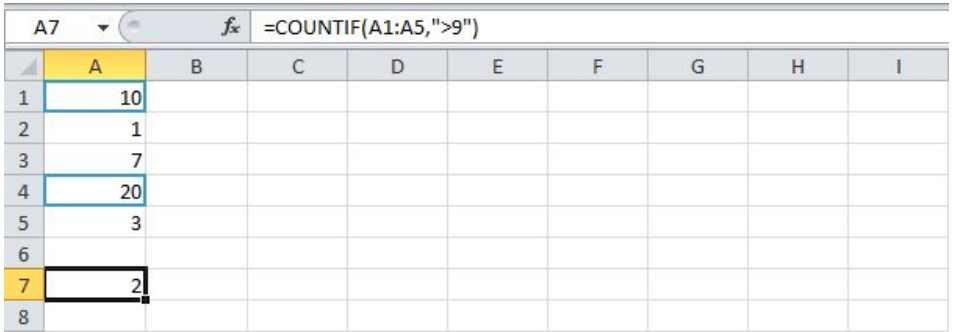
Count

To count the number of cells that contain numbers, use the COUNT function.



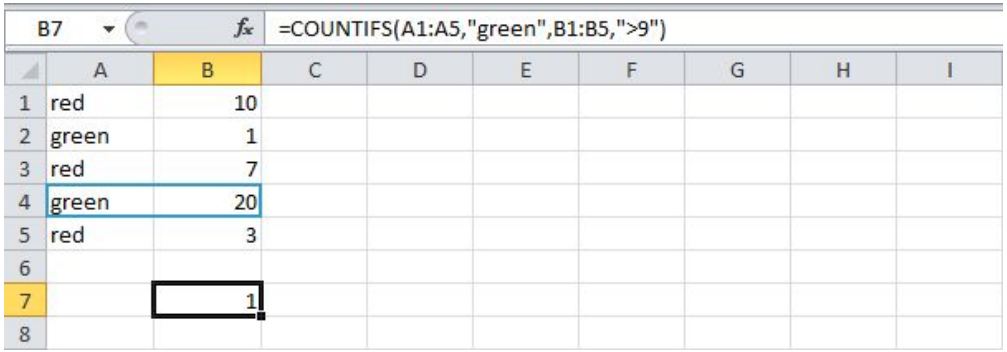
Countif

To count cells based on one criteria (for example, higher than 9), use the following COUNTIF function.



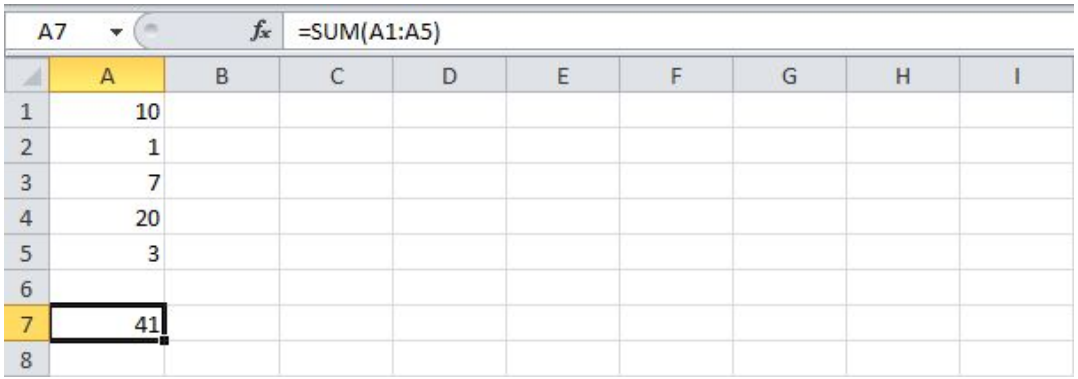
Countifs

To count cells based on multiple criteria (for example, green and higher than 9), use the following COUNTIFS function.



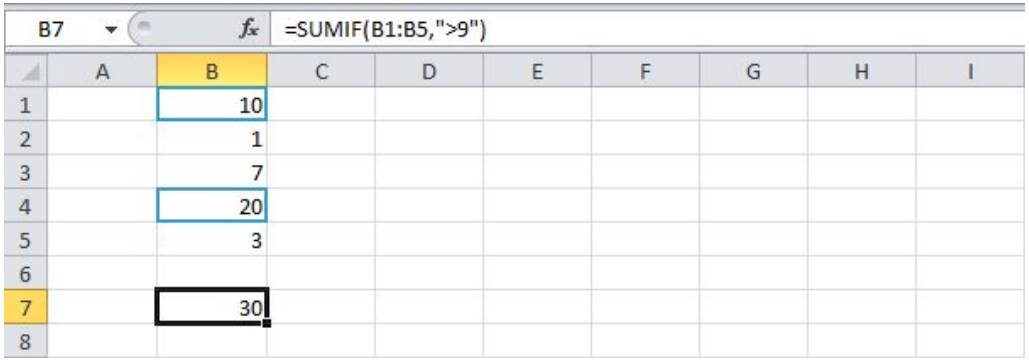
Sum

To sum a range of cells, use the SUM function.

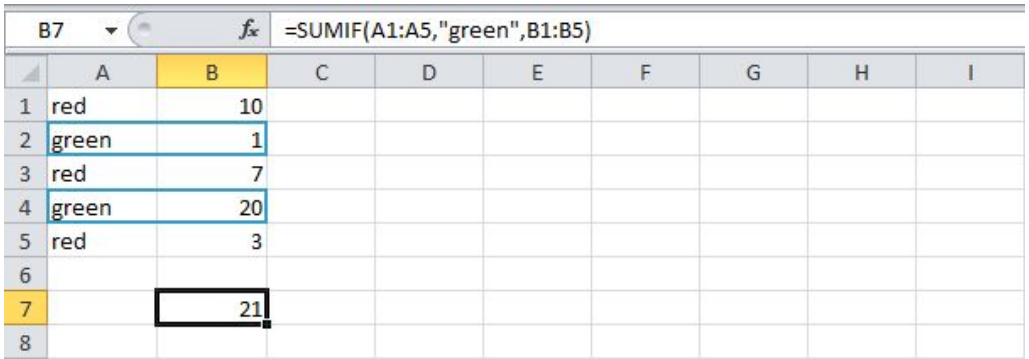


Sumif

To sum cells based on one criteria (for example, higher than 9), use the following SUMIF function (two arguments).

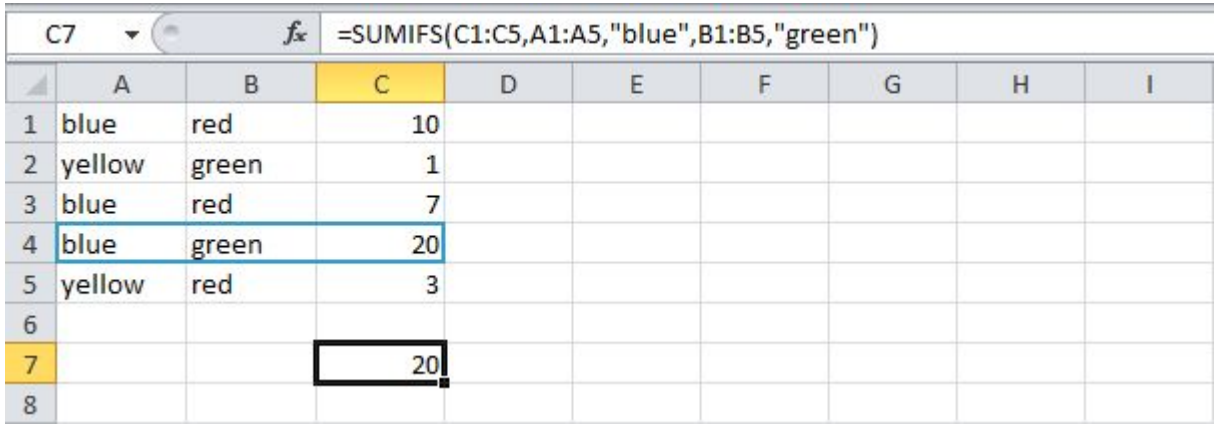


To sum cells based on one criteria (for example, green), use the following SUMIF function (three arguments, last argument is the range to sum).



Sumifs

To sum cells based on multiple criteria (for example, blue and green), use the following SUMIFS function (first argument is the range to sum).



	A	B	C	D	E	F	G	H	I
1	blue	red	10						
2	yellow	green	1						
3	blue	red	7						
4	blue	green	20						
5	yellow	red	3						
6									
7			20						
8									

General note: in a similar way, you can use the AVERAGEIF and AVERAGEIFS function to average cells based on one or multiple criteria.

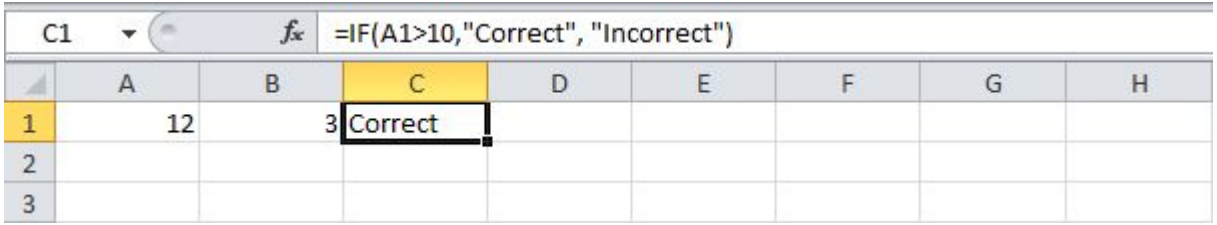
2. Logical Functions

If Function

The IF function checks whether a condition is met, and returns one value if TRUE and another value if FALSE.

1. Select cell C2 and enter the following function.

The IF function returns Correct because the value in cell A1 is higher than 10.

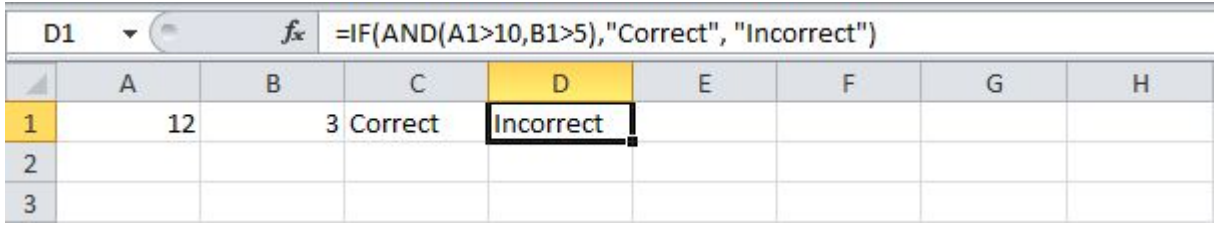


And Function

The AND Function returns TRUE if all conditions are true and returns FALSE if any of the conditions are false.

1. Select cell D2 and enter the following formula.

The AND function returns FALSE because the value in cell B2 is not higher than 5. As a result the IF function returns Incorrect.



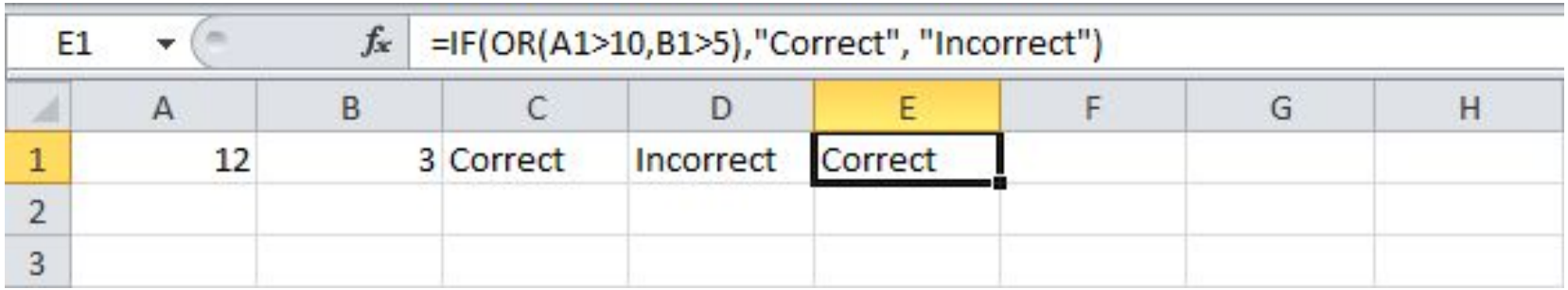
Or Function

The OR function returns TRUE if any of the conditions are TRUE and returns FALSE if all conditions are false.

1. Select cell E2 and enter the following formula.

The OR function returns TRUE because the value in cell A1 is higher than 10. As a result the IF function returns Correct.

General note: the AND and OR function can check up to 255 conditions.

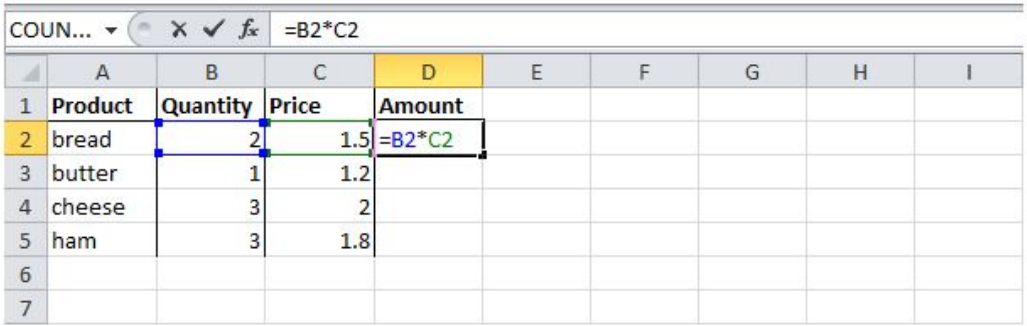


3. Cell References

Cell references in Excel are very important. Understand the difference between relative, absolute and mixed reference, and you are on your way to success.

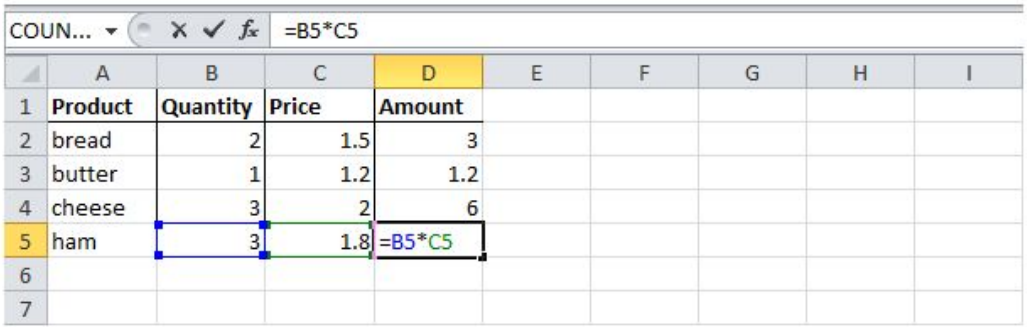
Relative Reference

By default, Excel uses relative reference. See the formula in cell D2 below. Cell D2 references (points to) cell B2 and cell C2. Both references are relative.



	A	B	C	D	E	F	G	H	I
1	**Product**	**Quantity**	**Price**	**Amount**					
2	bread	2	1.5	=B2*C2					
3	butter	1	1.2						
4	cheese	3	2						
5	ham	3	1.8						
6									
7									

1. Select cell D2, click on the lower right corner of cell D2 and drag it down to cell D5.



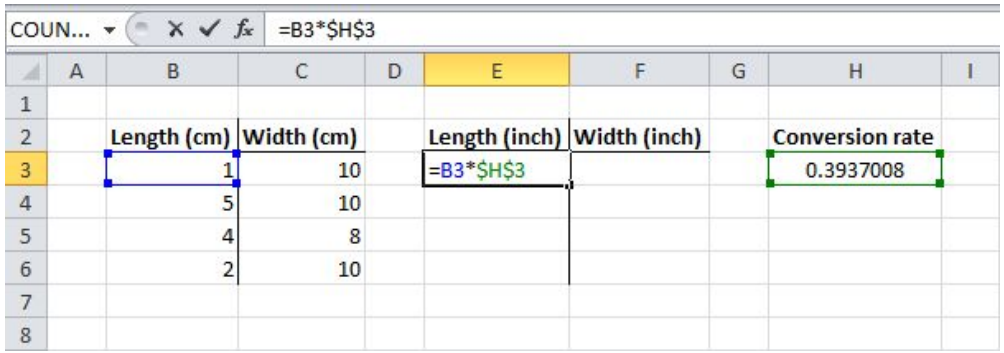
	A	B	C	D	E	F	G	H	I
1	**Product**	**Quantity**	**Price**	**Amount**					
2	bread	2	1.5	3					
3	butter	1	1.2	1.2					
4	cheese	3	2	6					
5	ham	3	1.8	=B5*C5					
6									
7									

Cell D3 references cell B3 and cell C3. Cell D4 references cell B4 and cell C4. Cell D5 references cell B5 and cell C5. In other words: each cell references its two neighbors on the left.

Absolute Reference

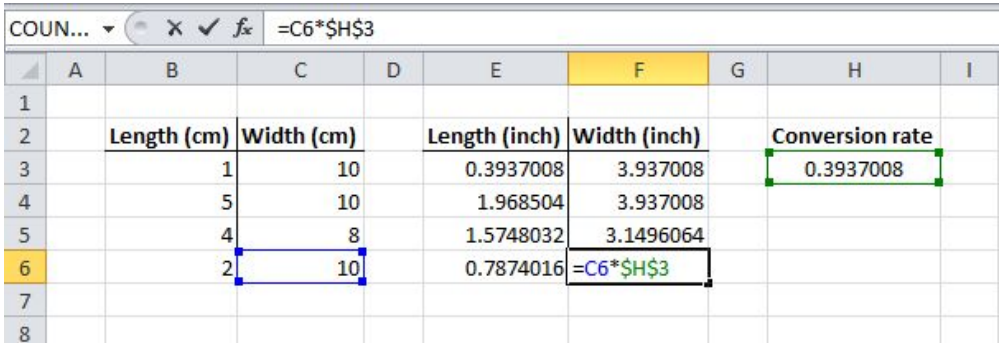
See the formula in cell E3 below.

1. To create an absolute reference to cell H3, place a $ symbol in front of the column letter and row number of cell H3 ($H$3) in the formula of cell E3.



	A	B	C	D	E	F	G	H	I
1									
2		Length (cm)	Width (cm)		Length (inch)	Width (inch)		Conversion rate	
3		1	10		=B3*$H$3			0.3937008	
4		5	10						
5		4	8						
6		2	10						
7									
8									

2. Now we can quickly drag this formula to the other cells.



	A	B	C	D	E	F	G	H	I
1									
2		Length (cm)	Width (cm)		Length (inch)	Width (inch)		Conversion rate	
3		1	10		0.3937008	3.937008		0.3937008	
4		5	10		1.968504	3.937008			
5		4	8		1.5748032	3.1496064			
6		2	10		0.7874016	=C6*$H$3			
7									
8									

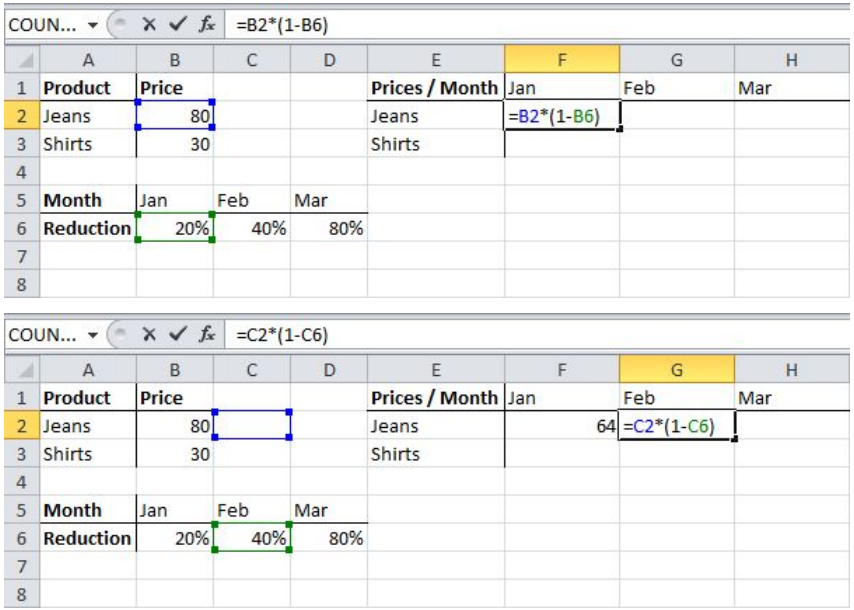
The reference to cell H3 is fixed (when we drag the formula down and across). As a result, the correct lengths and widths in inches are calculated.

Mixed Reference

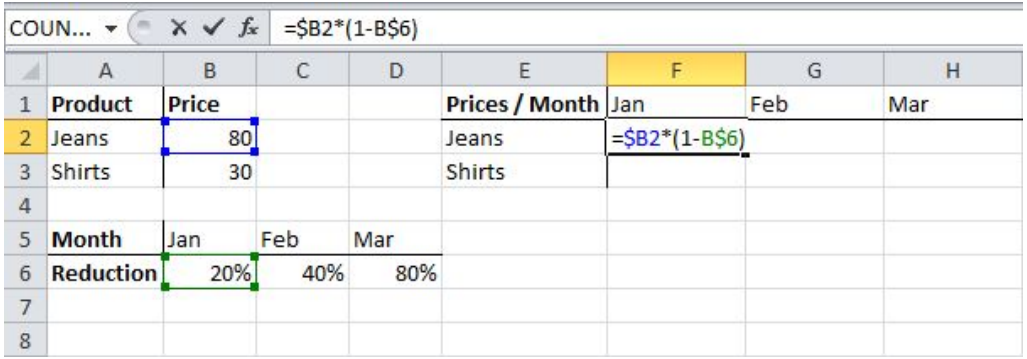
Sometimes we need a combination of relative and absolute reference (mixed reference).

1. See the formula in cell F2 below.

2. We want to copy this formula to the other cells quickly. Drag cell F2 across one cell, and look at the formula in cell G2.



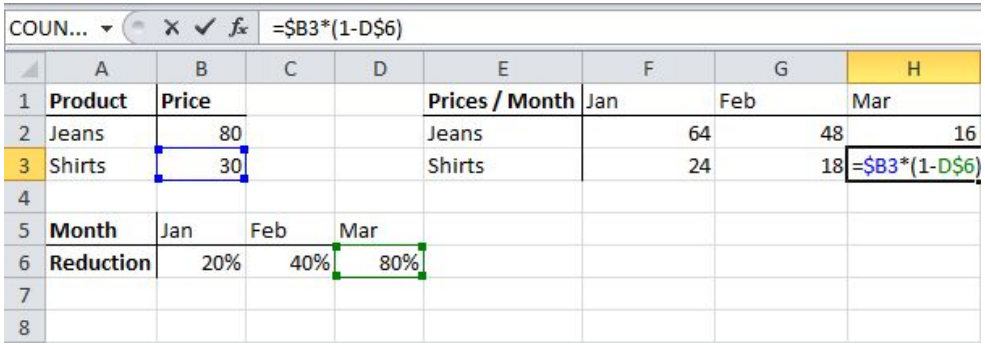
Do you see what happens? The reference to the price should be a fixed reference to column B. Solution: place a $ symbol in front of the column letter of cell B2 ($B2) in the formula of cell F2. In a similar way, when we drag cell F2 down, the reference to the reduction should be a fixed reference to row 6. Solution: place a $ symbol in front of the row number of cell B6 (B$6) in the formula of cell F2.



Result:

Note: we don't place a $ symbol in front of the row number of B2 (this way we allow the reference to change from B2 (Jeans) to B3 (Shirts) when we drag the formula down). In a similar way, we don't place a $ symbol in front of the column letter of B6 (this way we allow the reference to change from B6 (Jan) to C6 (Feb) and D6 (Mar) when we drag the formula across).

3. Now we can quickly drag this formula to the other cells.

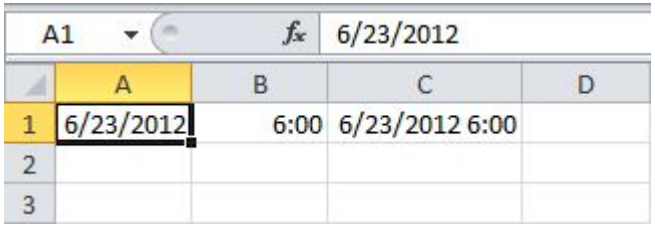


The references to column B and row 6 are fixed.

4. Date and Time Function

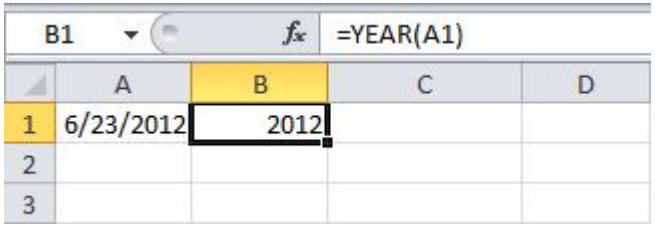
To enter a date in Excel, use the "/" or "-" characters. To enter a time, use the ":" (colon). You can also enter a date and a time in one cell.

Note: Dates are in US Format. Months first, Days second. This type of format depends on your windows regional settings.



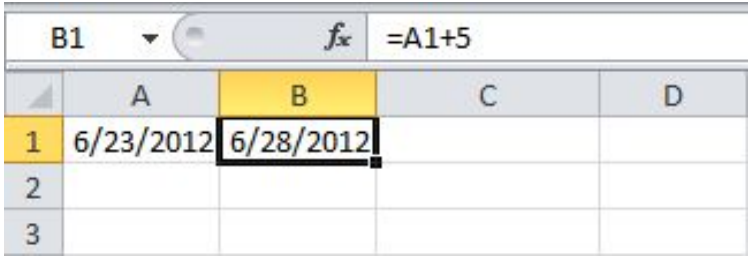
Year, Month, Day

To get the year of a date, use the YEAR function.

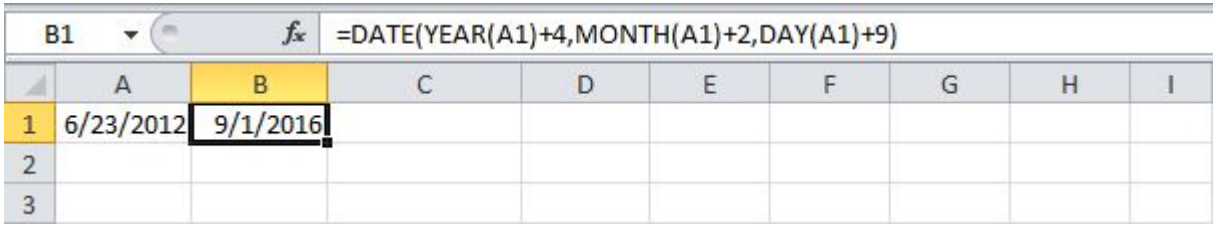


Date Function

1. To add a number of days to a date, use the following simple formula.



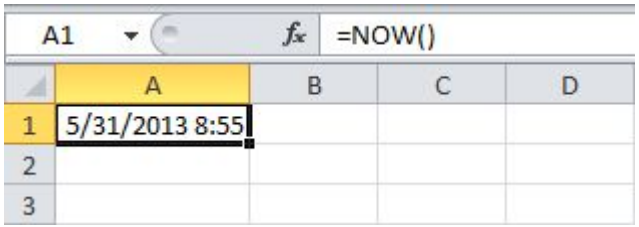
2. To add a number of years, months and/or days, use the DATE function.



Note: the DATE function accepts three arguments: year, month and day. Excel knows that 6 + 2 = 8 = August has 31 days and rolls over to the next month (23 August + 9 days = 1 September).

Current Date & Time

To get the current date and time, use the NOW function.

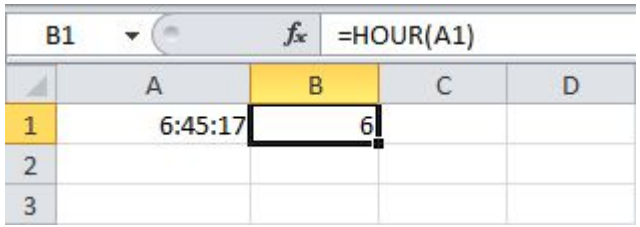


Note: use the TODAY function to get the current date only. Use NOW()-TODAY() to get the current time only (and apply a Time format).

Hour, Min, Sec

To return the hour, use the HOUR function.

Note: use the MINUTE and SECOND function to return the minute and second.

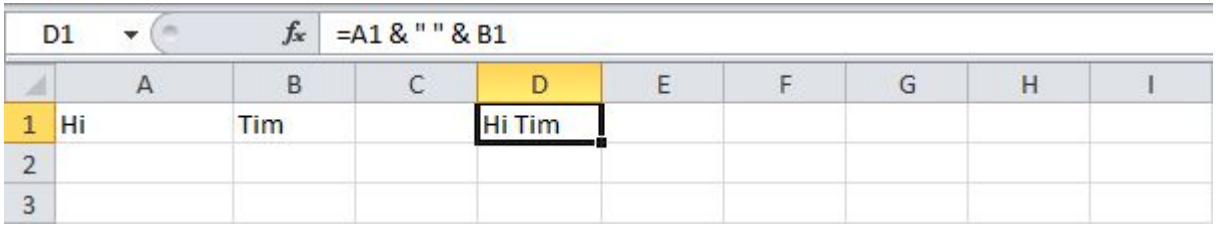


4. Text Functions

Excel has many functions to offer when it comes to manipulating text strings.

Join Strings

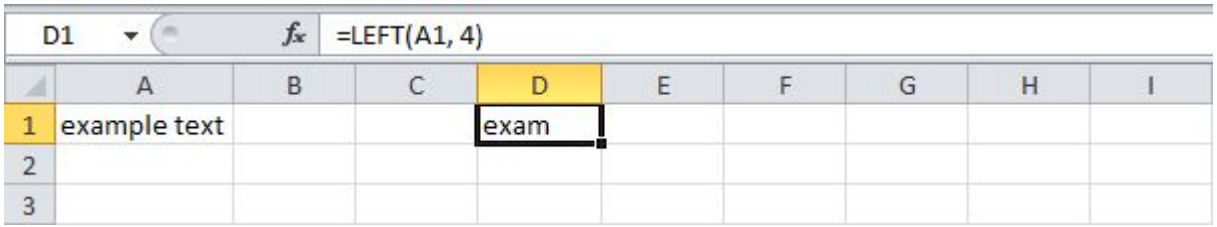
To join strings, use the & operator.



Note: to insert a space, use " "

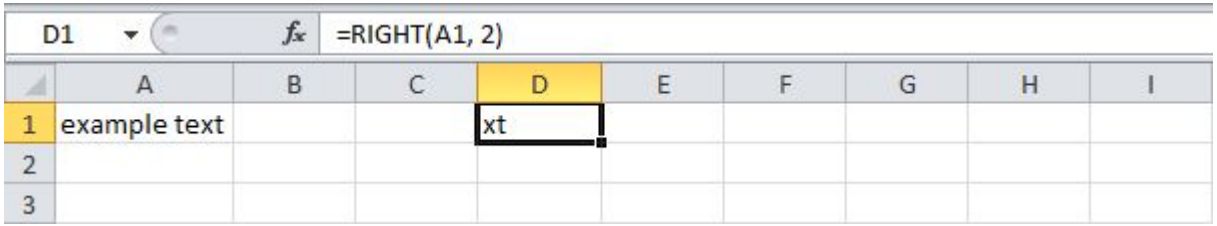
Left

To extract the leftmost characters from a string, use the LEFT function.



Right

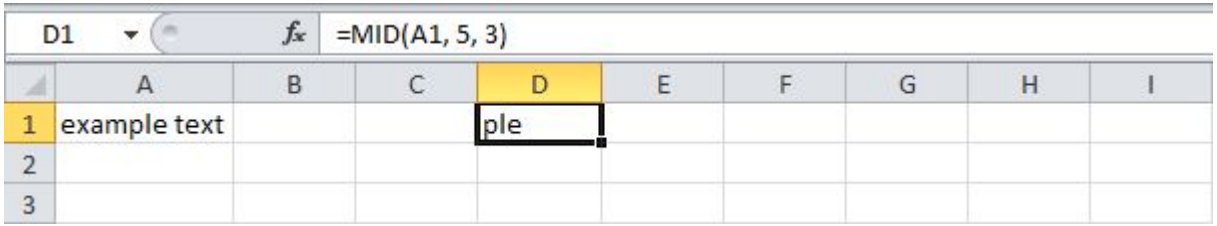
To extract the rightmost characters from a string, use the RIGHT function.



Mid

To extract a substring, starting in the middle of a string, use the MID function.

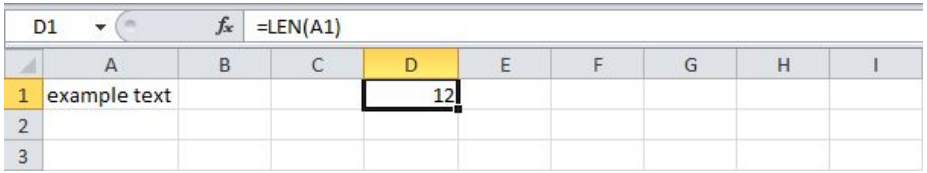
Note: started at position 5 (p) with length 3.



Len

To get the length of a string, use the LEN function.

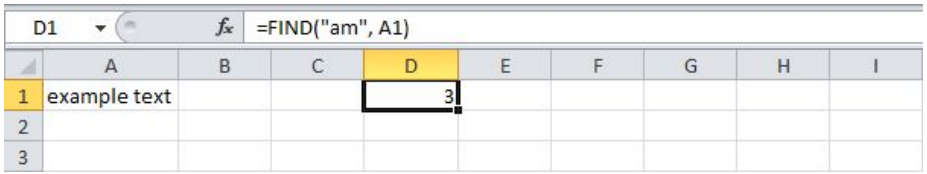
Note: space (position 8) included!



Find

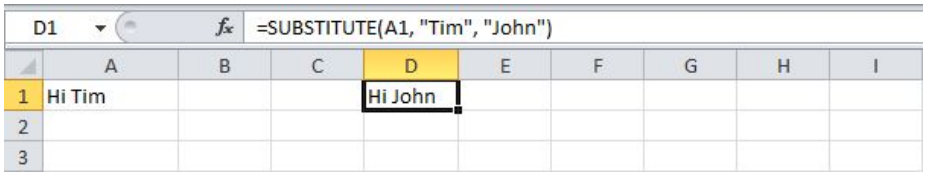
To find the position of a substring in a string, use the FIND function.

Note: string "am" found at position 3.



Substitute

To replace existing text with new text in a string, use the SUBSTITUTE function.



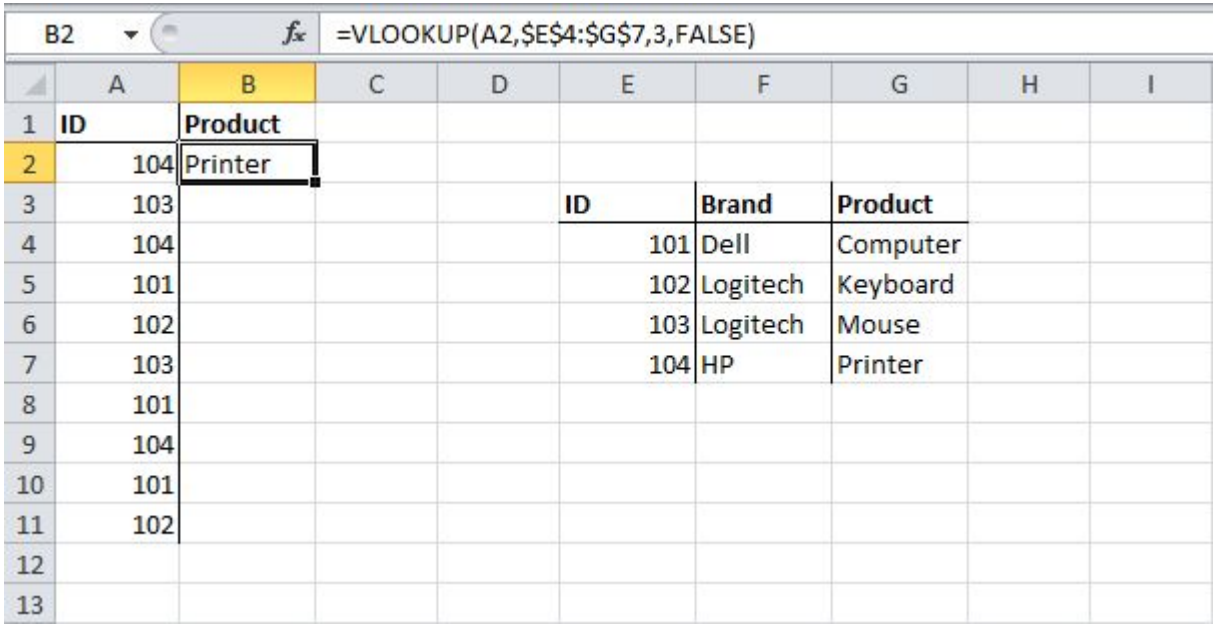
5. LookUp and Reference Functions

VLookup

The VLOOKUP (Vertical lookup) function looks for a value in the leftmost column of a table, and then returns a value in the same row from another column you specify.

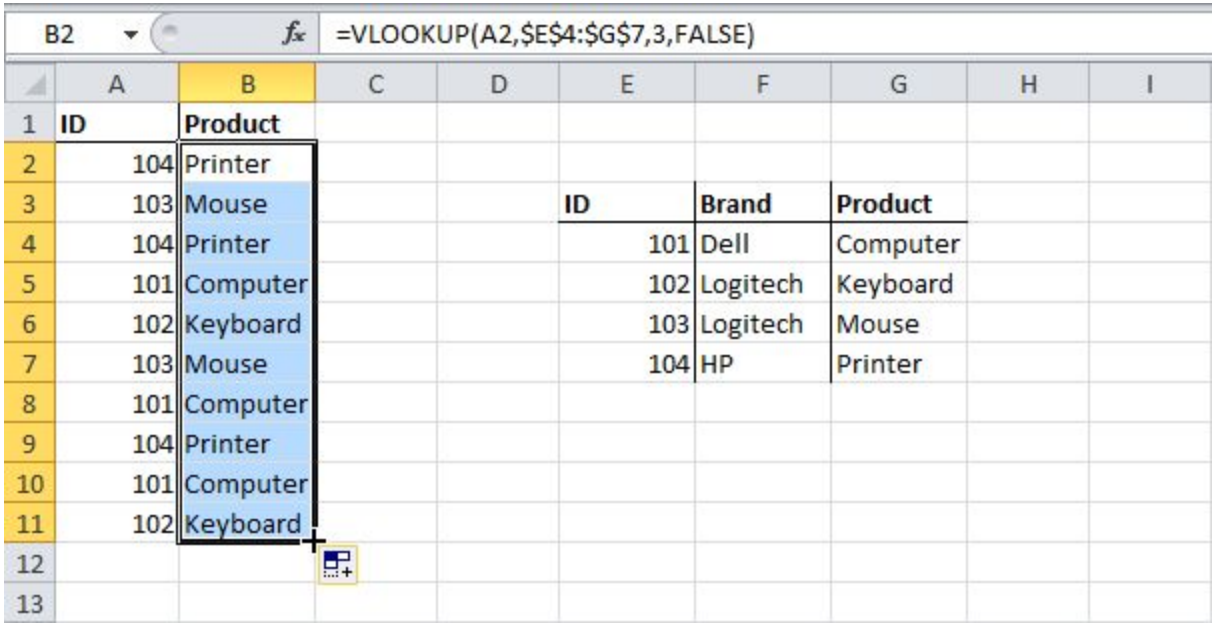
1. Insert the VLOOKUP function shown below.

Explanation: the VLOOKUP function looks for the ID (104) in the leftmost column of the range $E$4:$G$7 and returns the value in the same row from the third column (third argument is set to 3). The fourth argument is set to FALSE to return an exact match or a #N/A error if not found.



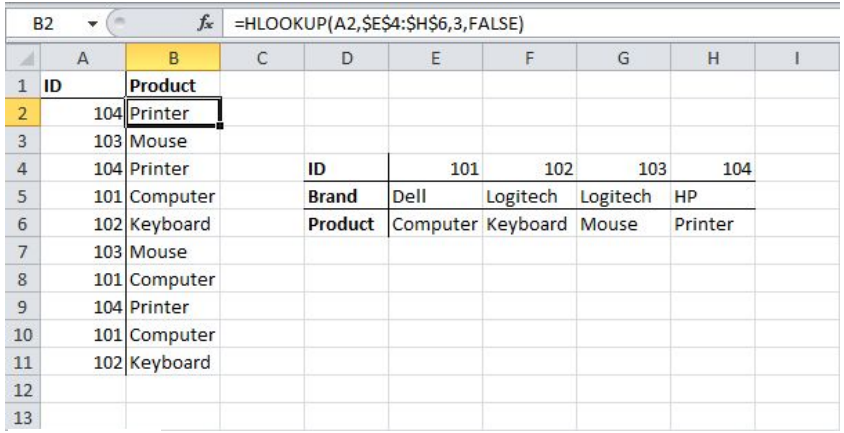
2. Drag the VLOOKUP function in cell B2 down to cell B11.

Note: when we drag the VLOOKUP function down, the absolute reference ($E$4:$G$7) stays the same, while the relative reference (A2) changes to A3, A4, A5, etc.



HLookup

In a similar way, you can use the HLOOKUP (Horizontal lookup) function.



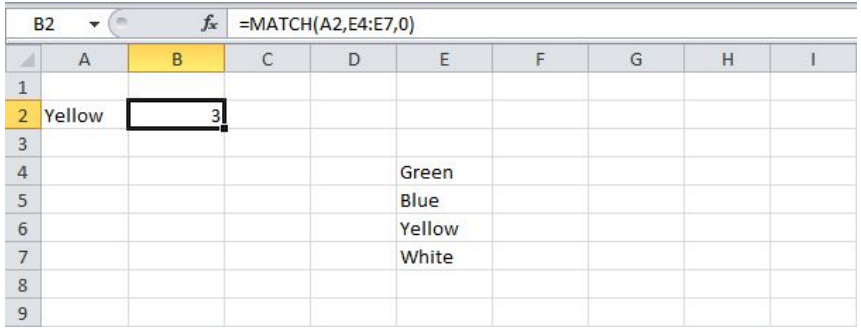
B2 =HLOOKUP(A2,$E$4:$H$6,3,FALSE)

	A	B	C	D	E	F	G	H	I
1	ID	Product							
2	104	Printer							
3	103	Mouse							
4	104	Printer		ID	101	102	103	104	
5	101	Computer		Brand	Dell	Logitech	Logitech	HP	
6	102	Keyboard		Product	Computer	Keyboard	Mouse	Printer	
7	103	Mouse							
8	101	Computer							
9	104	Printer							
10	101	Computer							
11	102	Keyboard							
12									
13									

Match

The MATCH function returns the position of a value in a given range.

Note: Yellow found at position 3 in the range E4:E7. The third argument is optional. Set this argument to 0 to return the position of the value that is exactly equal to lookup_value (A2) or a #N/A error if not found.



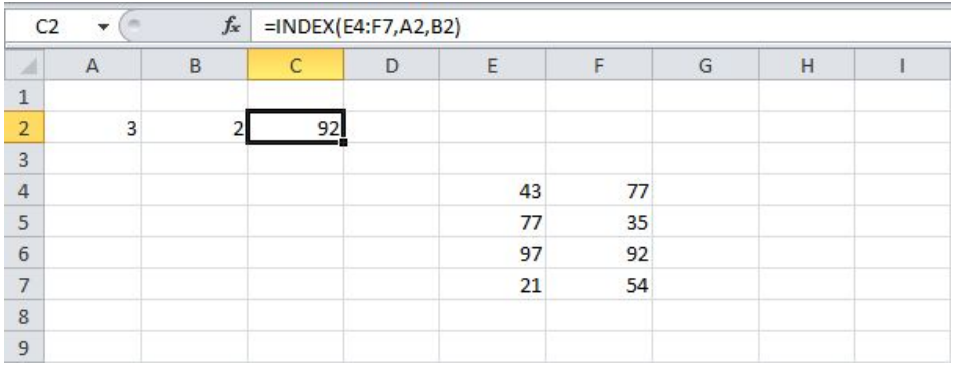
B2 =MATCH(A2,E4:E7,0)

	A	B	C	D	E	F	G	H	I
1									
2	Yellow	3							
3									
4					Green				
5					Blue				
6					Yellow				
7					White				
8									
9									

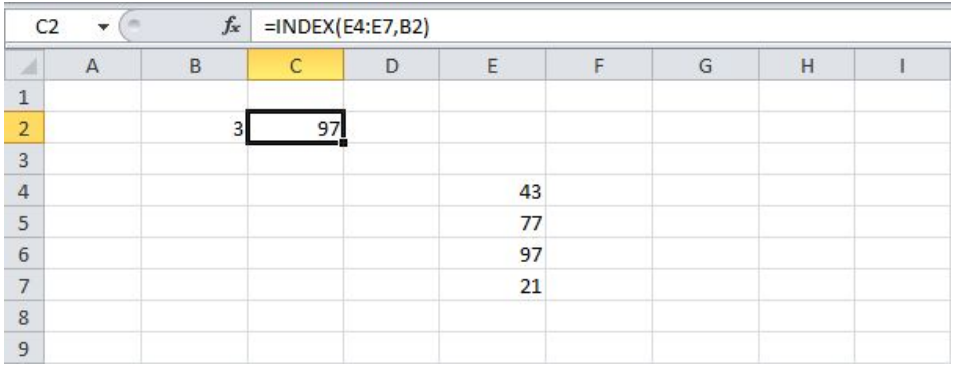
Index

The INDEX function returns a specific value in a two-dimensional or one-dimensional range.

Note: 92 found at the intersection of row 3 and column 2 in the range E4:F7.



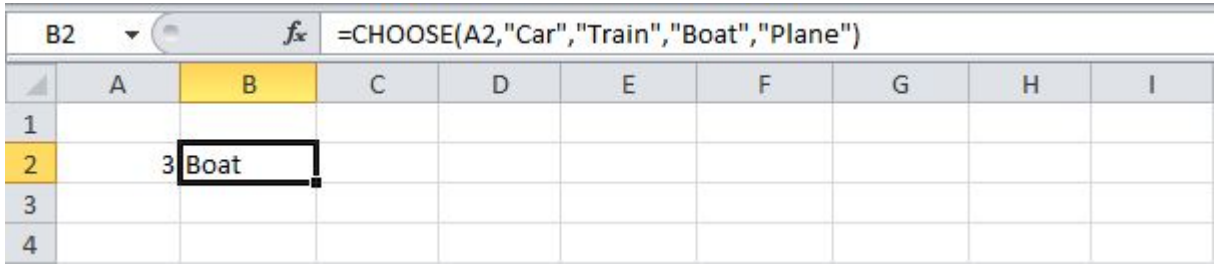
Note: 97 found at position 3 in the range E4:E7.



Choose

The CHOOSE function returns a value from a list of values, based on a position number.

Note: Boat found at position 3.



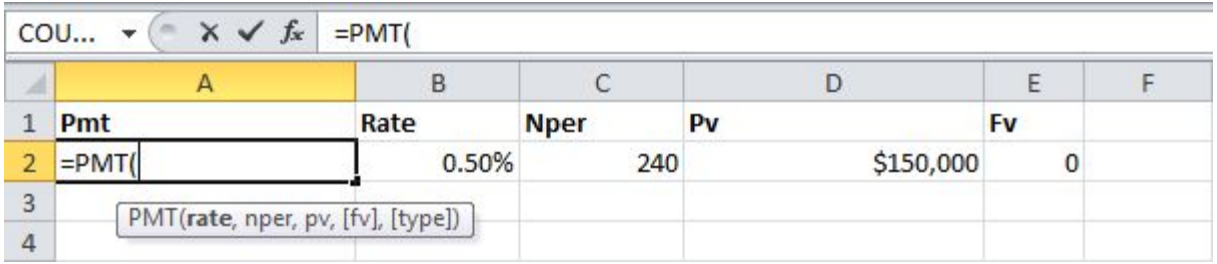
6. Financial Functions

To illustrate Excel's most popular financial functions, we consider a loan with monthly payments, an annual interest rate of 6%, a 20-year duration, a present value of $150,000 (amount borrowed) and a future value of 0 (that's what you hope to achieve when you pay off a loan).

We make monthly payments, so we use 6%/12 = 0.5% for Rate and 20*12 = 240 for Nper (total number of periods). If we make annual payments on the same loan, we use 6% for Rate and 20 for Nper.

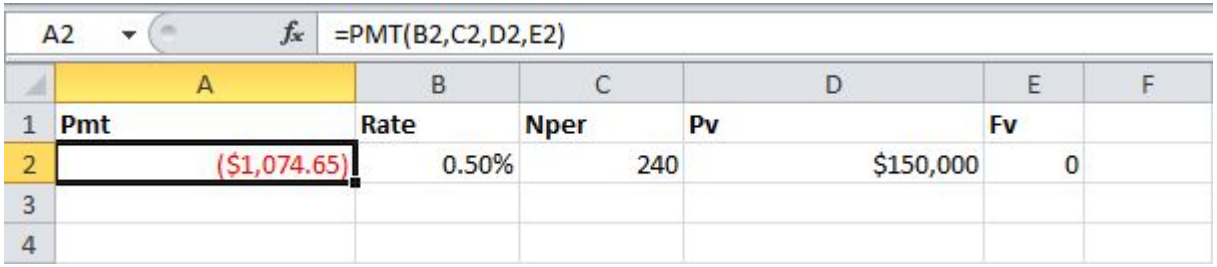
Pmt

Select cell A2 and insert the PMT function.



Note: The last two arguments are optional. For loans the Fv can be omitted (the future value of a loan equals 0, however, it's included here for clarification). If Type is omitted, it is assumed that payments are due at the end of the period.

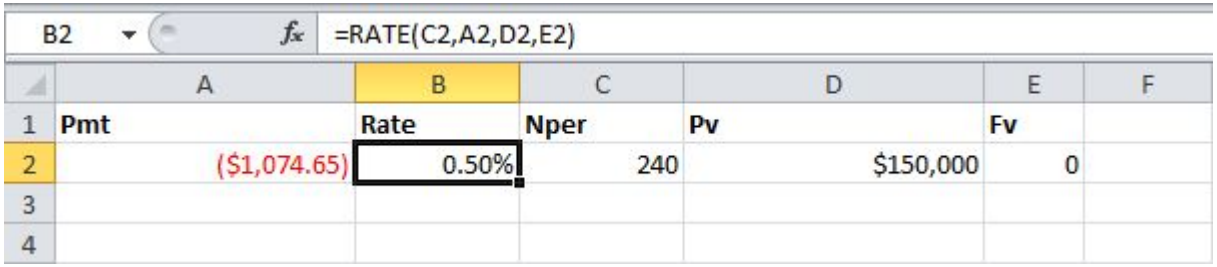
Result. The monthly payment equals $1,074.65.



Tip: when working with financial functions in Excel, always ask yourself the question, am I making a payment (negative) or am I receiving money (positive)? We pay off a loan of $150,000 (positive, we received that amount) and we make monthly payments of $1,074.65 (negative, we pay).

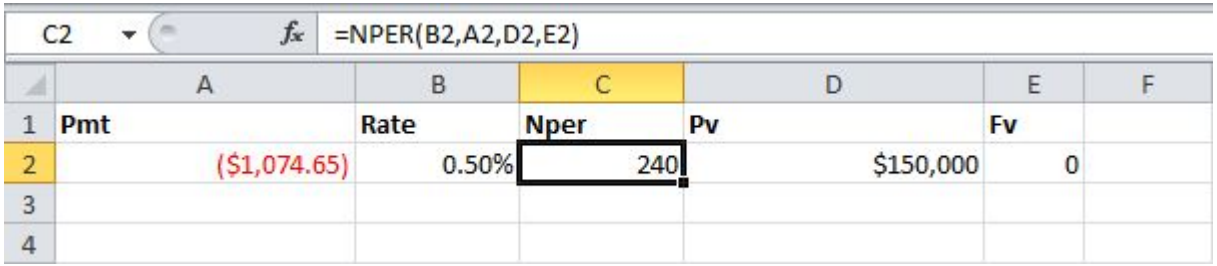
Rate

If Rate is the only unknown variable, we can use the RATE function to calculate the interest rate.



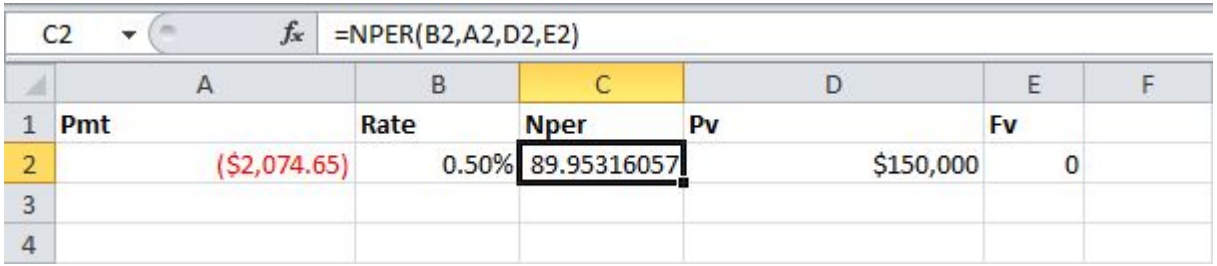
Nper

Or the NPER function. If we make monthly payments of $1,074.65 on a 20-year loan, with an annual interest rate of 6%, it takes 240 months to pay off this loan.



	A	B	C	D	E	F
1	**Pmt**	**Rate**	**Nper**	**Pv**	**Fv**	
2	($1,074.65)	0.50%	240	$150,000	0	
3						
4						

We already knew this, but we can change the monthly payment now to see how this affects the total number of periods.



	A	B	C	D	E	F
1	**Pmt**	**Rate**	**Nper**	**Pv**	**Fv**	
2	($2,074.65)	0.50%	89.95316057	$150,000	0	
3						
4						

Conclusion: if we make monthly payments of $2,074.65, it takes less than 90 months to pay off this loan.

Pv

Or the PV (Present Value) function. If we make monthly payments of $1,074.65 on a 20-year loan, with an annual interest rate of 6%, how much can we borrow? You already know the answer.

D2 =PV(B2,C2,A2,E2)

	A	B	C	D	E	F
1	**Pmt**	**Rate**	**Nper**	**Pv**	**Fv**	
2	($1,074.65)	0.50%	240	$150,000	0	
3						
4						

Fv

And we finish this chapter with the FV (Future Value) function. If we make monthly payments of $1,074.65 on a 20-year loan, with an annual interest rate of 6%, do we pay off this loan? Yes.

E2 =FV(B2,C2,A2,D2)

	A	B	C	D	E	F
1	**Pmt**	**Rate**	**Nper**	**Pv**	**Fv**	
2	($1,074.65)	0.50%	240	$150,000	0	
3						
4						

But, if we make monthly payments of only $1,000.00, we still have debt after 20 years.

E2 =FV(B2,C2,A2,D2)

	A	B	C	D	E	F
1	**Pmt**	**Rate**	**Nper**	**Pv**	**Fv**	
2	($1,000.00)	0.50%	240	$150,000	($34,489.78)	
3						
4						

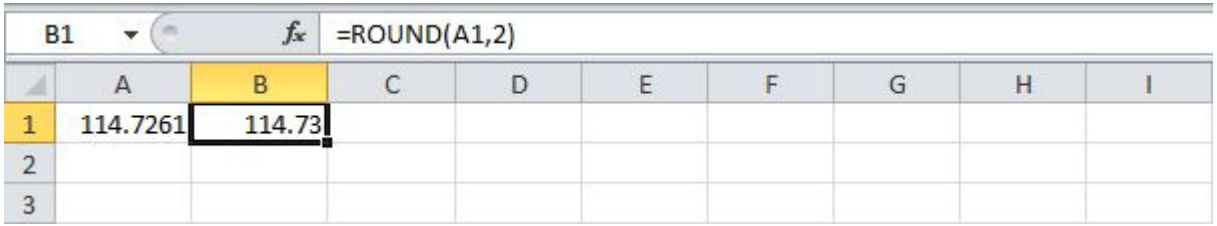
7. Round Function

This chapter illustrates three functions to round numbers in Excel. The ROUND, ROUNDUP and ROUNDDOWNfunction.

Before your start: if you round a number, you lose precision. If you don't want this, show fewer decimal places without changing the number itself.

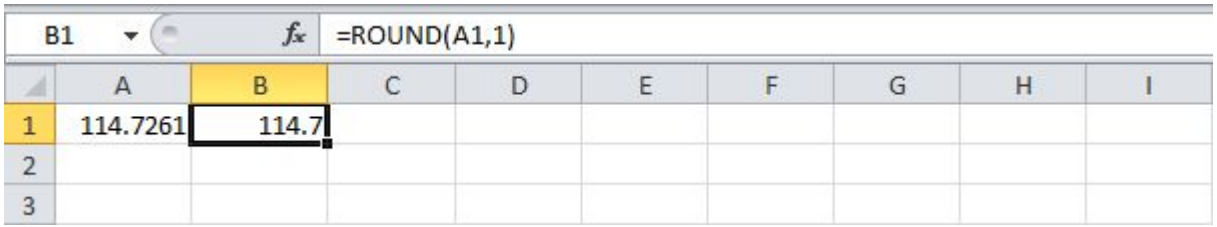
Round

1. Round a number to two decimal places.

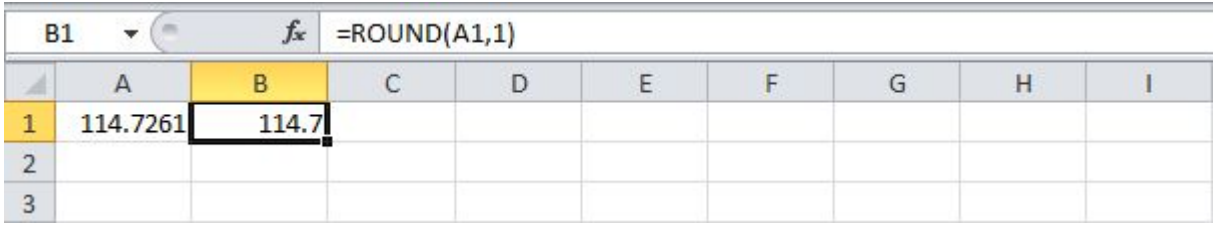


Note: 1, 2, 3, and 4 get rounded down. 5, 6, 7, 8, and 9 get rounded up. In this example, 114.7211, 114.7221, 114.7231 and 114.7241 get rounded down to 114.72 and 114.7251, 114.7261, 114.7271, 114.7281 and 114.7291 get rounded up to 114.73.

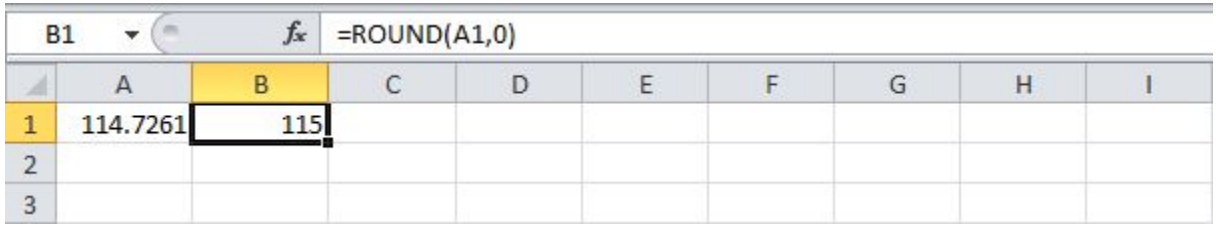
2. Round a number to one decimal place.



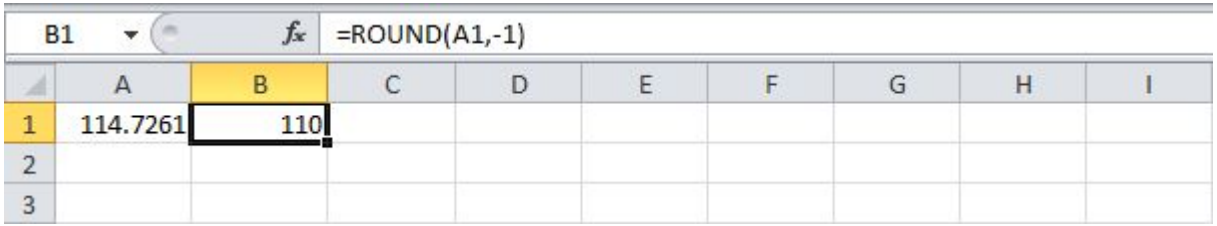
2. Round a number to one decimal place.



3. Round a number to the nearest integer.

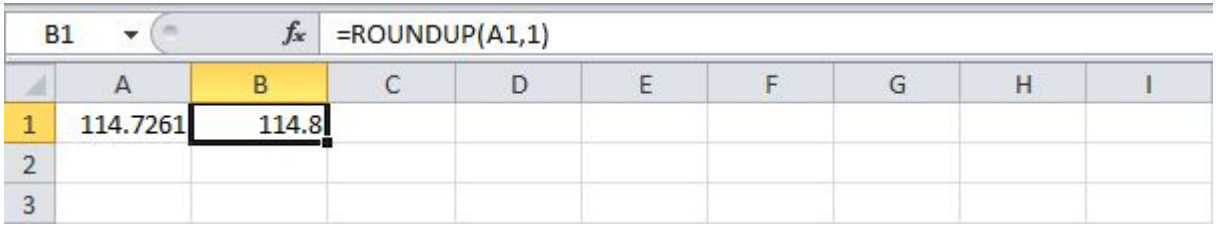


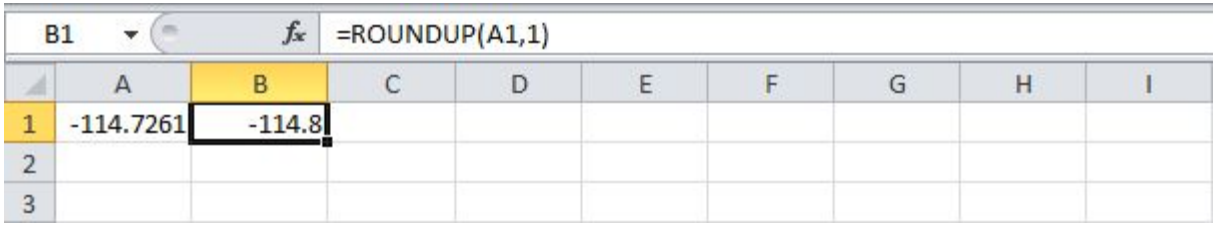
4. Round a number to the nearest 10.



RoundUp

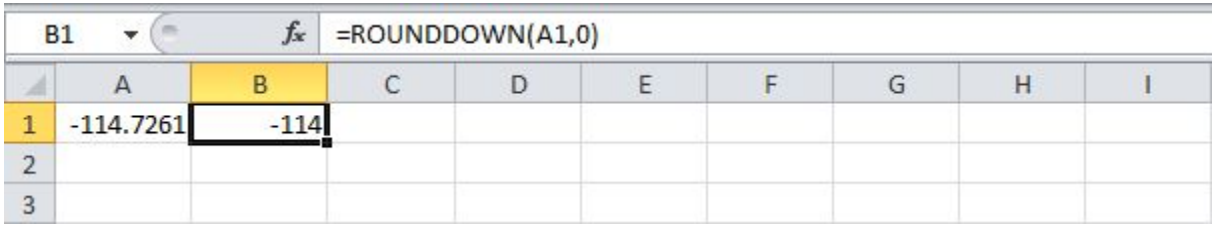
The ROUNDUP function always rounds a number up (away from zero). For example, round a number up to one decimal place.





RoundDown

The ROUNDDOWN function always rounds a number down (toward zero). For example, round a number down to the nearest integer.

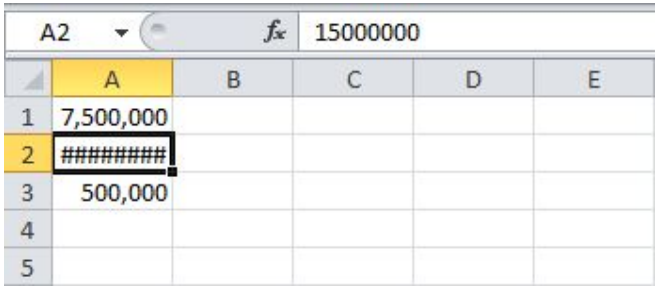


8. Formula and Errors

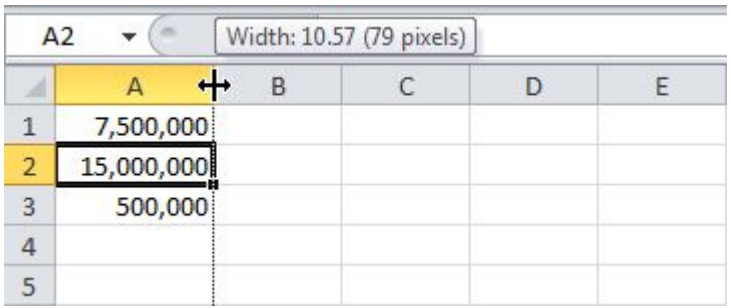
This chapter teaches you how to deal with some common formula errors in Excel.

error

When your cell contains this error code, the column isn't wide enough to display the value.



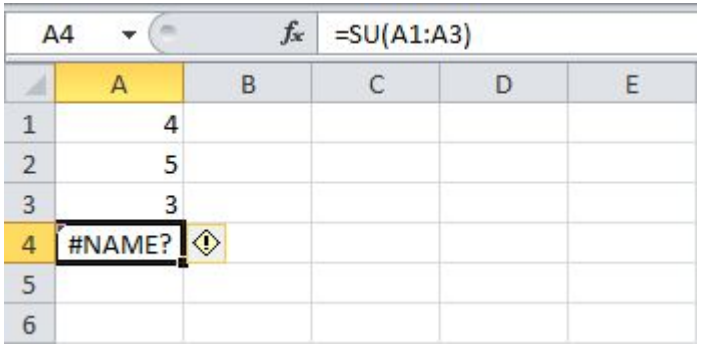
1. Click on the right border of the column A header and increase the column width.



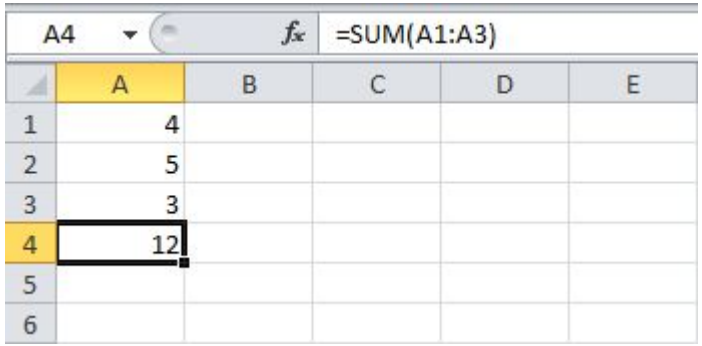
Tip: double click the right border of the column A header to automatically fit the widest cell in column A.

#NAME? error

The #NAME? error occurs when Excel does not recognize text in a formula.



1. Simply correct SU to SUM.

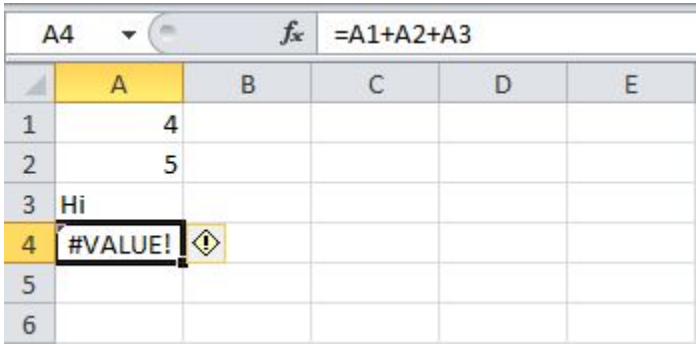


#VALUE! error

Excel displays the #VALUE! error when a formula has the wrong type of argument.

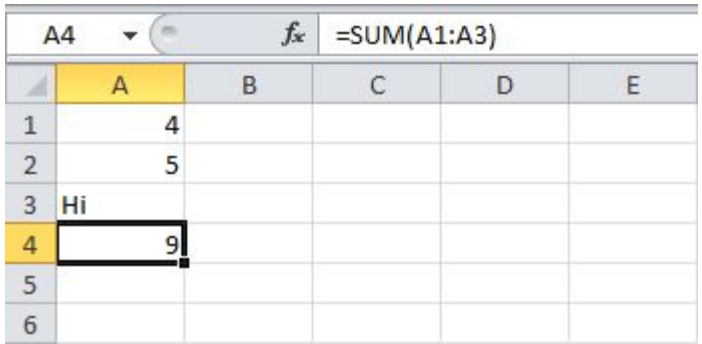
1a. Change the value of cell A3 to a number.

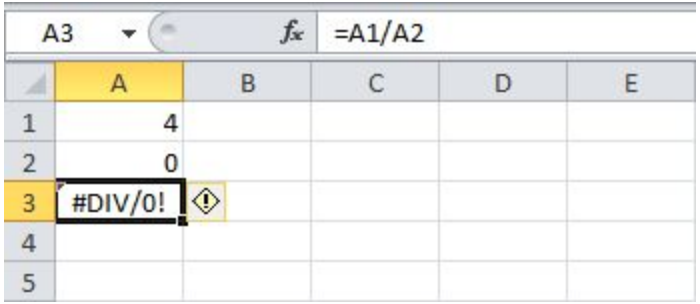
1b. Use a function to ignore cells that contain text.



#DIV/0! error

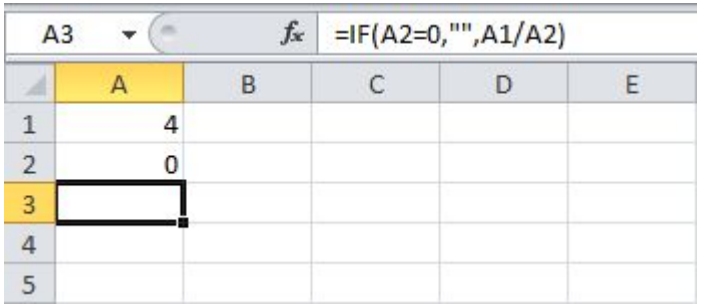
Excel displays the #DIV/0! error when a formula tries to divide a number by 0 or an empty cell.





1a. Change the value of cell A2 to a value that is not equal to 0.

1b. Prevent the error from being displayed by using the logical function IF.

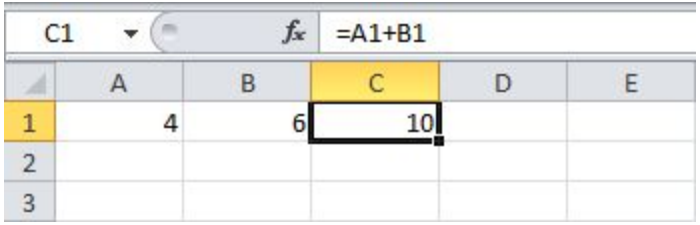


Explanation: if cell A2 equals 0, an empty string is displayed. If not, the result of the formula A1/A2 is displayed.

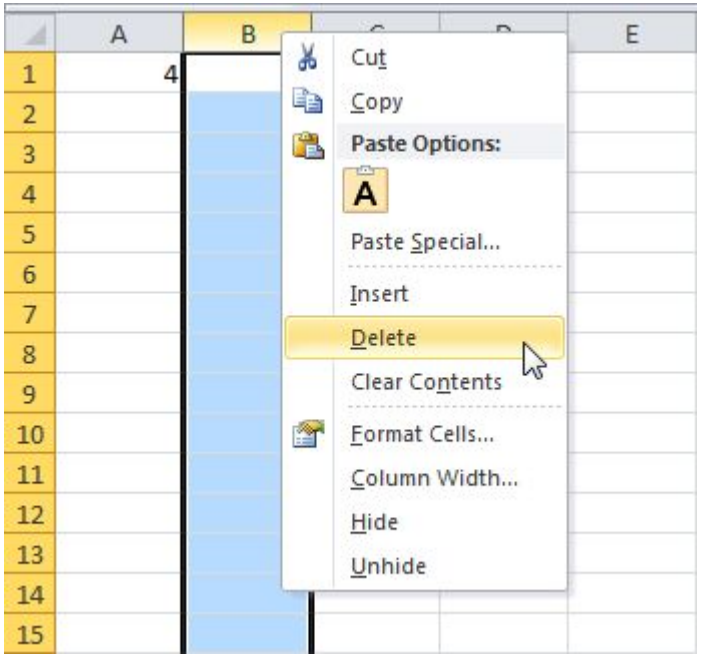
#REF! error

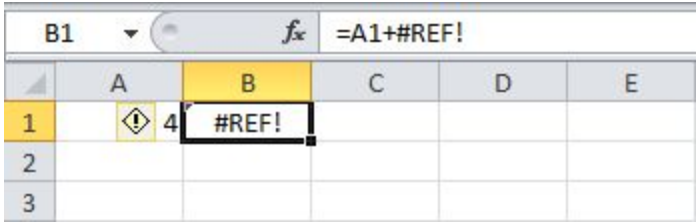
Excel displays the #REF! error when a formula refers to a cell that is not valid.

1. Cell C1 references cell A1 and cell B1.



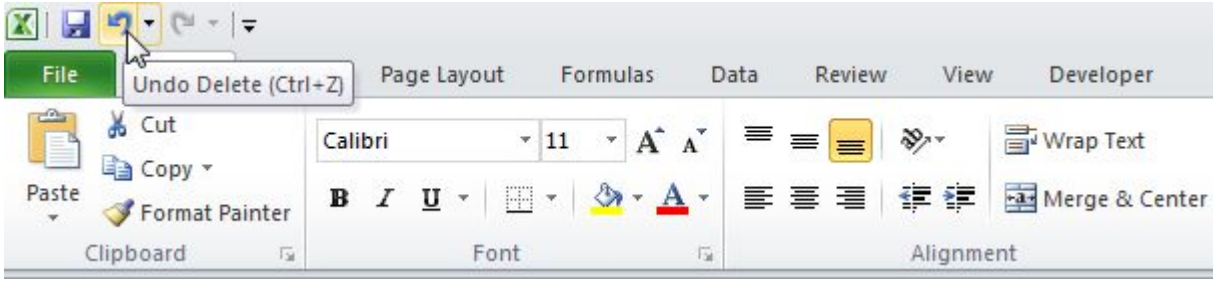
2. Delete column B. To achieve this, right click the column B header and click Delete.





3. Select cell B1. The reference to cell B1 is not valid anymore.

4. To fix this error, you can either delete +#REF! in the formula of cell B1 or you can undo your action by clicking Undo in the Quick Access Toolbar (or press CTRL + z).

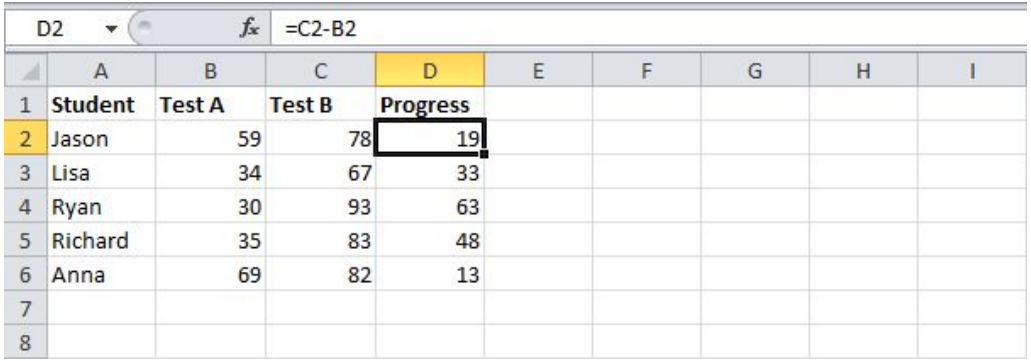


9. Array Formulas

Without Array Formula

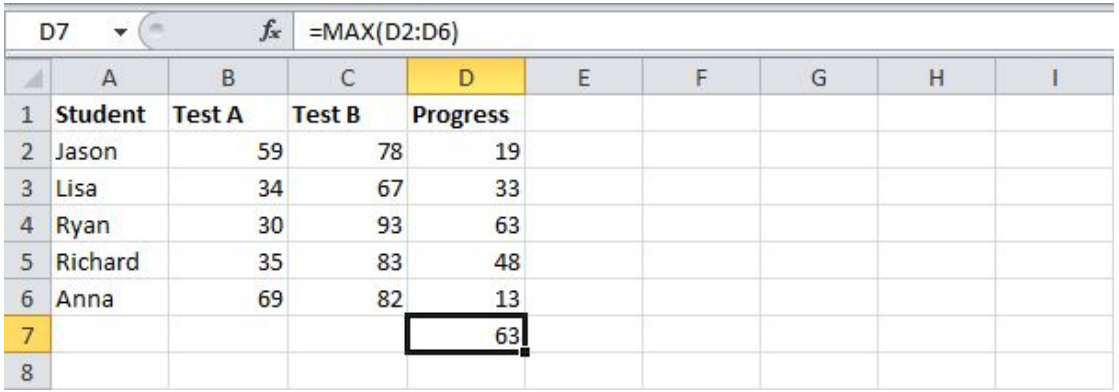
Without using an array formula, we would execute the following steps to find the greatest progress.

1. First, we would calculate the progress of each student.



	A	B	C	D	E	F	G	H	I
1	**Student**	**Test A**	**Test B**	**Progress**					
2	Jason	59	78	19					
3	Lisa	34	67	33					
4	Ryan	30	93	63					
5	Richard	35	83	48					
6	Anna	69	82	13					
7									
8									

2. Next, we would use the MAX function to find the greatest progress.

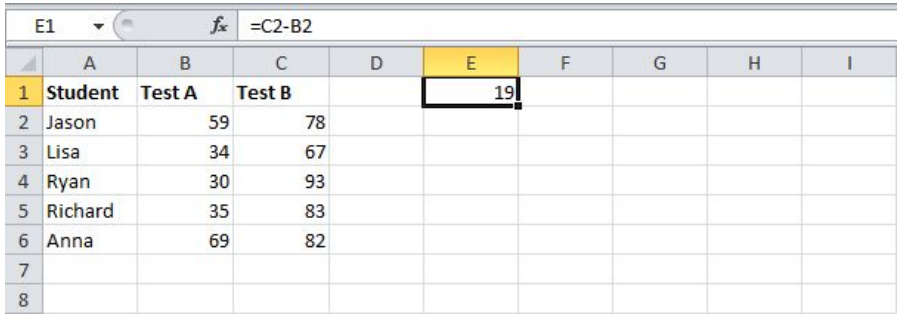


	A	B	C	D	E	F	G	H	I
1	**Student**	**Test A**	**Test B**	**Progress**					
2	Jason	59	78	19					
3	Lisa	34	67	33					
4	Ryan	30	93	63					
5	Richard	35	83	48					
6	Anna	69	82	13					
7				63					
8									

With Array Formula

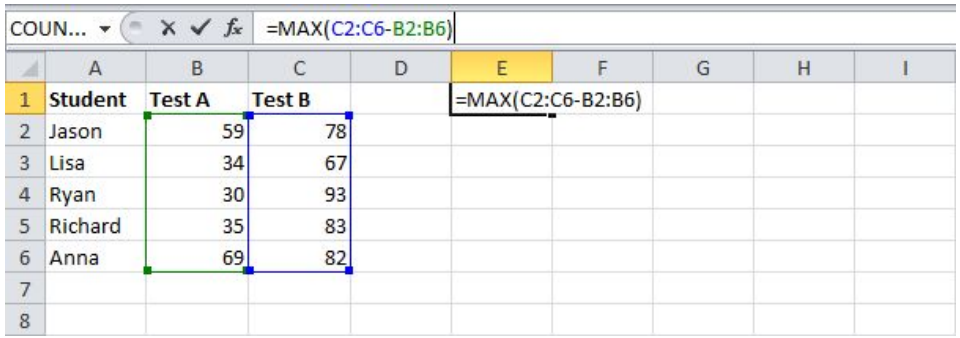
We don't need to store the range in column D. Excel can store this range in its memory. A range stored in Excel's memory is called an array constant.

1. We already know that we can find the progress of the first student by using the formula below.



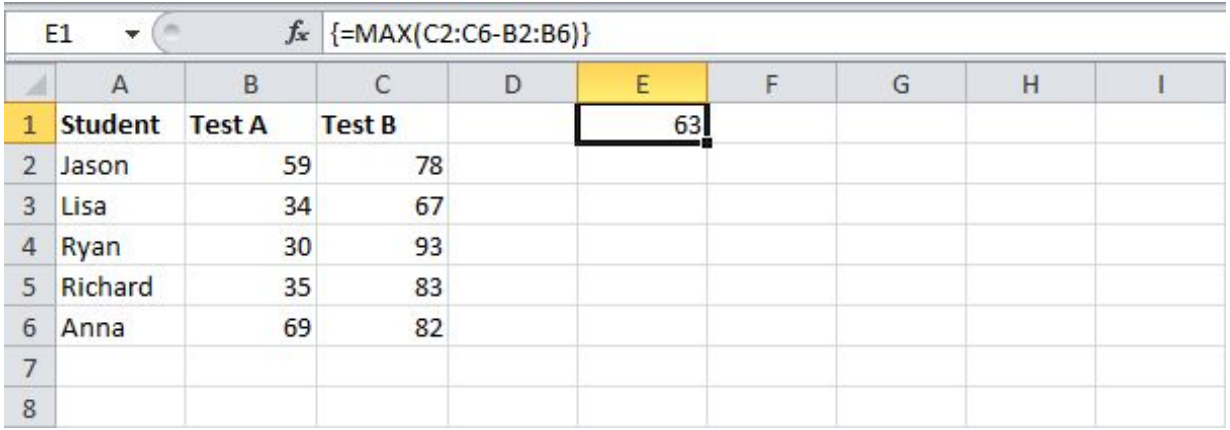
	A	B	C	D	E	F	G	H	I
1	**Student**	**Test A**	**Test B**		19				
2	Jason	59	78						
3	Lisa	34	67						
4	Ryan	30	93						
5	Richard	35	83						
6	Anna	69	82						
7									
8									

2. To find the greatest progress (don't be overwhelmed), we add the MAX function, replace C2 with C2:C6 and B2 with B2:B6.



	A	B	C	D	E	F	G	H	I
1	**Student**	**Test A**	**Test B**		=MAX(C2:C6-B2:B6)				
2	Jason	59	78						
3	Lisa	34	67						
4	Ryan	30	93						
5	Richard	35	83						
6	Anna	69	82						
7									
8									

3. Finish by pressing CTRL + SHIFT + ENTER.



Note: The formula bar indicates that this is an array formula by enclosing it in curly braces {}. Do not type these yourself. They will disappear when you edit the formula.

Explanation: The range (array constant) is stored in Excel's memory, not in an range. The array constant looks as follows:

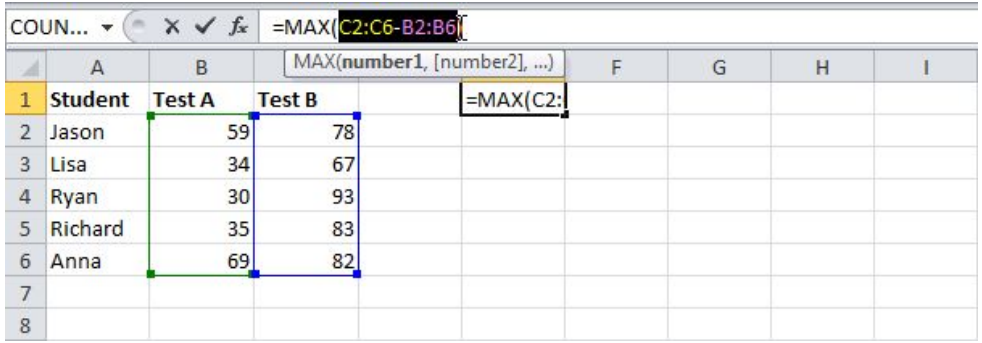
{19;33;63;48;13}

This array constant is used as an argument for the MAX function, giving a result of 63.

F9 Key

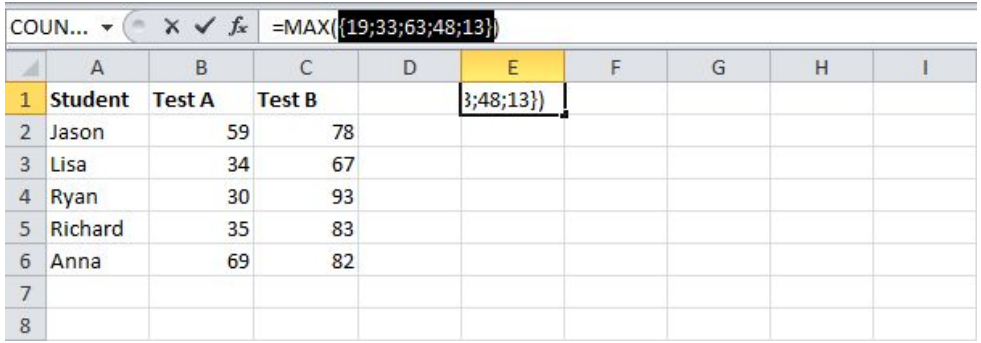
When working with array formulas, you can have a look at these array constants yourself.

1. Select C2:C6-B2:B6 in the formula.



2. Press F9.

That looks good. Elements in a vertical array constant are separated by semicolons. Elements in a horizontal array constant are separated by commas.



End

This is the end of the basic functions used for Data Analysis in MS Excel

Chapter 2

Intermediate Data Analytics using MS Excel

This section illustrates the powerful features Excel has to offer to analyze data.

This section illustrates the powerful features Excel has to offer to analyze data.

1 Sort: You can sort your Excel data on one column or multiple columns. You can sort in ascending or descending order.

2 Filter: Filter your Excel data if you only want to display records that meet certain criteria.

3 Conditional Formatting: Conditional formatting in Excel enables you to highlight cells with a certain color, depending on the cell's value.

4 Charts: A simple Excel chart can say more than a sheet full of numbers. As you'll see, creating charts is very easy.

5 Pivot Tables: Pivot tables are one of Excel's most powerful features. A pivot table allows you to extract the significance from a large, detailed data set.

6 Tables: Tables allow you to analyze your data in Excel quickly and easily.

7 What-If Analysis: What-If Analysis in Excel allows you to try out different values (scenarios) for formulas.

8 Solver: Excel includes a tool called solver that uses techniques from the operations research to find optimal solutions for all kind of decision problems.

9 Analysis ToolPak: The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

1. Sorting

You can sort your Excel data on one column or multiple columns. You can sort in ascending or descending order.

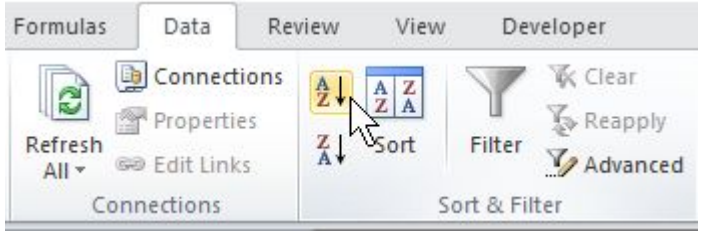
One Column

To sort on one column, execute the following steps.

1. Click any cell in the column you want to sort.

	A	B	C	D	E
1	**Last Name**	**Sales**	**Country**	**Quarter**	
2	Smith	$16,753.00	UK	Qtr 3	
3	Johnson	$14,808.00	USA	Qtr 4	
4	Williams	$10,644.00	UK	Qtr 2	
5	Jones	$1,390.00	USA	Qtr 3	
6	Brown	$4,865.00	USA	Qtr 4	
7	Williams	$12,438.00	UK	Qtr 1	
8	Johnson	$9,339.00	UK	Qtr 2	
9	Smith	$18,919.00	USA	Qtr 3	
10	Jones	$9,213.00	USA	Qtr 4	
11	Jones	$7,433.00	UK	Qtr 1	
12	Brown	$3,255.00	USA	Qtr 2	
13	Williams	$14,867.00	USA	Qtr 3	
14	Williams	$19,302.00	UK	Qtr 4	
15	Smith	$9,698.00	USA	Qtr 1	
16					
17					

2. To sort in ascending order, on the Data tab, click AZ.



Result:

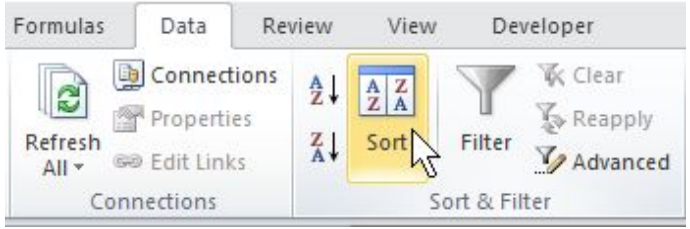
A2 Brown

	A	B	C	D	E
1	**Last Name**	**Sales**	**Country**	**Quarter**	
2	Brown	$4,865.00	USA	Qtr 4	
3	Brown	$3,255.00	USA	Qtr 2	
4	Johnson	$14,808.00	USA	Qtr 4	
5	Johnson	$9,339.00	UK	Qtr 2	
6	Jones	$1,390.00	USA	Qtr 3	
7	Jones	$9,213.00	USA	Qtr 4	
8	Jones	$7,433.00	UK	Qtr 1	
9	Smith	$16,753.00	UK	Qtr 3	
10	Smith	$18,919.00	USA	Qtr 3	
11	Smith	$9,698.00	USA	Qtr 1	
12	Williams	$10,644.00	UK	Qtr 2	
13	Williams	$12,438.00	UK	Qtr 1	
14	Williams	$14,867.00	USA	Qtr 3	
15	Williams	$19,302.00	UK	Qtr 4	
16					
17					

Multiple Columns

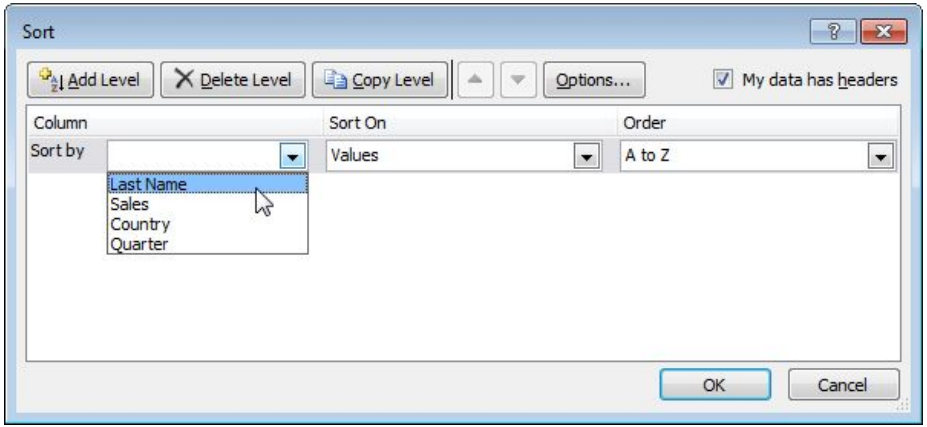
To sort on multiple columns, execute the following steps.

1. On the Data tab, click Sort.



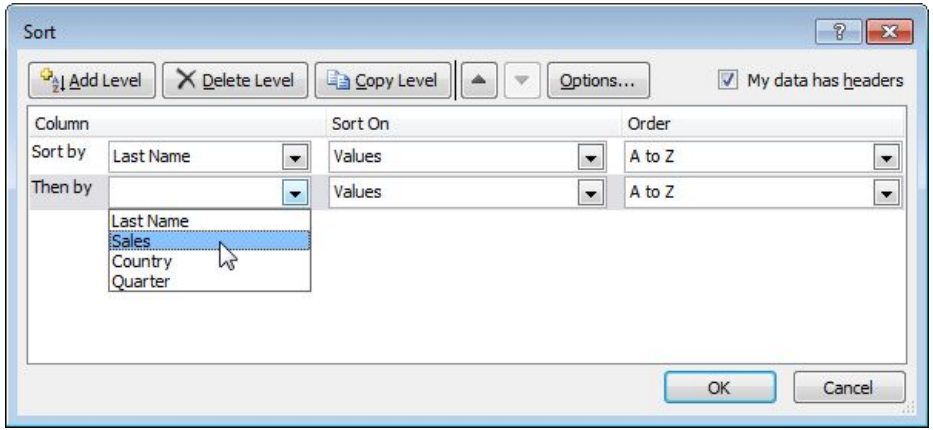
The Sort dialog box appears.

2. Select Last Name from the 'Sort by' drop-down list.



3. Click on Add Level.

4. Select Sales from the 'Then by' drop-down list.



5. Click OK.

Result. Records are sorted by Last Name first and Sales second.

A2 | Brown

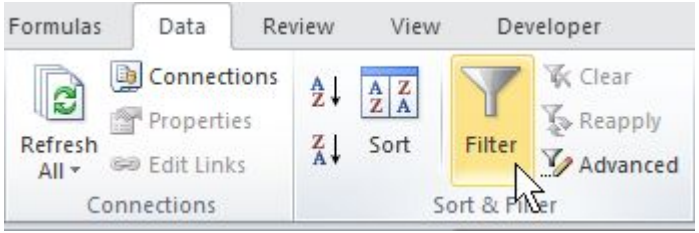
	A	B	C	D	E
1	Last Name	Sales	Country	Quarter	
2	Brown	$3,255.00	USA	Qtr 2	
3	Brown	$4,865.00	USA	Qtr 4	
4	Johnson	$9,339.00	UK	Qtr 2	
5	Johnson	$14,808.00	USA	Qtr 4	
6	Jones	$1,390.00	USA	Qtr 3	
7	Jones	$7,433.00	UK	Qtr 1	
8	Jones	$9,213.00	USA	Qtr 4	
9	Smith	$9,698.00	USA	Qtr 1	
10	Smith	$16,753.00	UK	Qtr 3	
11	Smith	$18,919.00	USA	Qtr 3	
12	Williams	$10,644.00	UK	Qtr 2	
13	Williams	$12,438.00	UK	Qtr 1	
14	Williams	$14,867.00	USA	Qtr 3	
15	Williams	$19,302.00	UK	Qtr 4	
16					
17					

2. Filter

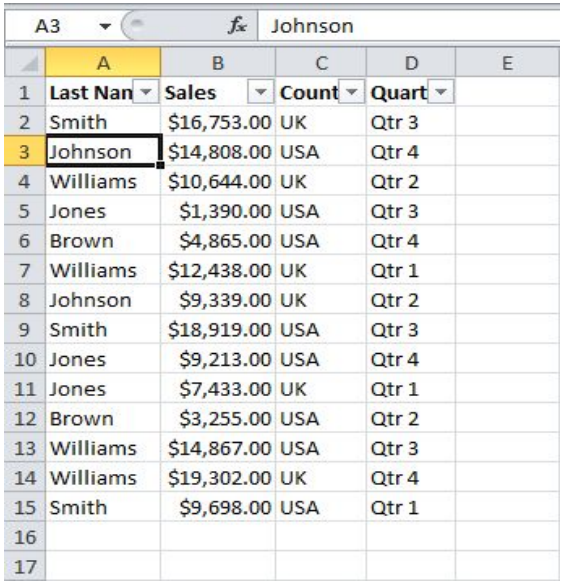
Filter your Excel data if you only want to display records that meet certain criteria.

1. Click any single cell inside a data set.

2. On the Data tab, click Filter.



Arrows in the column headers appear.

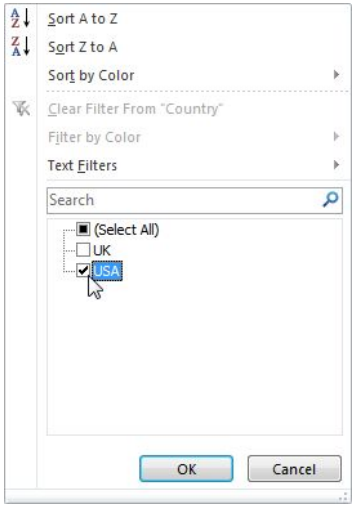


A3 | Johnson

	A	B	C	D	E
1	Last Nan	Sales	Count	Quart	
2	Smith	$16,753.00	UK	Qtr 3	
3	Johnson	$14,808.00	USA	Qtr 4	
4	Williams	$10,644.00	UK	Qtr 2	
5	Jones	$1,390.00	USA	Qtr 3	
6	Brown	$4,865.00	USA	Qtr 4	
7	Williams	$12,438.00	UK	Qtr 1	
8	Johnson	$9,339.00	UK	Qtr 2	
9	Smith	$18,919.00	USA	Qtr 3	
10	Jones	$9,213.00	USA	Qtr 4	
11	Jones	$7,433.00	UK	Qtr 1	
12	Brown	$3,255.00	USA	Qtr 2	
13	Williams	$14,867.00	USA	Qtr 3	
14	Williams	$19,302.00	UK	Qtr 4	
15	Smith	$9,698.00	USA	Qtr 1	
16					
17					

3. Click the arrow next to Country.

4. Click on Select All to clear all the check boxes, and click the check box next to USA.



5. Click OK.

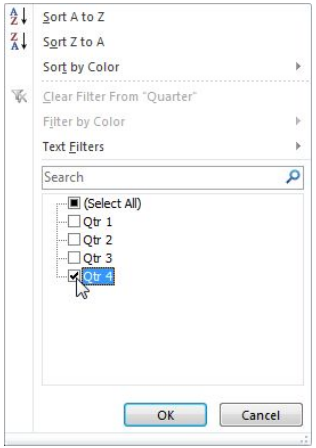
Result. Excel only displays the sales in the USA.

A3 Johnson

	A	B	C	D	E
1	Last Nan	Sales	Count	Quart	
3	Johnson	$14,808.00	USA	Qtr 4	
5	Jones	$1,390.00	USA	Qtr 3	
6	Brown	$4,865.00	USA	Qtr 4	
9	Smith	$18,919.00	USA	Qtr 3	
10	Jones	$9,213.00	USA	Qtr 4	
12	Brown	$3,255.00	USA	Qtr 2	
13	Williams	$14,867.00	USA	Qtr 3	
15	Smith	$9,698.00	USA	Qtr 1	
16					
17					

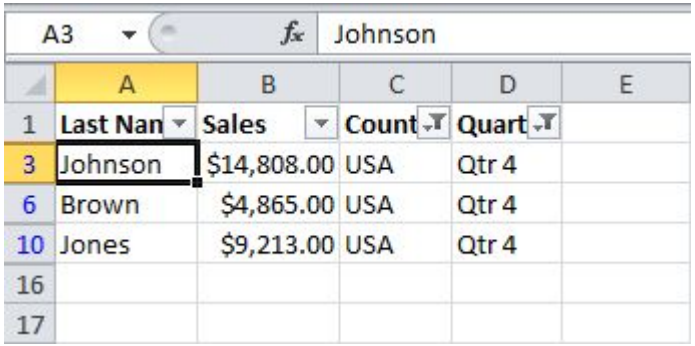
6. Click the arrow next to Quarter.

7. Click on Select All to clear all the check boxes, and click the check box next to Qtr 4.



8. Click OK.

Result. Excel only displays the sales in the USA in Qtr 4.

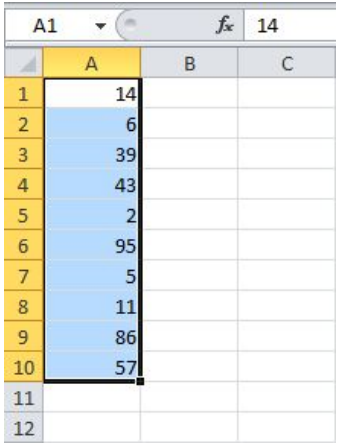


3. Conditional Formatting

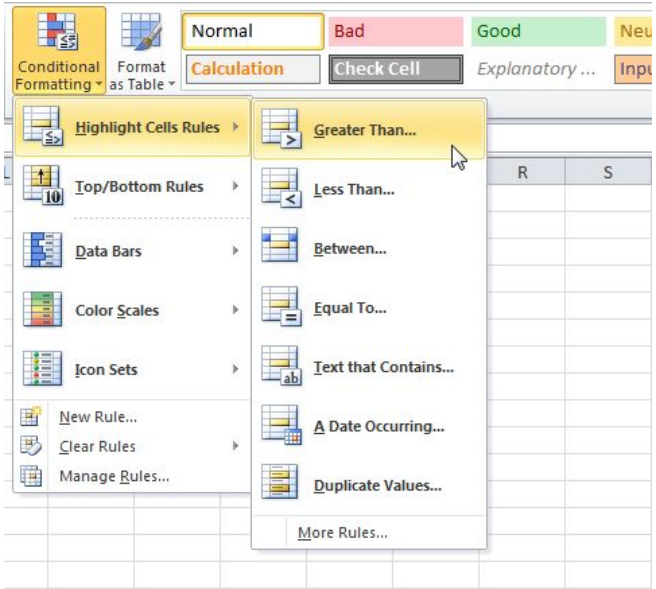
Highlight Cells Rules

To highlight cells that are greater than a value, execute the following steps.

1. Select the range A1:A10.



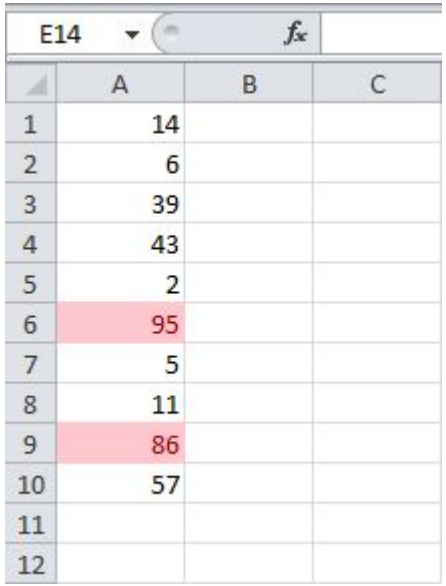
2. On the Home tab, click Conditional Formatting, Highlight Cells Rules, Greater Than...



3. Enter the value 80 and select a formatting style.

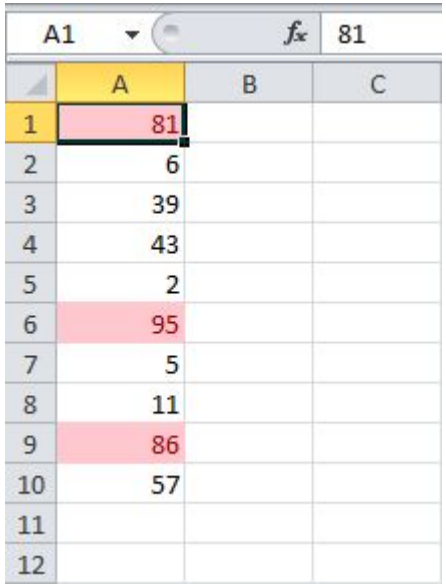
4. Click OK.

Result. Excel highlights the cells that are greater than 80.

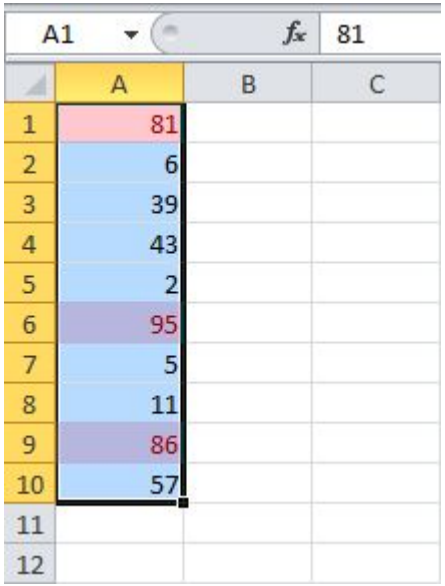


5. Change the value of cell A1 to 81.

Result. Excel changes the format of cell A1 automatically.



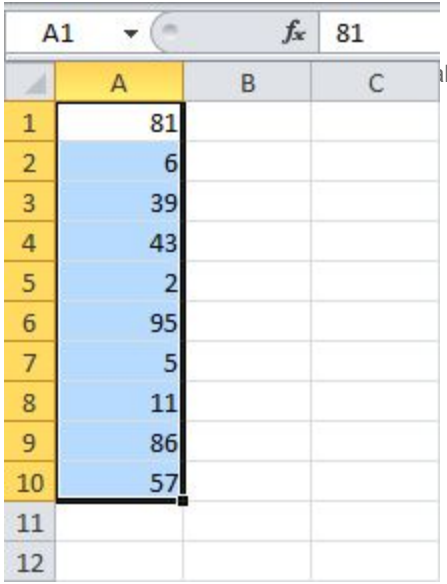
Note: you can also highlight cells that are less than a value, between a low and high value, etc.



Clear Rules

To clear a conditional formatting rule, execute the following steps.

1. Select the range A1:A10.

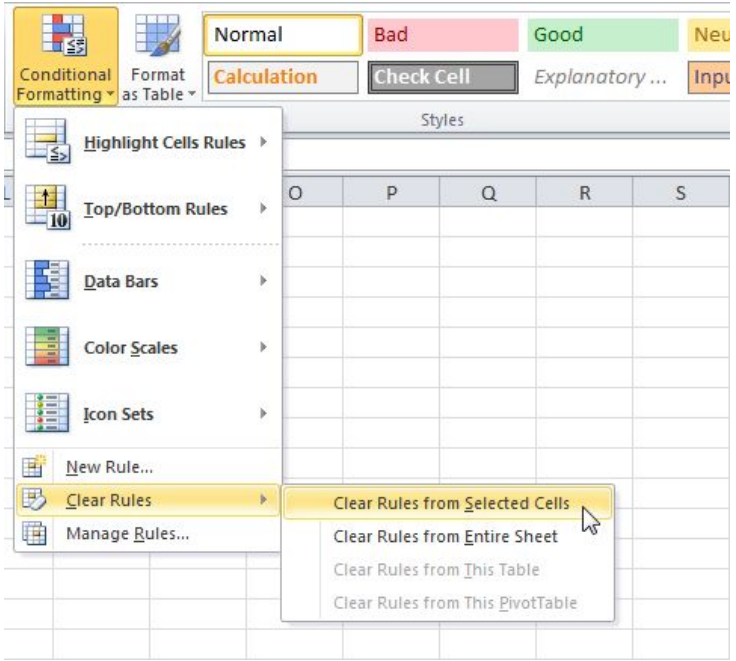


l Formatting, Clear Rules, Clear Rules from Selected Cells.

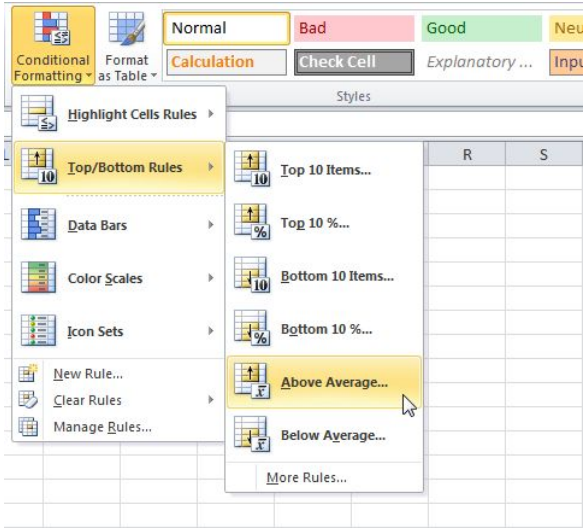
Top/Bottom Rules

To highlight cells that are above the average of the cells, execute the following steps.

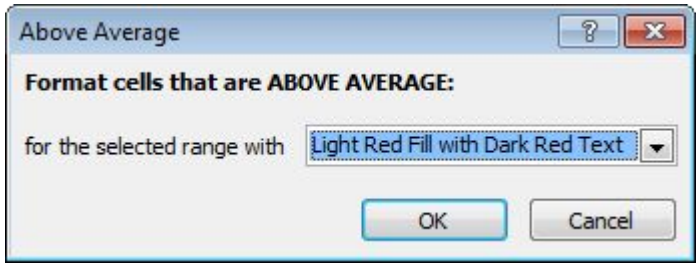
1. Select the range A1:A10.



2. On the Home tab, click Conditional Formatting, Top/Bottom Rules, Above Average...



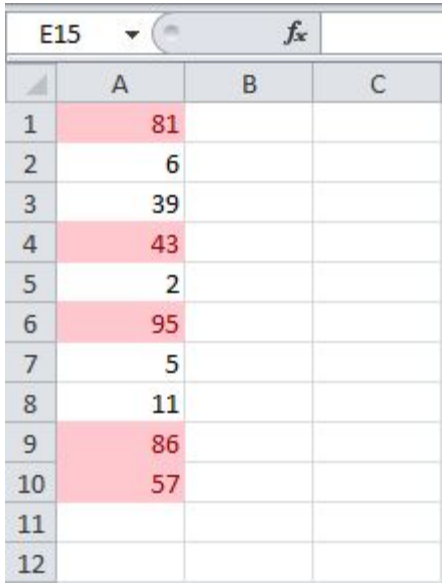
3. Select a formatting style.



4. Click OK.

Result. Excel calculates the average (42.5) and formats the cells that are above this average.

Note: you can also highlight the top 10 items, the top 10 %, etc. The sky is the limit!



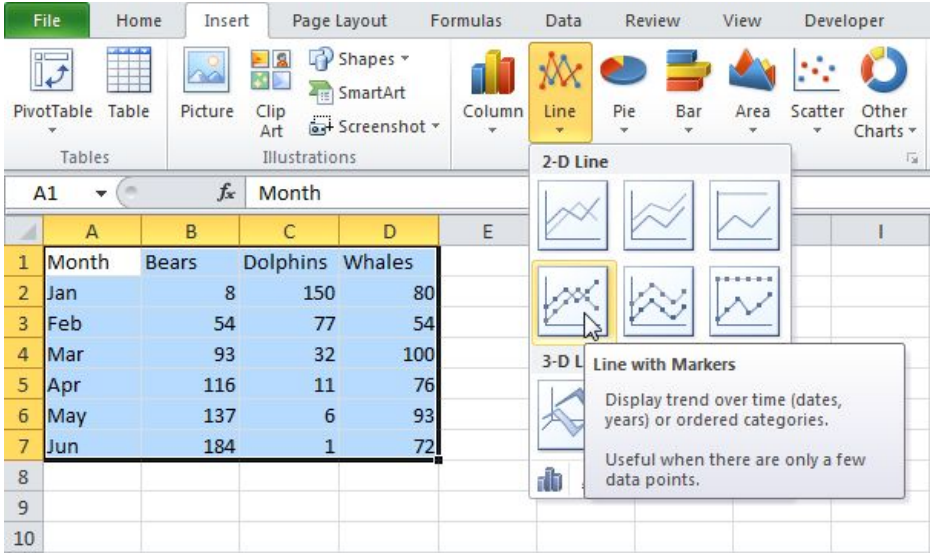
3. Charts

Create a Chart

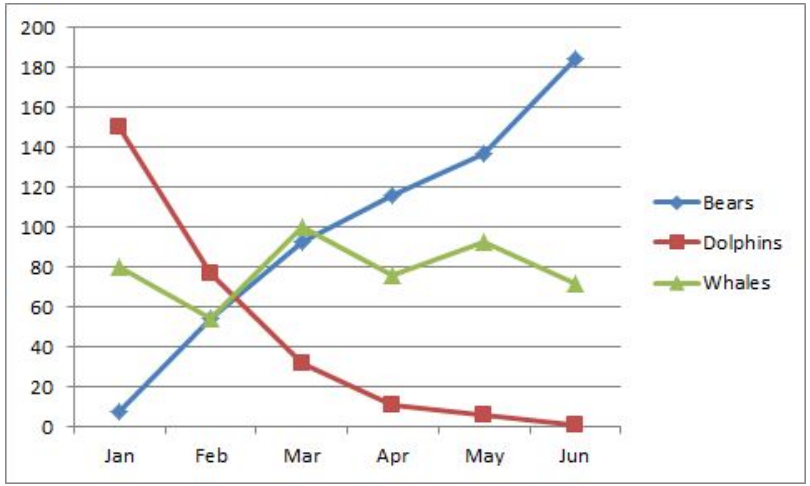
To create a line chart, execute the following steps.

1. Select the range A1:D7.

2. On the Insert tab, in the Charts group, choose Line, and select Line with Markers.



Result:

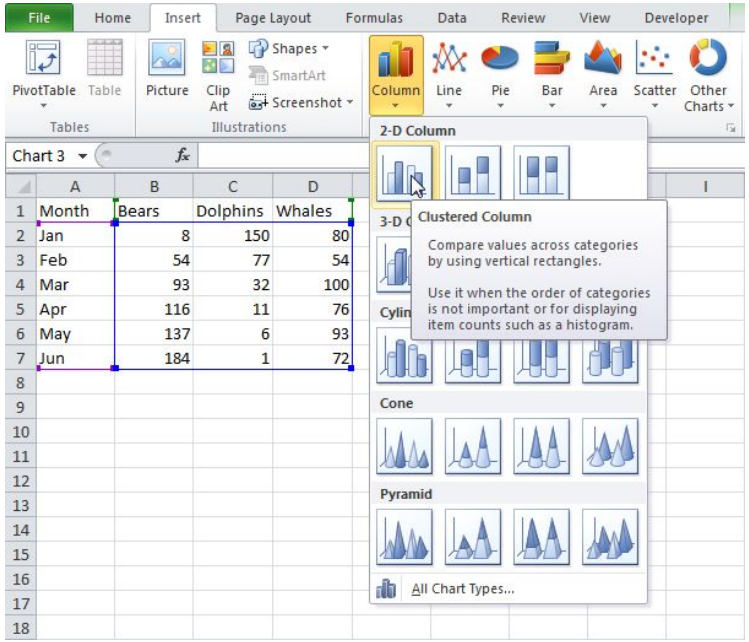


Change Chart Type

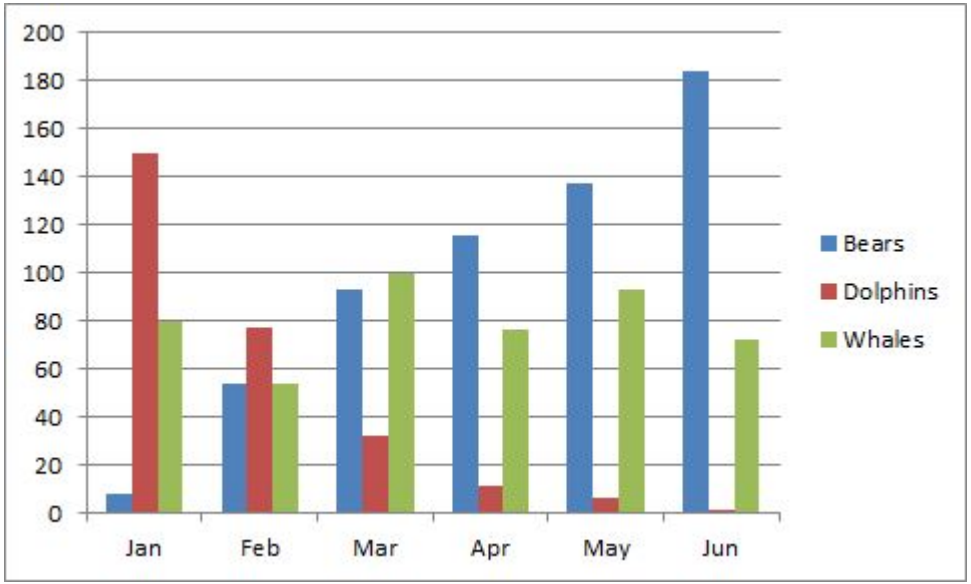
You can easily change to a different type of chart at any time.

1. Select the chart.

2. On the Insert tab, in the Charts group, choose Column, and select Clustered Column.



Result:

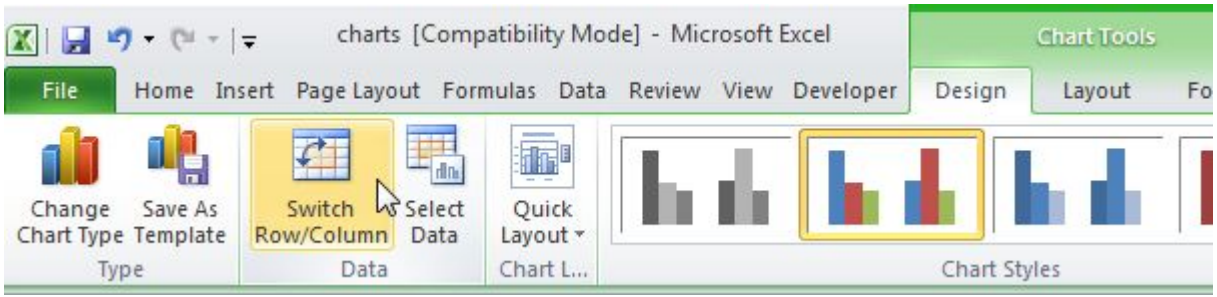


Switch Row/Column

If you want the animals, displayed on the vertical axis, to be displayed on the horizontal axis instead, execute the following steps.

1. Select the chart. The Chart Tools contextual tab activates.

2. On the Design tab, click Switch Row/Column.



Result:

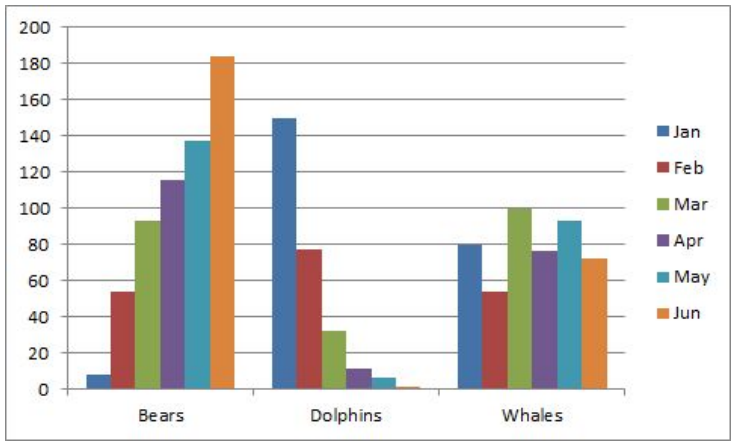
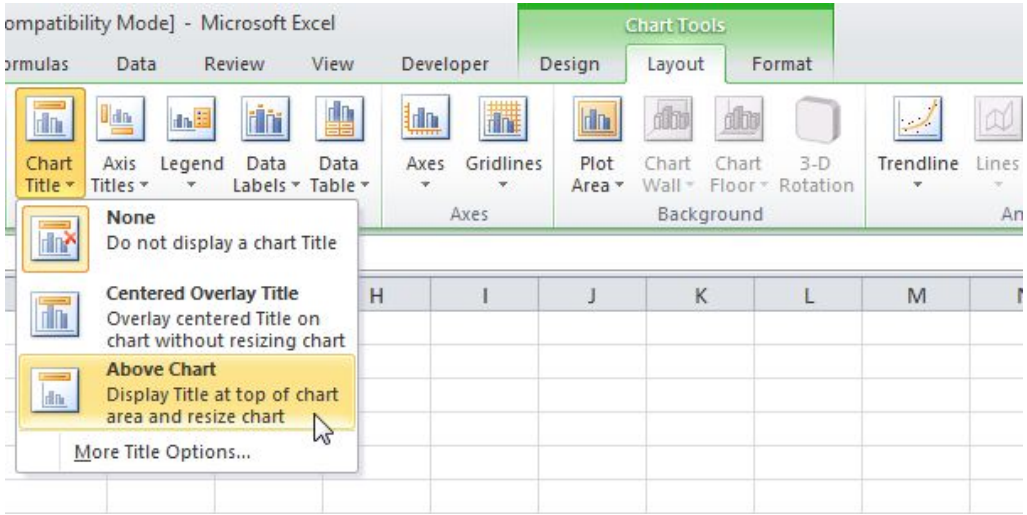


Chart Title

To add a chart title, execute the following steps.

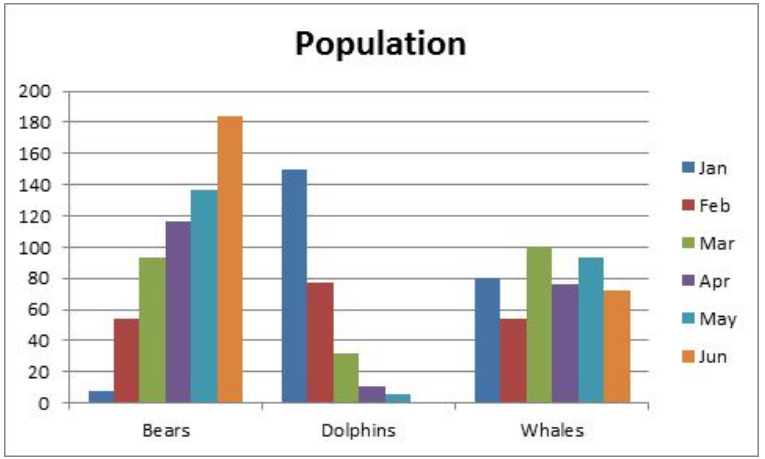
1. Select the chart. The Chart Tools contextual tab activates.

2. On the Layout tab, click Chart Title, Above Chart.



3. Enter a title. For example, Population.

Result:

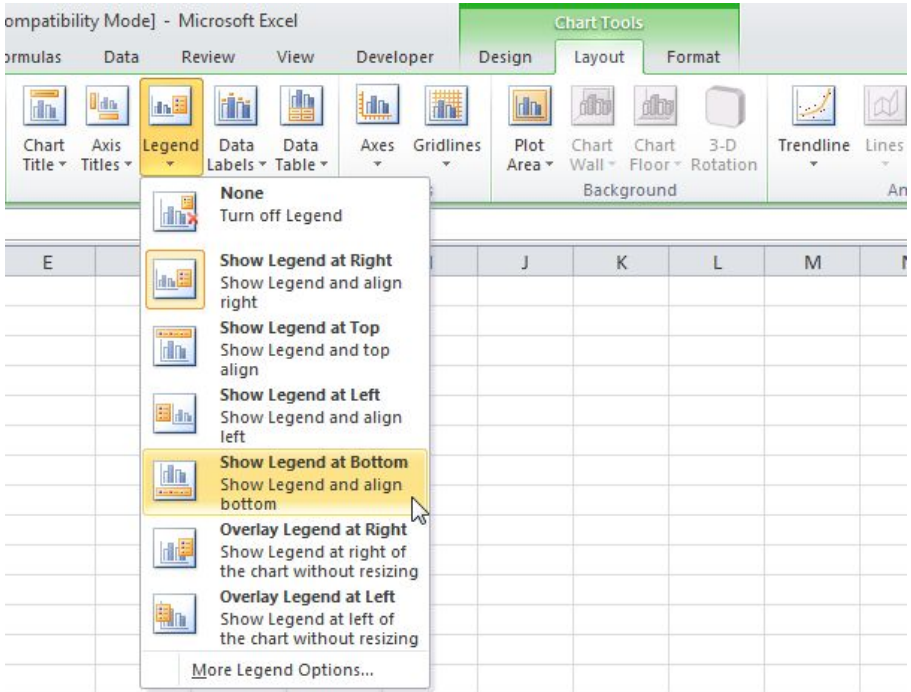


Legend Position

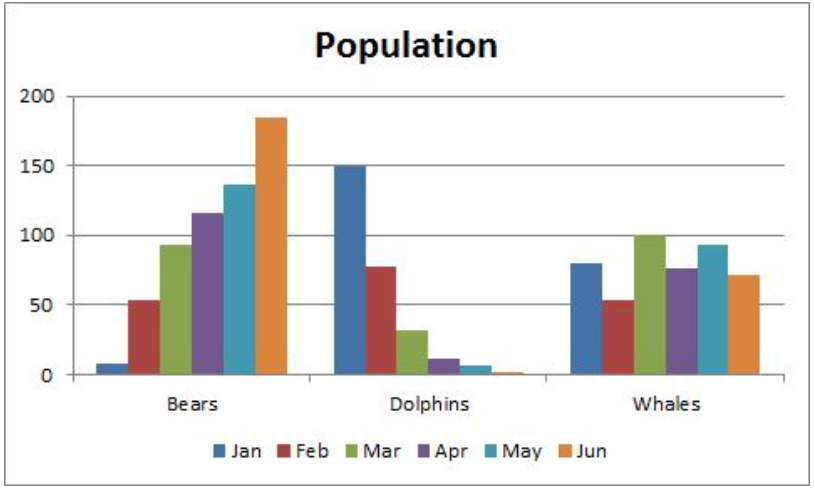
By default, the legend appears to the right of the chart. To move the legend to the bottom of the chart, execute the following steps.

1. Select the chart. The Chart Tools contextual tab activates.

2. On the Layout tab, click Legend, Show Legend at Bottom.



Result:



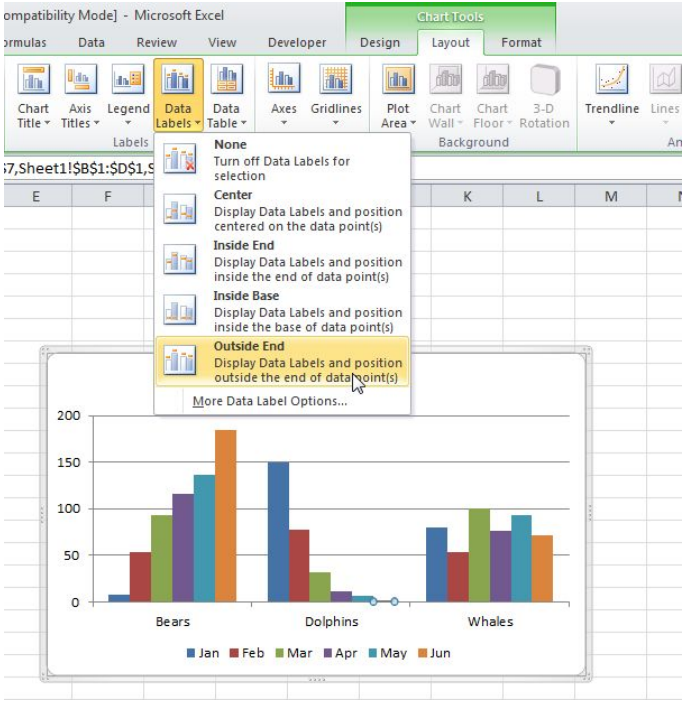
Data Labels

You can use data labels to focus your readers' attention on a single data series or data point.

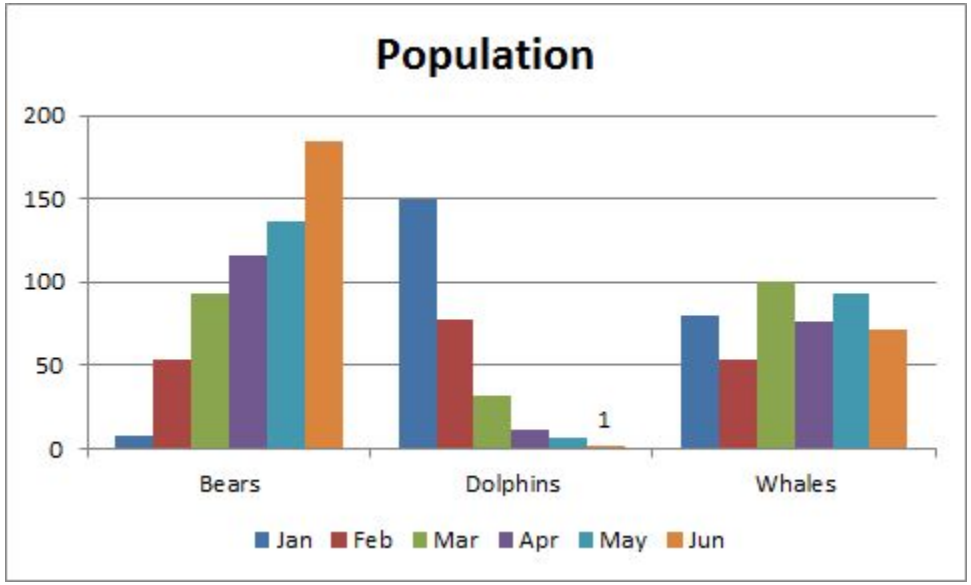
1. Select the chart. The Chart Tools contextual tab activates.

2. Click an orange bar to select the Jun data series. Click again on an orange bar to select a single data point.

3. On the Layout tab, click Data Labels, Outside End.



Result:


Population
200
150
100
50
0
1
Bears
Dolphins
Whales
Jan
Feb
Mar
Apr
May
Jun

4. Pivot Tables

Pivot tables are one of Excel's most powerful features. A pivot table allows you to extract the significance from a large, detailed data set.

Our data set consists of 214 rows and 6 fields. Order ID, Product, Category, Amount, Date and Country.

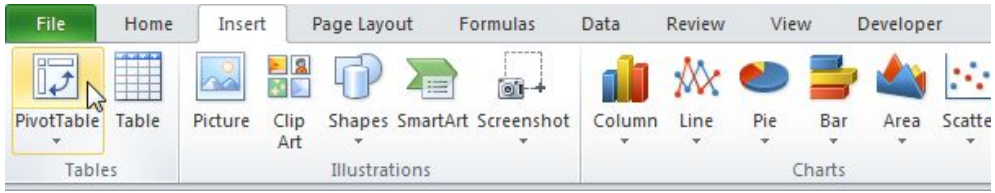
	A	B	C	D	E	F	G	H
1	Order ID	Product	Category	Amount	Date	Country		
2	1	Carrots	Vegetables	$4,270	1/6/2012	United States		
3	2	Broccoli	Vegetables	$8,239	1/7/2012	United Kingdom		
4	3	Banana	Fruit	$617	1/8/2012	United States		
5	4	Banana	Fruit	$8,384	1/10/2012	Canada		
6	5	Beans	Vegetables	$2,626	1/10/2012	Germany		
7	6	Orange	Fruit	$3,610	1/11/2012	United States		
8	7	Broccoli	Vegetables	$9,062	1/11/2012	Australia		
9	8	Banana	Fruit	$6,906	1/16/2012	New Zealand		
10	9	Apple	Fruit	$2,417	1/16/2012	France		

Insert a Pivot Table

To insert a pivot table, execute the following steps.

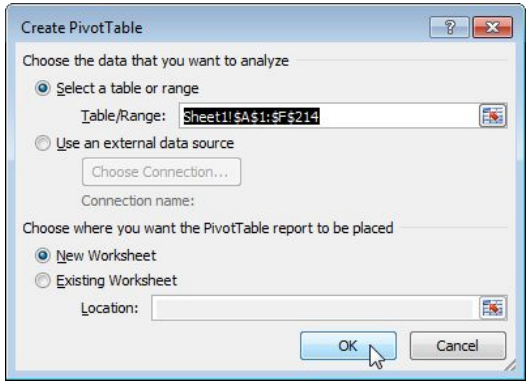
1. Click any single cell inside the data set.

2. On the Insert tab, click PivotTable.



The following dialog box appears. Excel automatically selects the data for you. The default location for a new pivot table is New Worksheet.

3. Click OK.



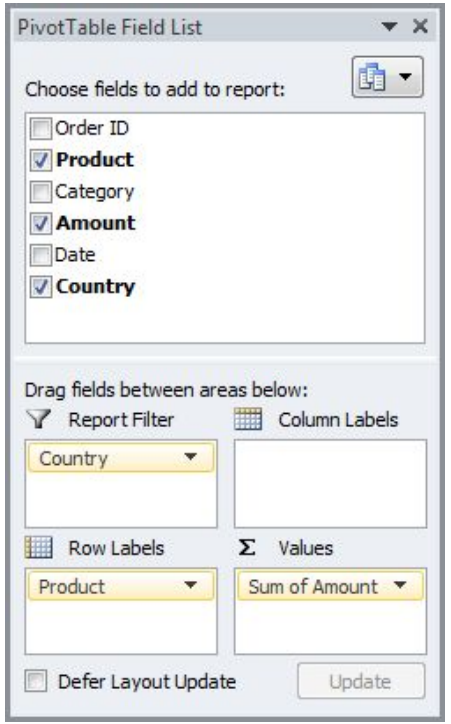
Drag fields

The PivotTable field list appears. To get the total amount exported of each product, drag the following fields to the different areas.

1. Product Field to the Row Labels area.

2. Amount Field to the Values area.

3. Country Field to the Report Filter area.



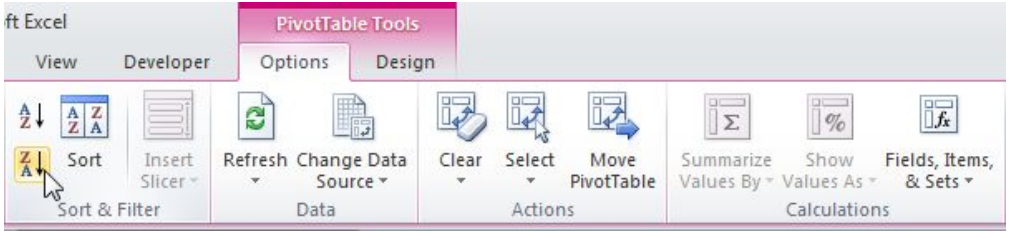
Below you can find the pivot table. Bananas are our main export product. That's how easy pivot tables can be!

A3 | Sum of Amount

	A	B	C	D
1	Country	(All)		
2				
3	Sum of Amount			
4	Product	Total		
5	Apple	191257		
6	Banana	340295		
7	Beans	57281		
8	Broccoli	142439		
9	Carrots	136945		
10	Mango	57079		
11	Orange	104438		
12	Grand Total	1029734		
13				
14				

Sort

To get Banana at the top of the list, sort the pivot table.



1. Click any cell inside the Total column.

2. The PivotTable Tools contextual tab activates. On the Options tab, click the Sort Largest to Smallest button (ZA).

Result.

B5 fx 340295

	A	B	C	D
1	Country	(All)		
2				
3	Sum of Amount			
4	Product	Total		
5	Banana	340295		
6	Apple	191257		
7	Broccoli	142439		
8	Carrots	136945		
9	Orange	104438		
10	Beans	57281		
11	Mango	57079		
12	Grand Total	1029734		
13				
14				

Filter

Because we added the Country field to the Report Filter area, we can filter this pivot table by Country. For example, which products do we export the most to France?

1. Click the filter drop-down and select France.

Result. Apples are our main export product to France.

B5 fx 80193

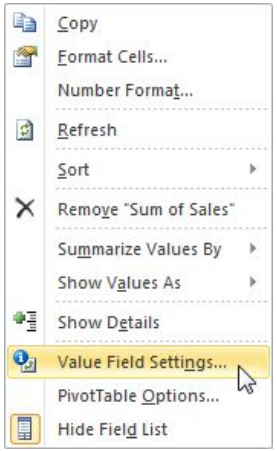
	A	B	C	D
1	Country	France		
2				
3	Sum of Amount			
4	Product	Total		
5	Apple	80193		
6	Banana	36094		
7	Carrots	9104		
8	Mango	7388		
9	Broccoli	5341		
10	Orange	2256		
11	Beans	680		
12	Grand Total	141056		
13				
14				

Change Summary Calculation

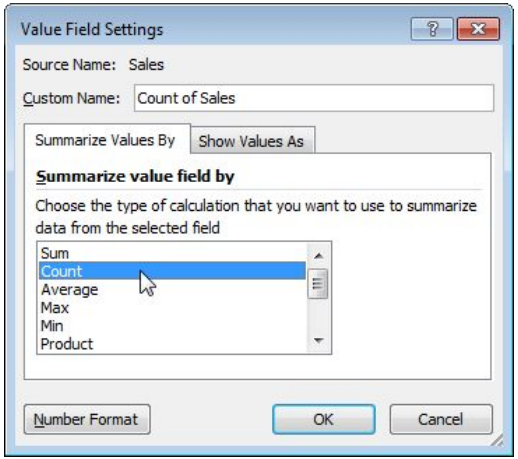
By default, Excel summarizes your data by either summing or counting the items. To change the type of calculation that you want to use, execute the following steps.

1. Click any cell inside the Total column.

2. Right click and click on Value Field Settings...

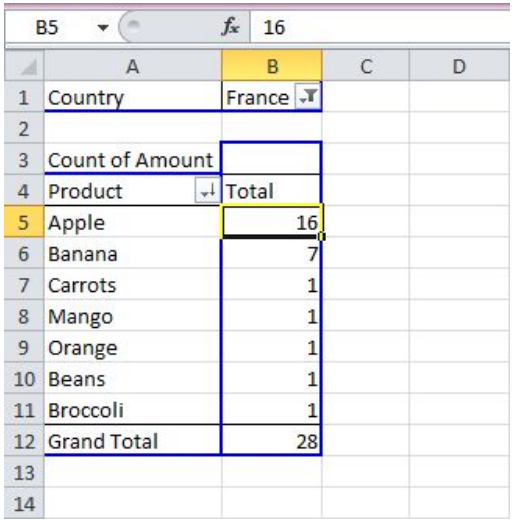


3. Choose the type of calculation you want to use. For example, click Count.



4. Click OK.

Result. 16 out of the 28 orders to France were 'Apple' orders.



	A	B	C	D
1	Country	France		
2				
3	Count of Amount			
4	Product	Total		
5	Apple	16		
6	Banana	7		
7	Carrots	1		
8	Mango	1		
9	Orange	1		
10	Beans	1		
11	Broccoli	1		
12	Grand Total	28		
13				
14				

Two-dimensional Pivot Table

If you drag a field to the Row Labels area and Column Labels area, you can create a two-dimensional pivot table.

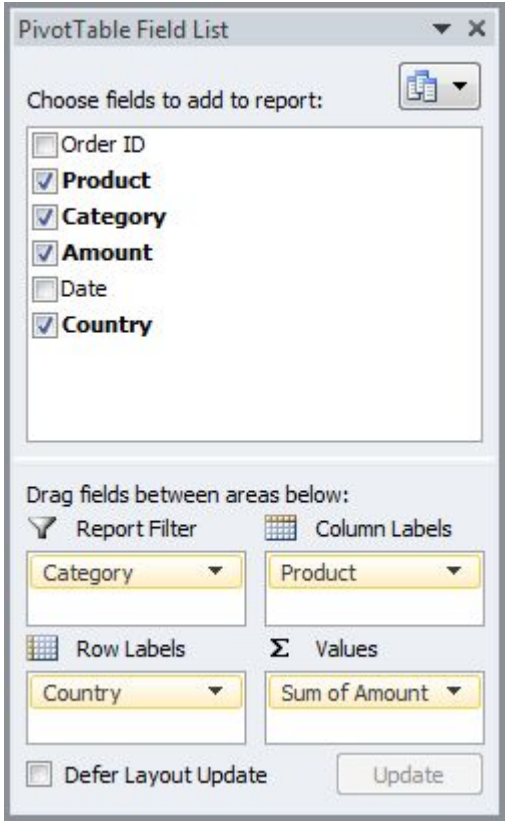
For example, to get the total amount exported to each country, of each product, drag the following fields to the different areas.

1. Country Field to the Row Labels area.

2. Product Field to the Column Labels area.

3. Amount Field to the Values area.

4. Category Field to the Report Filter area.

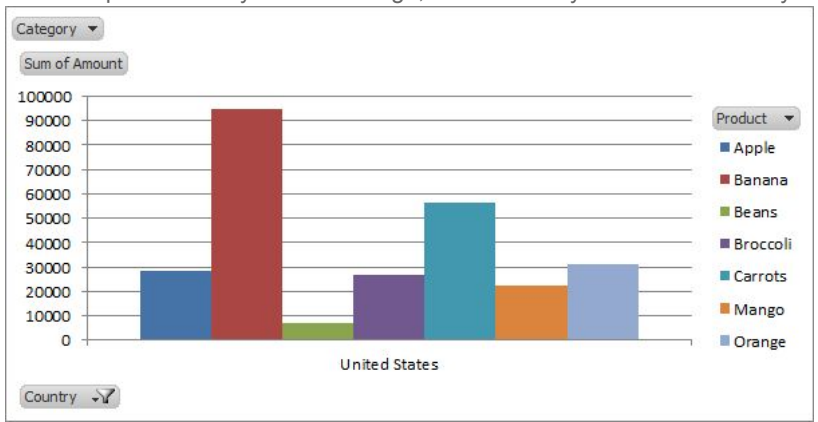


Below you can find the two-dimensional pivot table.

A3 | Sum of Amount

	A	B	C	D	E	F	G	H	I
1	Category	(All)							
2									
3	Sum of Amount	Product							
4	Country	Apple	Banana	Beans	Broccoli	Carrots	Mango	Orange	Grand Total
5	Australia	20634	52721	14433	17953	8106	9186	8680	131713
6	Canada	24867	33775		12407		3767	19929	94745
7	France	80193	36094	680	5341	9104	7388	2256	141056
8	Germany	9082	39686	29905	37197	21636	8775	8887	155168
9	New Zealand	10332	40050		4390			12010	66782
10	United Kingdom	17534	42908	5100	38436	41815	5600	21744	173137
11	United States	28615	95061	7163	26715	56284	22363	30932	267133
12	Grand Total	191257	340295	57281	142439	136945	57079	104438	1029734
13									
14									

To easily compare these numbers, create a pivot chart and apply a filter. Maybe this is one step too far for you at this stage, but it shows you one of the many other powerful pivot table features Excel has to offer.



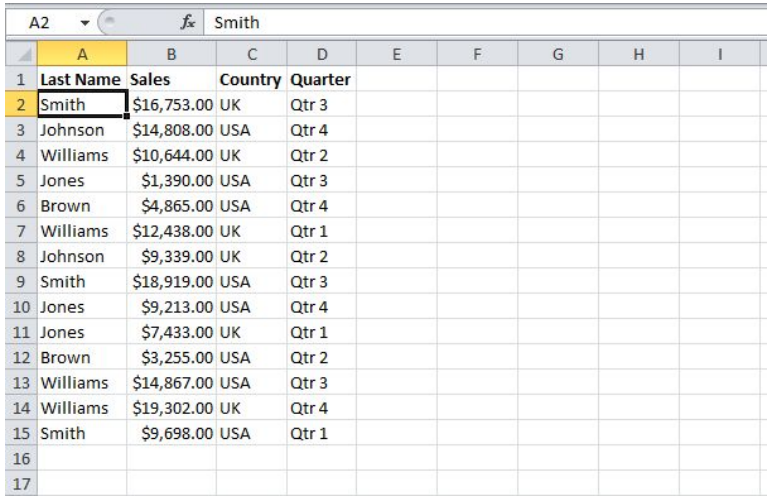
5. Tables

Tables allow you to analyze your data in Excel quickly and easily. Learn how to insert, sort and filter a table, and how to display a total row at the end of a table.

Insert a Table

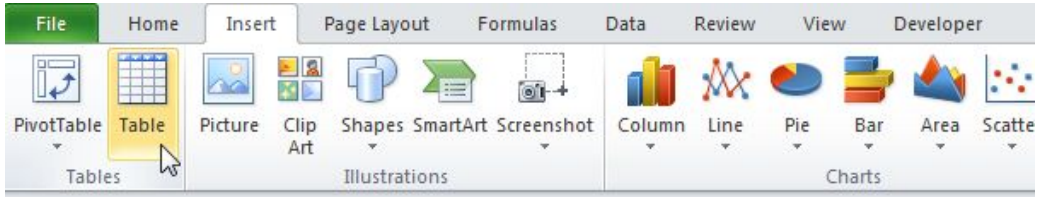
To insert a table, execute the following steps.

1. Click any single cell inside the data set.

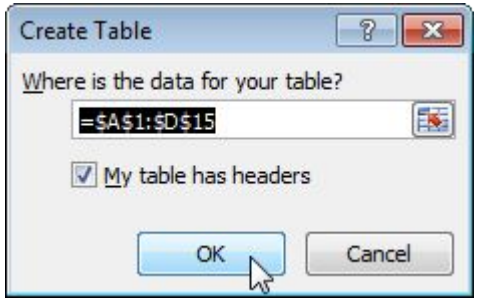


	A	B	C	D	E	F	G	H	I
1	**Last Name**	**Sales**	**Country**	**Quarter**					
2	Smith	$16,753.00	UK	Qtr 3					
3	Johnson	$14,808.00	USA	Qtr 4					
4	Williams	$10,644.00	UK	Qtr 2					
5	Jones	$1,390.00	USA	Qtr 3					
6	Brown	$4,865.00	USA	Qtr 4					
7	Williams	$12,438.00	UK	Qtr 1					
8	Johnson	$9,339.00	UK	Qtr 2					
9	Smith	$18,919.00	USA	Qtr 3					
10	Jones	$9,213.00	USA	Qtr 4					
11	Jones	$7,433.00	UK	Qtr 1					
12	Brown	$3,255.00	USA	Qtr 2					
13	Williams	$14,867.00	USA	Qtr 3					
14	Williams	$19,302.00	UK	Qtr 4					
15	Smith	$9,698.00	USA	Qtr 1					
16									
17									

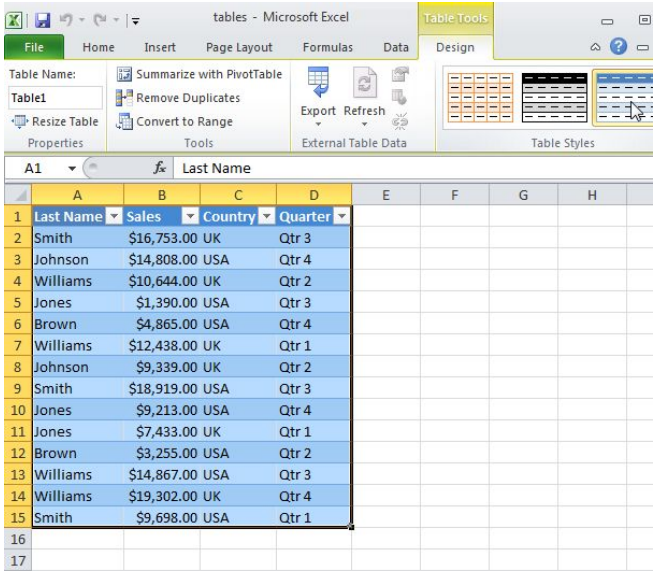
2. On the Insert tab, click Table.



3. Excel automatically selects the data for you. Check 'My table has headers' and click on OK.



Result. Excel creates a nicely formatted table for you. This may still seem like a normal data range to you but many powerful features are now just a click of a button away.



Note: the Table Tools contextual tab (with the underlying Design tab selected) is the starting point for working with tables. If at any time you lose this tab, simply click any cell within the table and it will activate again. Choose a table style you like. Hover over a table style and Excel gives you a life preview.

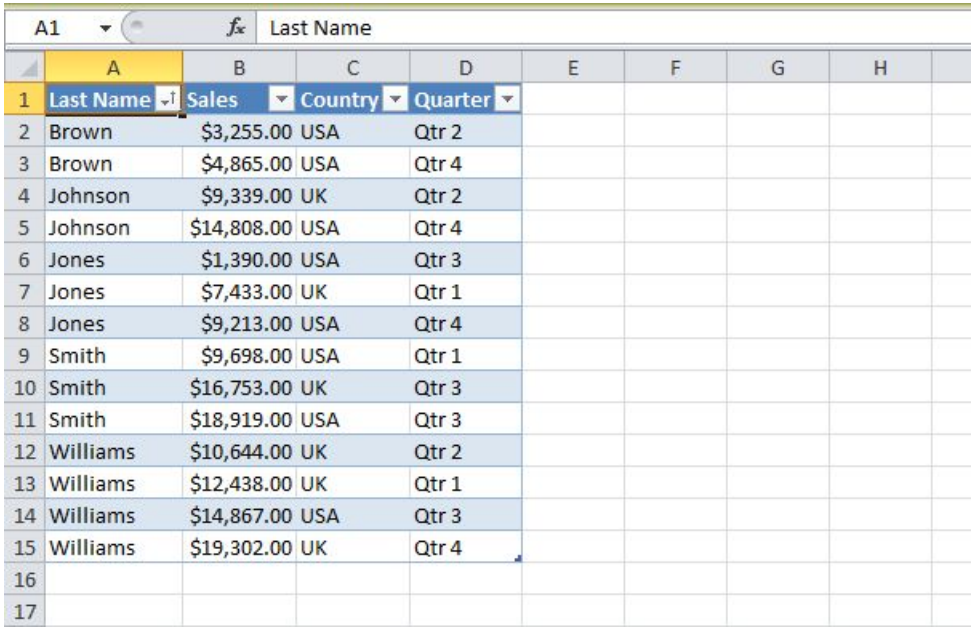
Sort a Table

To sort by Last Name first and Sales second, first sort by Sales, next sort by Last Name (the exact opposite).

1. Click the arrow next to Sales and click Sort Smallest to Largest.

2. Click the arrow next to Last Name and click Sort A to Z.

Result.



A1 | Last Name

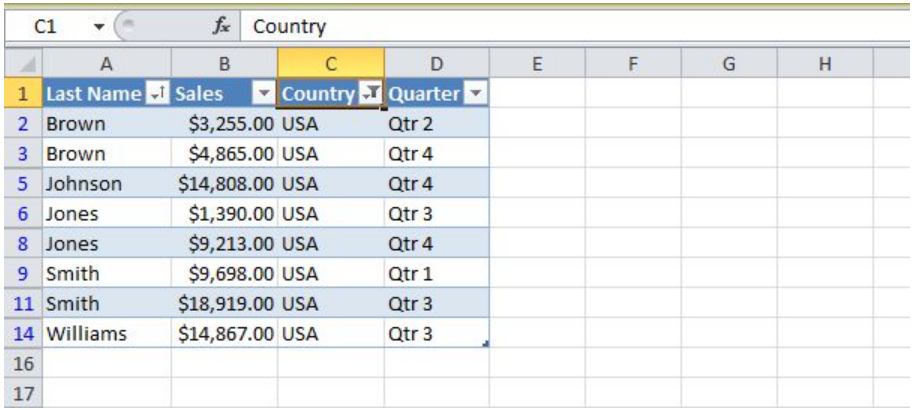
	A	B	C	D
1	Last Name	Sales	Country	Quarter
2	Brown	$3,255.00	USA	Qtr 2
3	Brown	$4,865.00	USA	Qtr 4
4	Johnson	$9,339.00	UK	Qtr 2
5	Johnson	$14,808.00	USA	Qtr 4
6	Jones	$1,390.00	USA	Qtr 3
7	Jones	$7,433.00	UK	Qtr 1
8	Jones	$9,213.00	USA	Qtr 4
9	Smith	$9,698.00	USA	Qtr 1
10	Smith	$16,753.00	UK	Qtr 3
11	Smith	$18,919.00	USA	Qtr 3
12	Williams	$10,644.00	UK	Qtr 2
13	Williams	$12,438.00	UK	Qtr 1
14	Williams	$14,867.00	USA	Qtr 3
15	Williams	$19,302.00	UK	Qtr 4
16				
17				

Filter a Table

To filter a table, execute the following steps.

1. Click the arrow next to Country and only check USA.

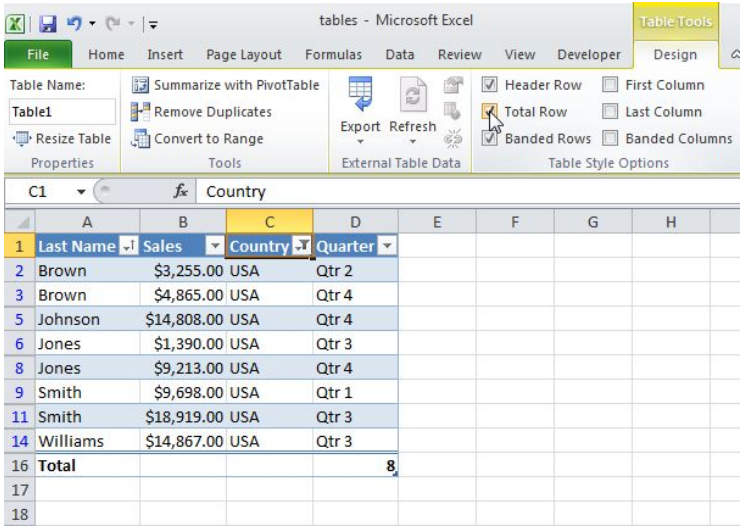
Result.



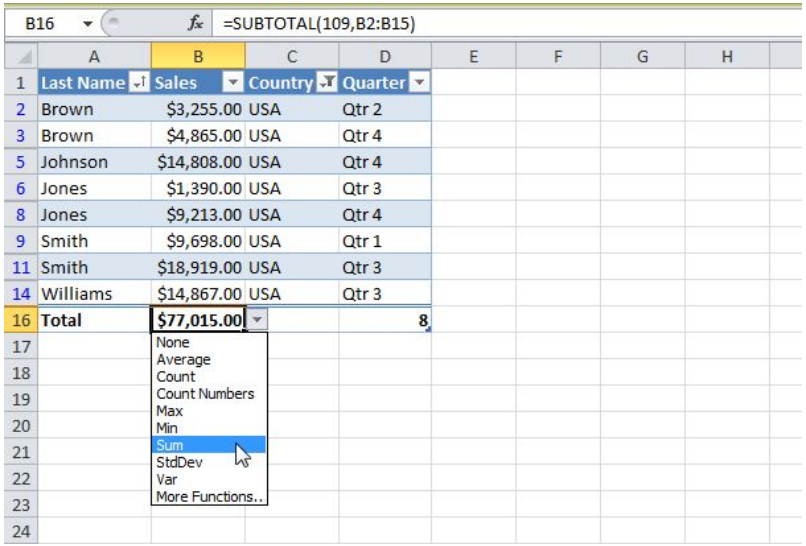
Total Row

To display a total row at the end of the table, execute the following steps.

1. On the Design tab, in the Table Style Options group, check Total Row.



2. Click any cell in the last row to calculate the Total (Average, Count, Max, Min, Sum etc.) of a column. For example, calculate the sum of the Sales column.

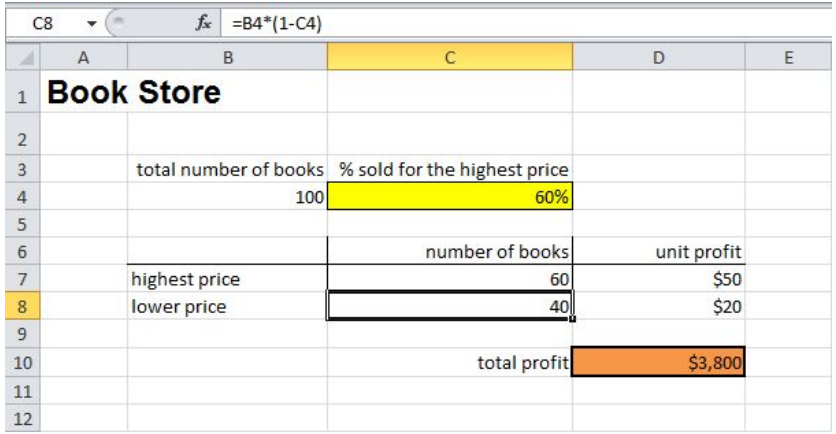


Note: in the formula bar see how Excel uses the SUBTOTAL function to calculate the sum. 109 is the argument for Sum if you use the SUBTOTAL function. Excel uses this function (and not the standard SUM function) to correctly calculate table totals of filtered tables

6. What If Analysis

What-If Analysis in Excel allows you to try out different values (scenarios) for formulas. The following example helps you master what-if analysis quickly and easily.

Assume you own a book store and have 100 books in storage. You sell a certain % for the highest price of $50 and a certain % for the lower price of $20.



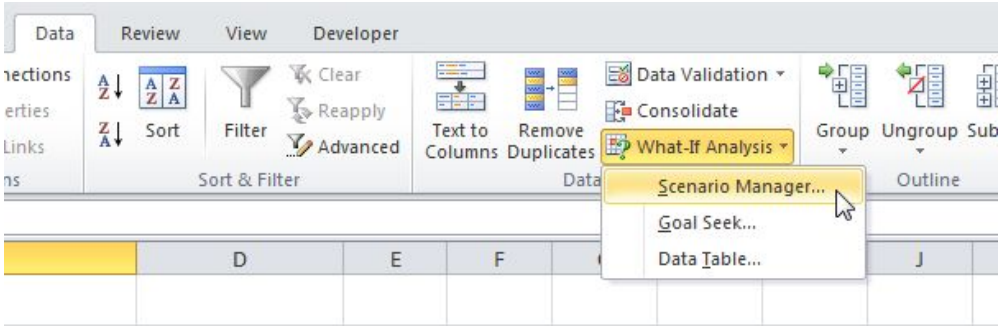
If you sell 60% for the highest price, cell D10 calculates a total profit of 60 * $50 + 40 * $20 = $3800.

Create Different Scenarios

But what if you sell 70% for the highest price? And what if you sell 80% for the highest price? Or 90%, or even 100%? Each different percentage is a different scenario. You can use the Scenario Manager to create these scenarios.

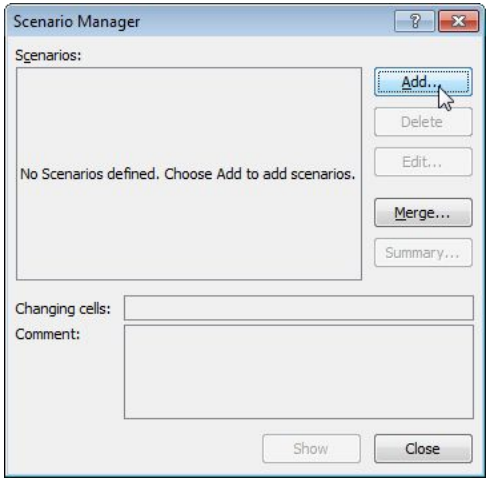
Note: You can simply type in a different percentage into cell C4 to see the corresponding result of a scenario in cell D10. However, what-if analysis enables you to easily compare the results of different scenarios. Read on.

1. On the Data tab, click What-If Analysis and select Scenario Manager from the list.

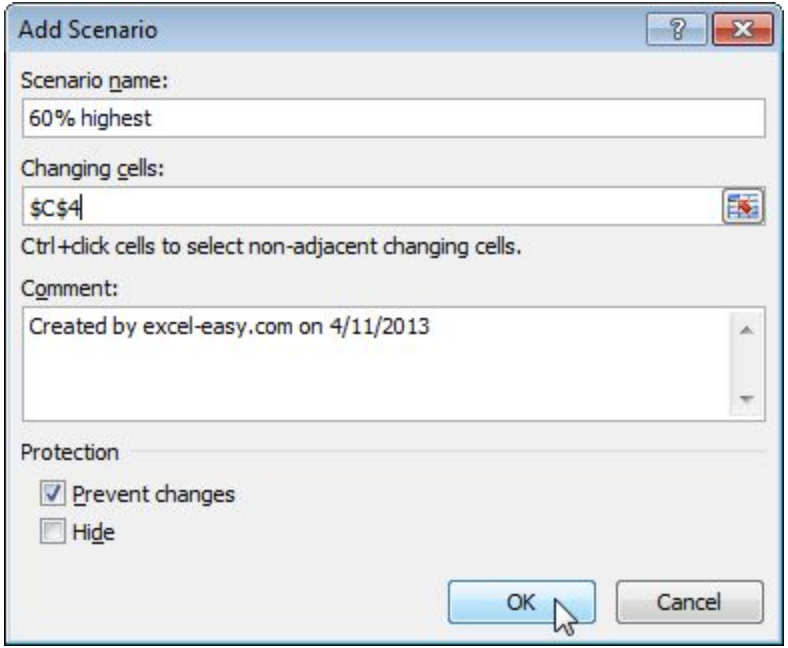


The Scenario Manager dialog box appears.

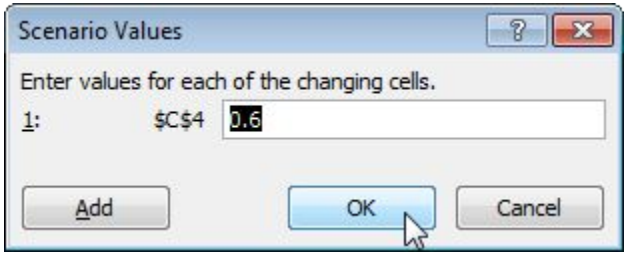
2. Add a scenario by clicking on Add.



3. Type a name (60% highest), select cell C4 (% sold for the highest price) for the Changing cells and click on OK.

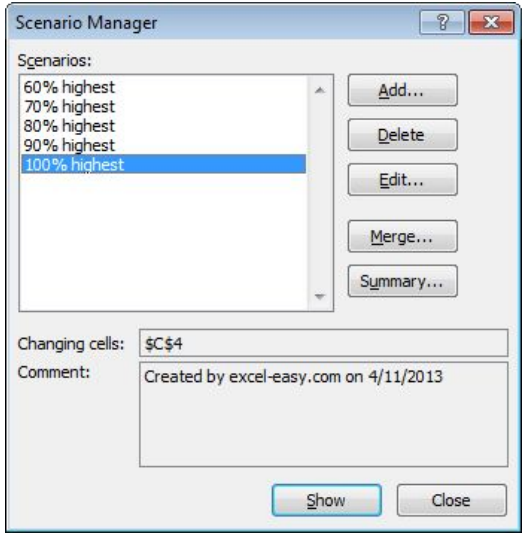


4. Enter the corresponding value 0.6 and click on OK again.



5. Next, add 4 other scenarios (70%, 80%, 90% and 100%).

Finally, your Scenario Manager should be consistent with the picture below:



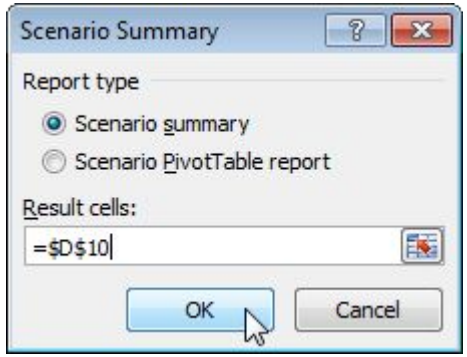
Note: to see the result of a scenario, select the scenario and click on the Show button. Excel will change the value of cell C4 accordingly for you to see the corresponding result on the sheet.

Scenario Summary

To easily compare the results of these scenarios, execute the following steps.

1. Click the Summary button in the Scenario Manager.

2. Next, select cell D10 (total profit) for the result cell and click on OK.



Result:

Scenario Summary						
	Current Values:	60% highest	70% highest	80% highest	90% highest	100% highest
Changing Cells:						
$C$4	60%	60%	70%	80%	90%	100%
Result Cells:						
$D$10	$3,800	$3,800	$4,100	$4,400	$4,700	$5,000

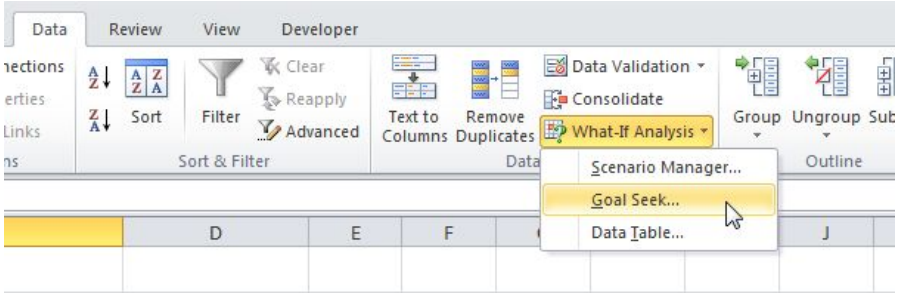
Notes: Current Values column represents values of changing cells at time Scenario Summary Report was created. Changing cells for each scenario are highlighted in gray.

Conclusion: if you sell 70% for the highest price, you obtain a total profit of $4100, if you sell 80% for the highest price, you obtain a total profit of $4400, etc. That's how easy what-if analysis in Excel can be.

Goal Seek

What if you want to know how many books you need to sell for the highest price, to obtain a total profit of exactly $4700? You can use Excel's Goal Seek feature to find the answer.

1. On the Data tab, click What-If Analysis, Goal Seek.



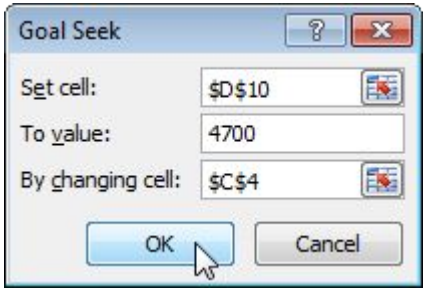
The Goal Seek dialog box appears.

2. Select cell D10.

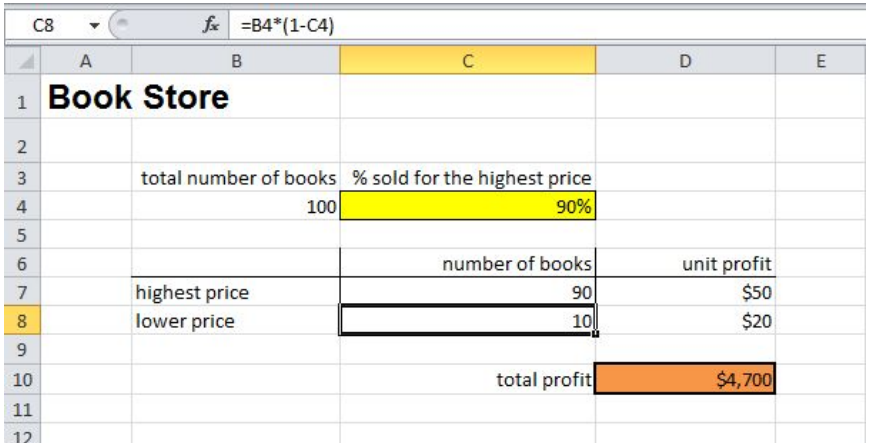
3. Click in the 'To value' box and type 4700.

4. Click in the 'By changing cell' box and select cell C4.

5. Click OK.



Result. You need to sell 90% of the books for the highest price to obtain a total profit of exactly $4700.



7. Solver in Excel

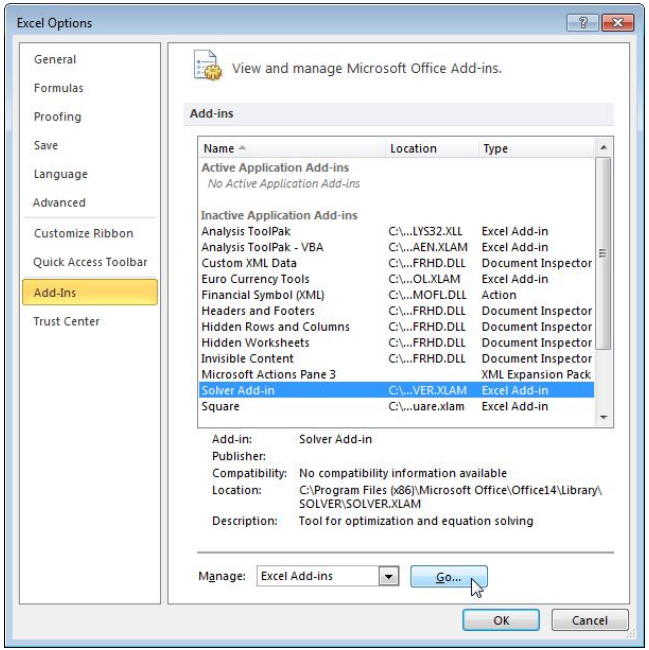
Excel includes a tool called solver that uses techniques from the operations research to find optimal solutions for all kind of decision problems.

Load the Solver Add-in

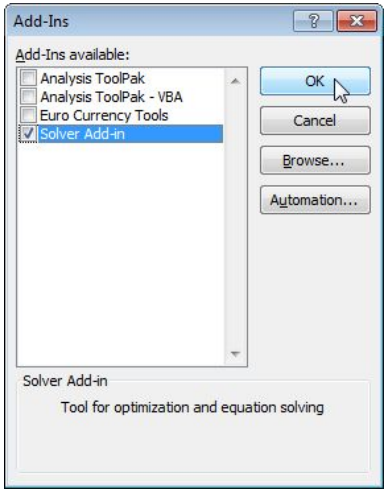
To load the solver add-in, execute the following steps.

1. On the green File tab, click Options.

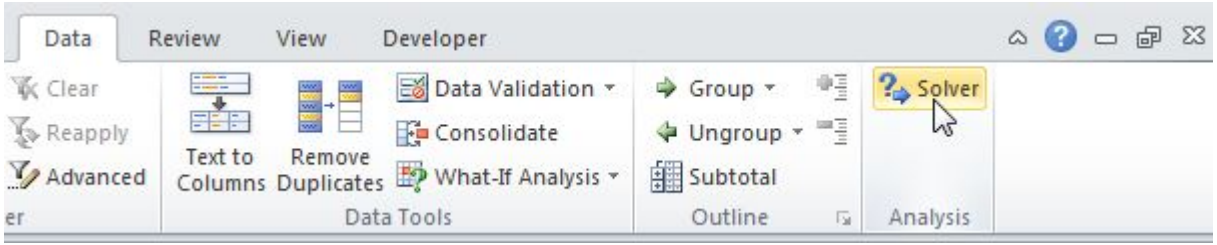
2. Under Add-ins, select Solver Add-in and click on the Go button.



3. Check Solver Add-in and click OK.



4. You can find the Solver on the Data tab.



Formulate the Model

The model we are going to solve looks as follows in Excel.

	A	B	C	D	E	F	G	H	I	J
1		Cycle Trader								
2										
3			Bicycles	Mopeds	Child Seats					
4		Unit Profit	100	300	50					
5							Resources		Resources	
6							Used		Available	
7		Capital	300	1200	120		0	≤	93000	
8		Storage	0.5	1	0.5		0	≤	101	
9										
10										
11			Bicycles	Mopeds	Child Seats				Total Profit	
12		Order Size	0	0	0				0	
13										
14										

Range Name	Cells
UnitProfit	C4:E4
OrderSize	C12:E12
ResourcesUsed	G7:G8
ResourcesAvailable	I7:I8
TotalProfit	I12

1. To formulate this linear programming model, answer the following three questions.

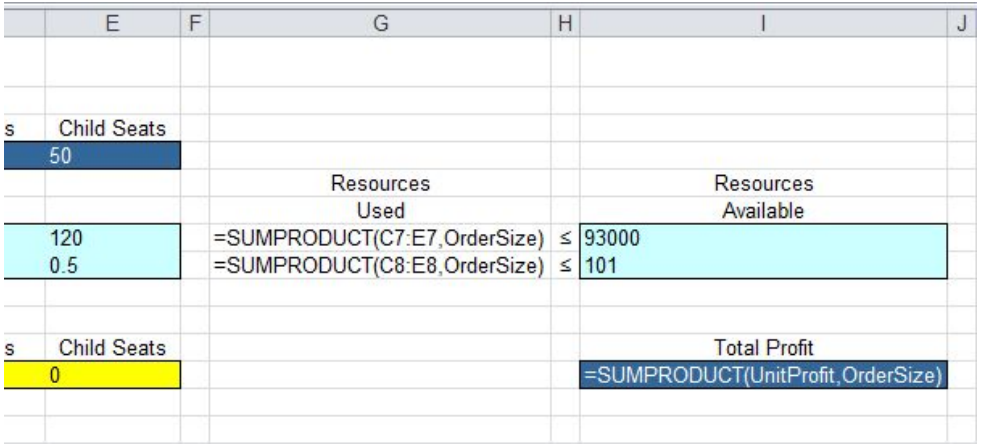
a. What are the decisions to be made? For this problem, we need Excel to find out how much to order of each product (bicycles, mopeds and child seats).

b. What are the constraints on these decisions? The constrains here are that the amount of capital and storage used by the products cannot exceed the limited amount of capital and storage (resources) available. For example, each bicycle uses 300 units of capital and 0.5 unit of storage.

c. What is the overall measure of performance for these decisions? The overall measure of performance is the total profit of the three products, so the objective is to maximize this quantity.

2. To make the model easier to understand, name the following ranges.

3. Insert the following three SUMPRODUCT functions.



	E	F	G	H	I	J
s	Child Seats					
	50					
			Resources Used		Resources Available	
	120		=SUMPRODUCT(C7:E7,OrderSize)	≤	93000	
	0.5		=SUMPRODUCT(C8:E8,OrderSize)	≤	101	
s	Child Seats				Total Profit	
	0				=SUMPRODUCT(UnitProfit,OrderSize)	

Explanation: The amount of capital used equals the sumproduct of the range C7:E7 and OrderSize. The amount of storage used equals the sumproduct of the range C8:E8 and OrderSize. Total Profit equals the sumproduct of UnitProfit and OrderSize.

Trial and Error

With this formulation, it becomes easy to analyze any trial solution.

For example, if we order 20 bicycles, 40 mopeds and 100 child seats, the total amount of resources used does not exceed the amount of resources available. This solution has a total profit of 19000.

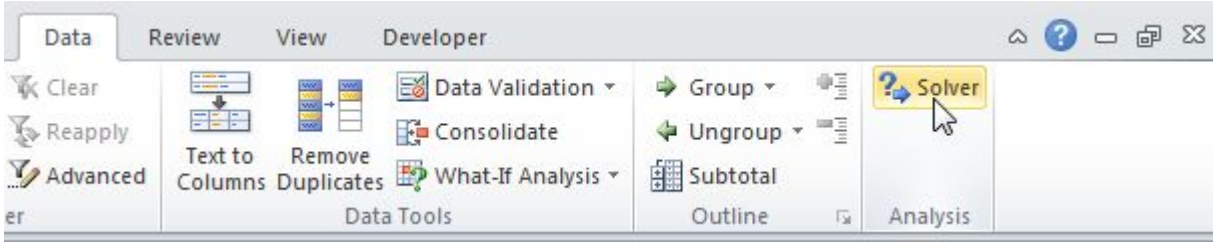
	A	B	C	D	E	F	G	H	I	J
1	Cycle Trader									
2										
3			Bicycles	Mopeds	Child Seats					
4		Unit Profit	100	300	50					
5							Resources		Resources	
6							Used		Available	
7		Capital	300	1200	120		66000	≤	93000	
8		Storage	0.5	1	0.5		100	≤	101	
9										
10										
11			Bicycles	Mopeds	Child Seats				Total Profit	
12		Order Size	20	40	100				19000	
13										
14										

It is not necessary to use trial and error. We shall describe next how the Excel Solver can be used to quickly find the optimal solution.

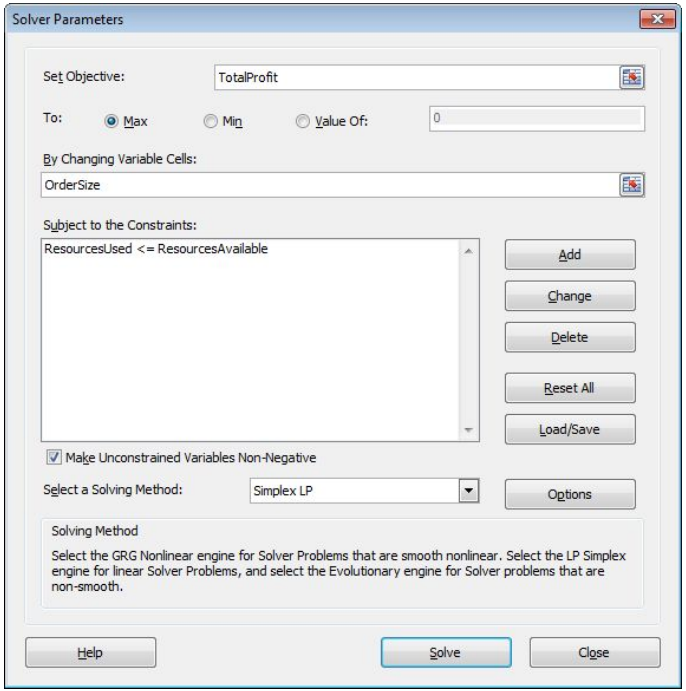
Solve the Model

To find the optimal solution, execute the following steps.

1. On the Data tab, click Solver.



Enter the solver parameters (read on). The result should be consistent with the picture below.



You have the choice of typing the range names or clicking on the cells in the spreadsheet.

2. Enter TotalProfit for the Objective.

3. Click Max.

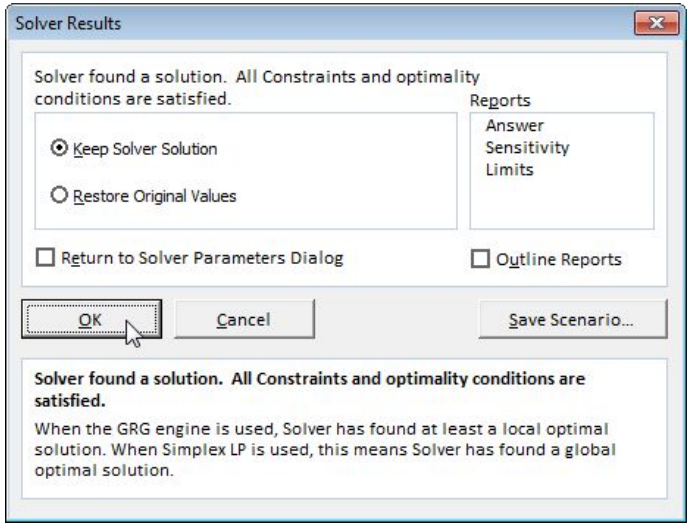
4. Enter OrderSize for the Changing Variable Cells.

5. Click Add to enter the following constraint.

6. Check 'Make Unconstrained Variables Non-Negative' and select 'Simplex LP'.

7. Finally, click Solve.

Result:



The optimal solution:

	A	B	C	D	E	F	G	H	I	J
1		Cycle Trader								
2										
3			Bicycles	Mopeds	Child Seats					
4		Unit Profit	100	300	50					
5							Resources Used		Resources Available	
6										
7		Capital	300	1200	120		93000	≤	93000	
8		Storage	0.5	1	0.5		101	≤	101	
9										
10										
11			Bicycles	Mopeds	Child Seats				Total Profit	
12		Order Size	94	54	0				25600	
13										
14										

Conclusion: it is optimal to order 94 bicycles and 54 mopeds.

This solution gives the maximum profit of 25600. This solution uses all the resources available.

7. Analysis ToolPak in Excel

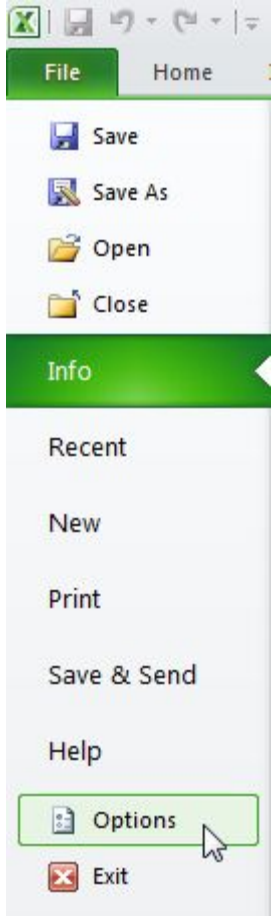
The Analysis ToolPak is an Excel add-in program that provides data analysis tools for financial, statistical and engineering data analysis.

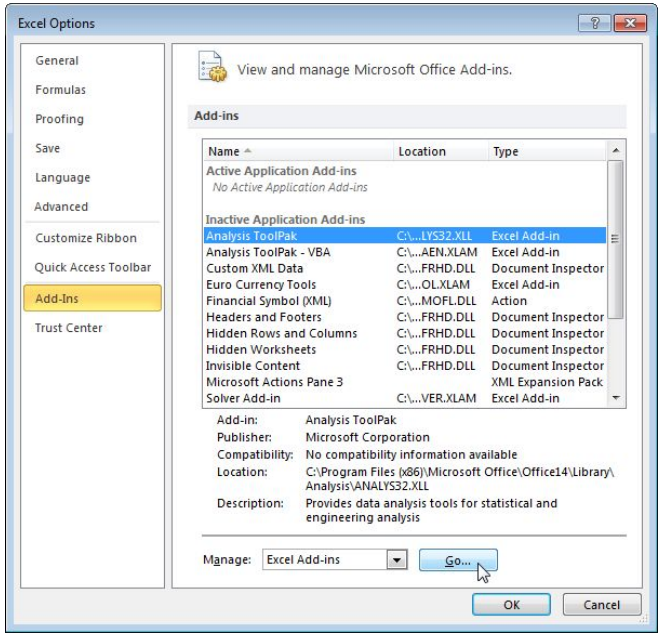
To load the Analysis ToolPak add-in, execute the following steps.

1. Click on the green File tab. The File tab in Excel 2010 replaces the Office Button (or File Menu) in previous versions of Excel.

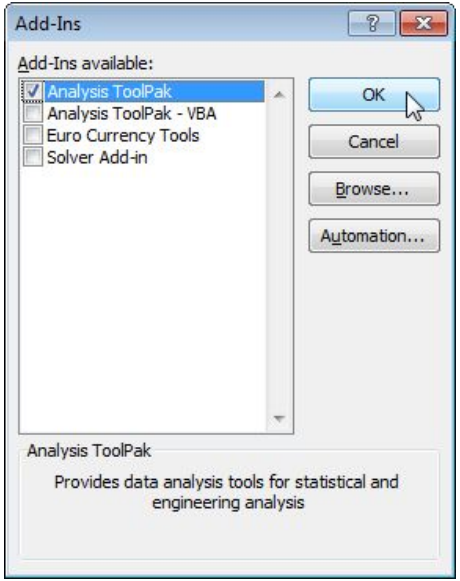
2. Click on Options.

3. Under Add-ins, select Analysis ToolPak and click on the Go button.

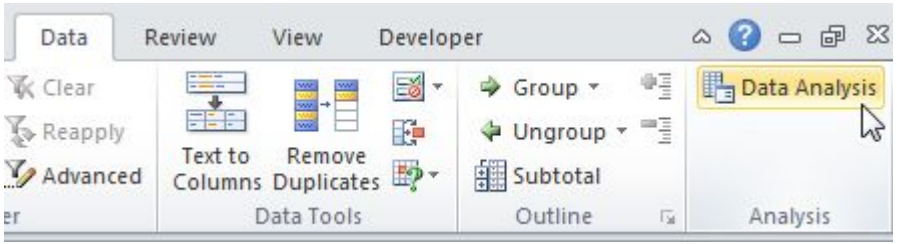




4. Check Analysis ToolPak and click on OK.

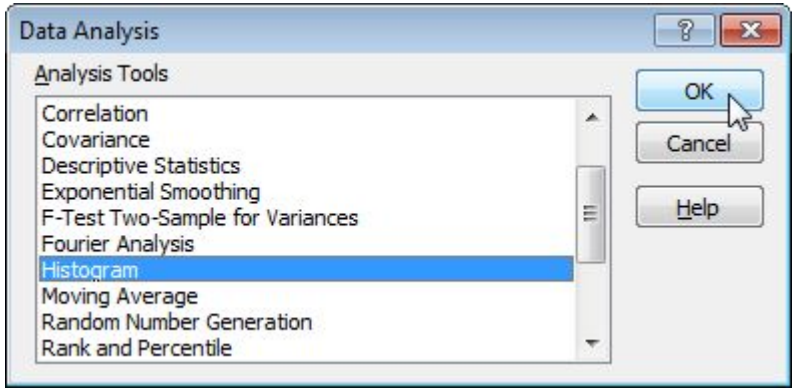


5. On the Data tab, you can now click on Data Analysis.



The following dialog box below appears.

6. For example, select Histogram and click OK to create a Histogram in Excel.



End

This is the end of the Intermediate Level Data Analysis in MS Excel

Chapter 3

Statistical Data Analysis in Excel

This section illustrates the statistical features Excel has to offer to analyze data.

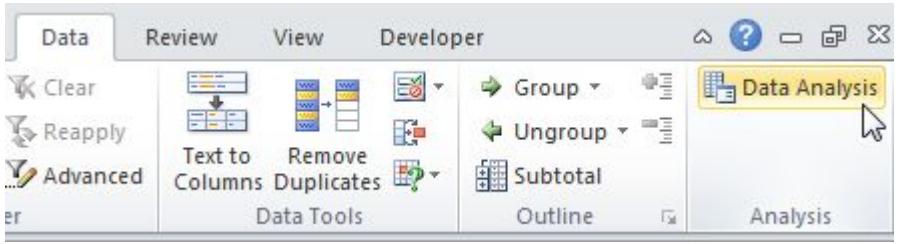
1. Histogram

This example teaches you how to create a histogram in Excel.

1. First, enter the bin numbers (upper levels) in the range C3:C7.

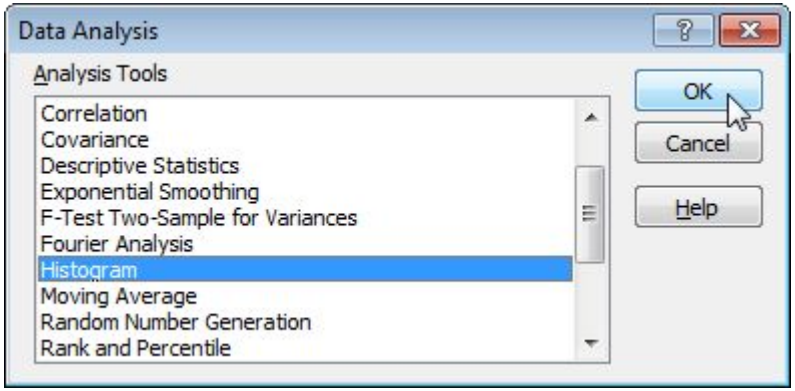
	A	B	C	D
1	Number of students			
2	22			
3	29		20	
4	40		25	
5	30		30	
6	48		35	
7	24		40	
8	21			
9	19			
10	24			
11	22			
12	25			
13	52			
14	35			
15	40			
16	31			
17	37			
18	21			
19	23			
20				
21				

2. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Load the Analysis ToolPak add-in.

3. Select Histogram and click OK

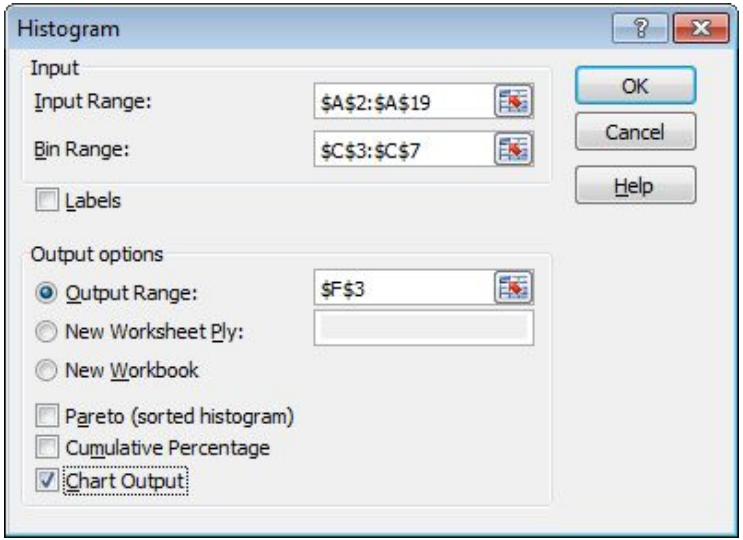


4. Select the range A2:A19.

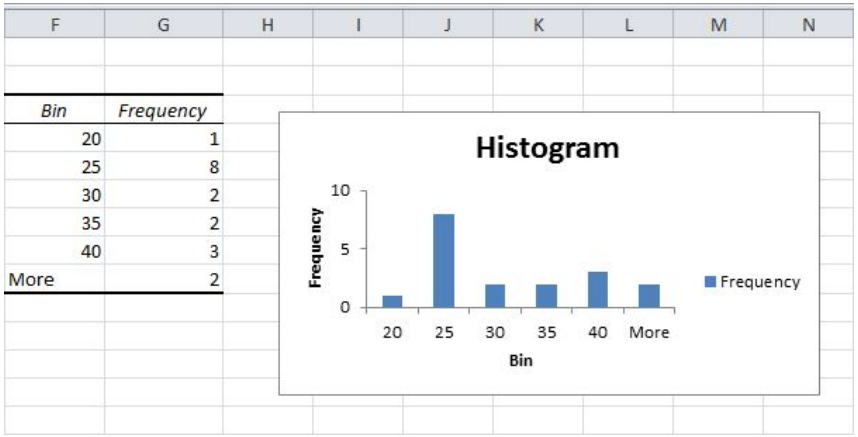
5. Click in the Bin Range box and select the range C3:C7.

6. Click the Output Range option button, click in the Output Range box and select cell F3.

7. Check Chart Output.



8. Click OK.

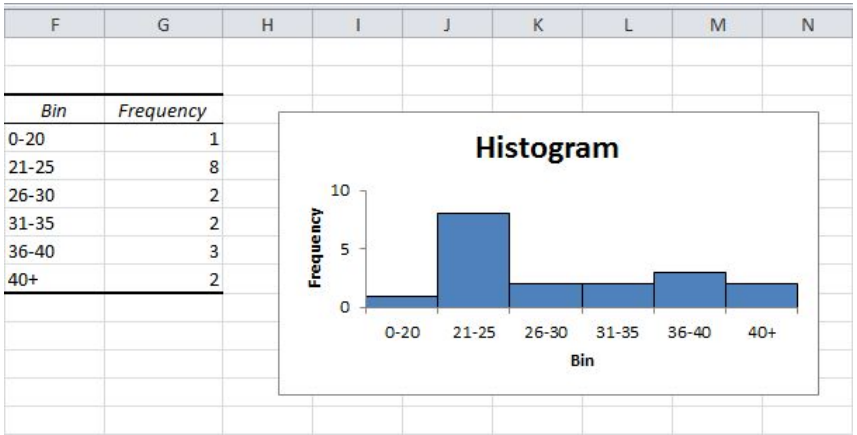


9. Click the legend on the right side and press Delete.

10. Properly label your bins.

11. To remove the space between the bars, right click a bar, select Format Data Series and change the Gap Width to 0%. Select Border Color to add a border.

Result:



2. Descriptive Statistics

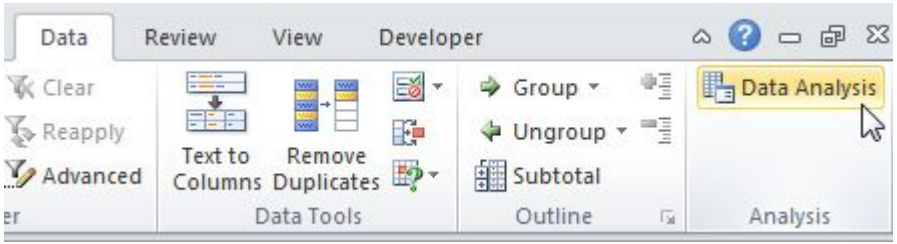
You can use the Analysis Toolpak add-in to generate descriptive statistics. For example, you may have the scores of 14 participants for a test.

M26

	A	B
1	Scores	
2	82	
3	93	
4	91	
5	69	
6	96	
7	61	
8	88	
9	58	
10	59	
11	100	
12	93	
13	71	
14	78	
15	98	
16		
17		

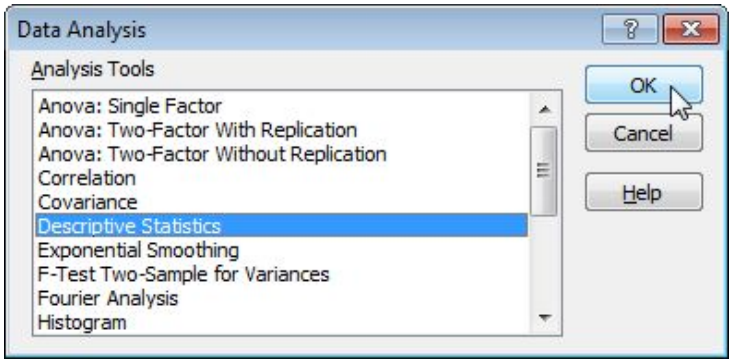
To generate descriptive statistics for these scores, execute the following steps.

1. On the Data tab, click Data Analysis.



Note: can't find the Data Analysis button? Click here to load the Analysis ToolPak add-in.

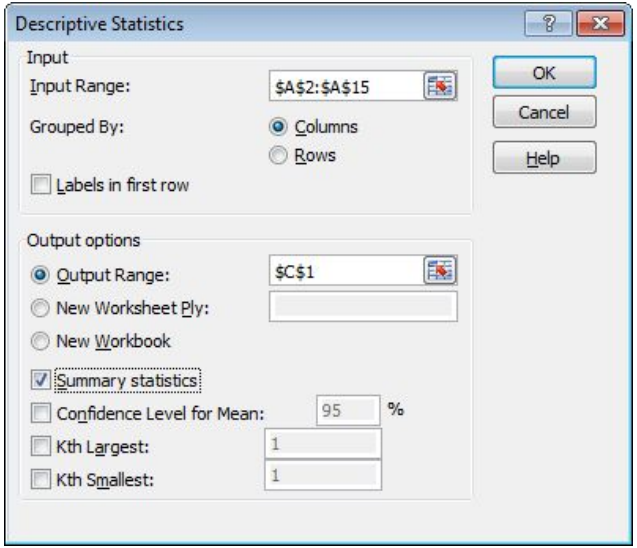
2. Select Descriptive Statistics and click OK.



3. Select the range A2:A15 as the Input Range.

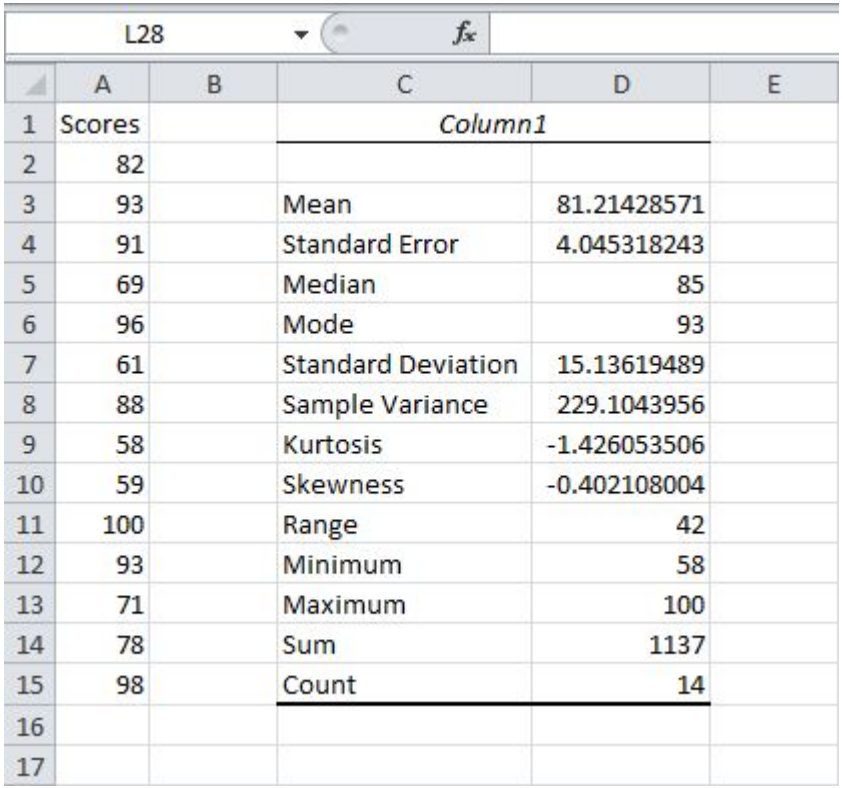
4. Select cell C1 as the Output Range.

5. Make sure Summary statistics is checked.



6. Click OK.

Result:



L28

	A	B	C	D	E
1	Scores		*Column1*		
2	82				
3	93		Mean	81.21428571	
4	91		Standard Error	4.045318243	
5	69		Median	85	
6	96		Mode	93	
7	61		Standard Deviation	15.13619489	
8	88		Sample Variance	229.1043956	
9	58		Kurtosis	-1.426053506	
10	59		Skewness	-0.402108004	
11	100		Range	42	
12	93		Minimum	58	
13	71		Maximum	100	
14	78		Sum	1137	
15	98		Count	14	
16					
17					

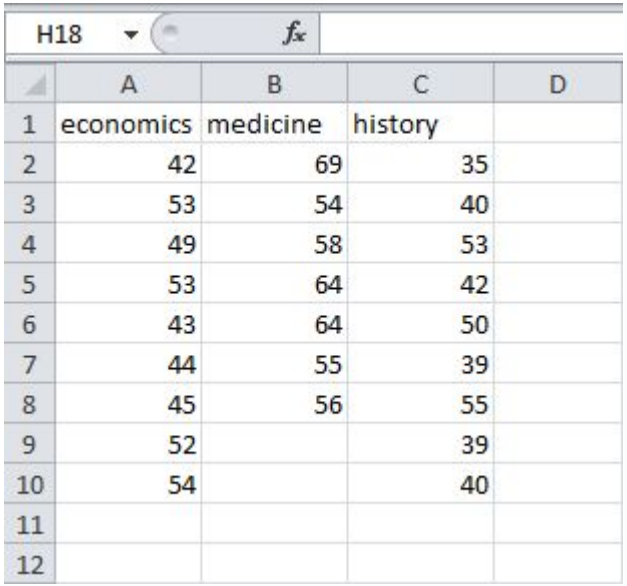
3. ANOVA

This example teaches you how to perform a single factor ANOVA (analysis of variance) in Excel. A single factor or one-way ANOVA is used to test the null hypothesis that the means of several populations are all equal.

Below you can find the salaries of people who have a degree in economics, medicine or history.

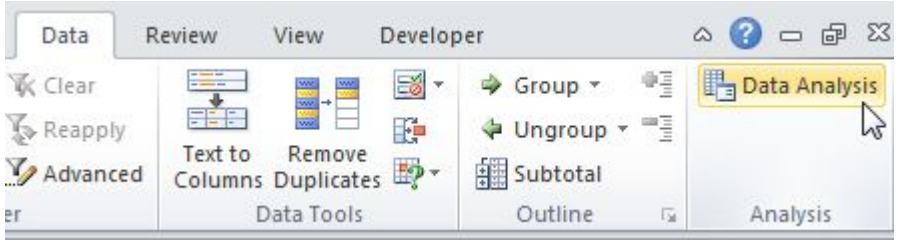
H_0: $\mu_1 = \mu_2 = \mu_3$

H_1: at least one of the means is different.

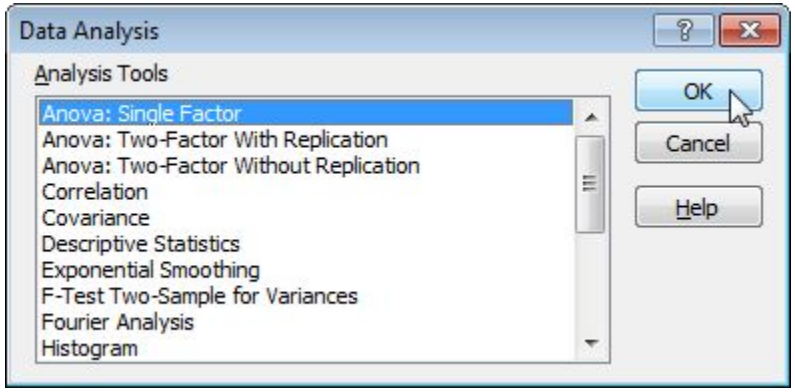


To perform a single factor ANOVA, execute the following steps.

1. On the Data tab, click Data Analysis.

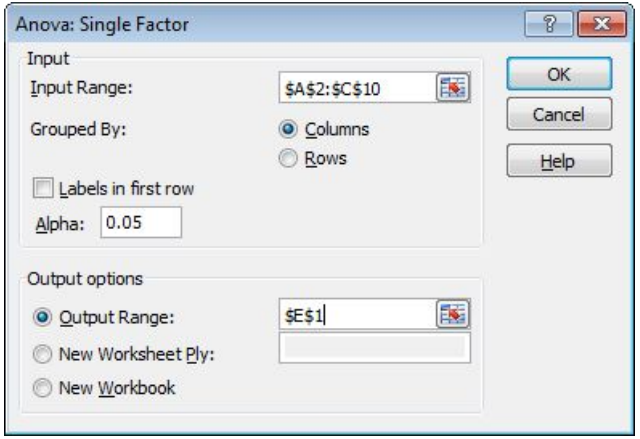


2. Select Anova: Single Factor and click OK.



3. Click in the Input Range box and select the range A2:C10.

4. Click in the Output Range box and select cell E1.



5. Click OK.

Result:

E	F	G	H	I	J	K
Anova: Single Factor						
SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Column 1	9	435	48.33333	23.5		
Column 2	7	420	60	32.33333		
Column 3	9	393	43.66667	50.5		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
Within Groups	786	22	35.72727			
Total	1871.84	24				

Conclusion: if F > F crit, we reject the null hypothesis. This is the case, 15.196 > 3.443. Therefore, we reject the null hypothesis. The means of the three populations are not all equal. At least one of the means is different. However, the ANOVA does not tell you where the difference lies. You need a t-Test to test each pair of means.

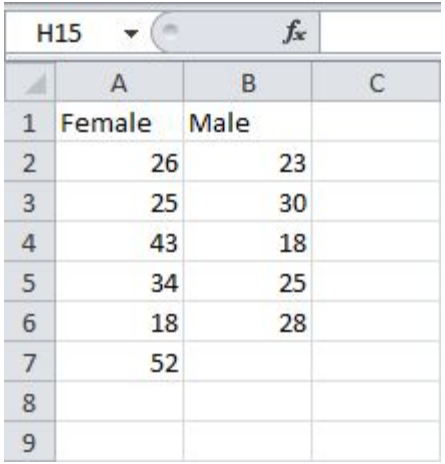
4. T-Test

This example teaches you how to perform a t-Test in Excel. The t-Test is used to test the null hypothesis that the means of two populations are equal.

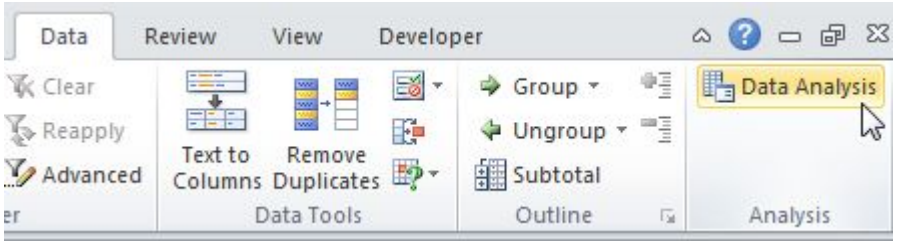
Below you can find the study hours of 6 female students and 5 male students.

H_0: $\mu_1 - \mu_2 = 0$

H_1: $\mu_1 - \mu_2 \neq 0$



	A	B	C
1	Female	Male	
2	26	23	
3	25	30	
4	43	18	
5	34	25	
6	18	28	
7	52		
8			
9			

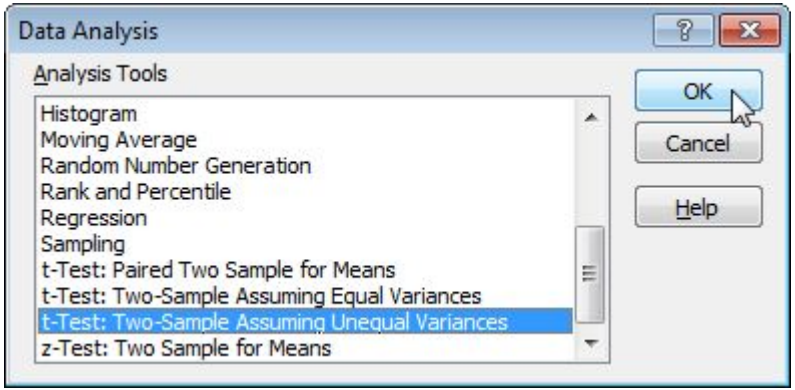


To perform a t-Test, execute the following steps.

1. First, perform an F-Test to determine if the variances of the two populations are equal. This is not the case.

2. On the Data tab, click Data Analysis.

3. Select t-Test: Two-Sample Assuming Unequal Variances and click OK.

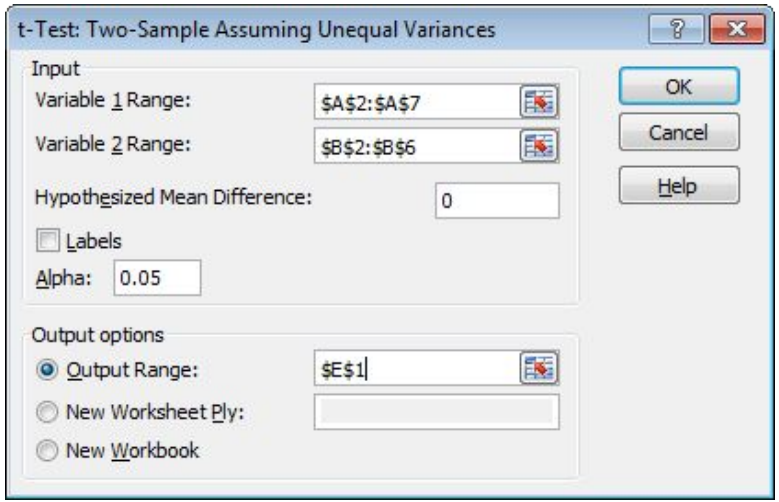


4. Click in the Variable 1 Range box and select the range A2:A7.

5. Click in the Variable 2 Range box and select the range B2:B6.

6. Click in the Hypothesized Mean Difference box and type 0 (H_0: $\mu_1 - \mu_2 = 0$).

7. Click in the Output Range box and select cell E1.



8. Click OK.

Result:

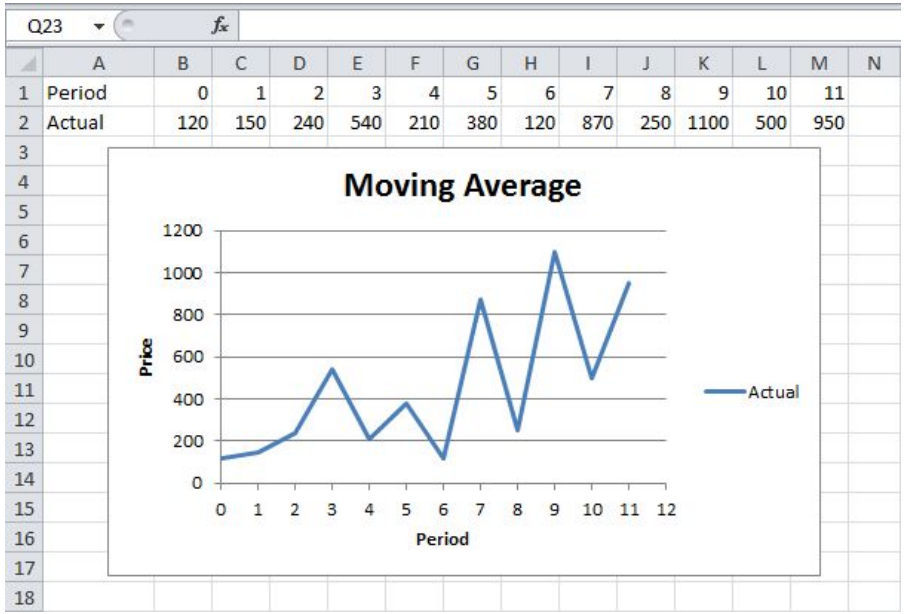
E	F	G
t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	33	24.8
Variance	160	21.7
Observations	6	5
Hypothesized Mean Difference	0	
df	7	
t Stat	1.47260514	
P(T<=t) one-tail	0.092170202	
t Critical one-tail	1.894578605	
P(T<=t) two-tail	0.184340405	
t Critical two-tail	2.364624252	

Conclusion: We do a two-tail test (inequality). If t Stat < -t Critical two-tail or t Stat > t Critical two-tail, we reject the null hypothesis. This is not the case, -2.365 < 1.473 < 2.365. Therefore, we do not reject the null hypothesis. The observed difference between the sample means (33 - 24.8) is not convincing enough to say that the average number of study hours between female and male students differ significantly.

4. Moving Average

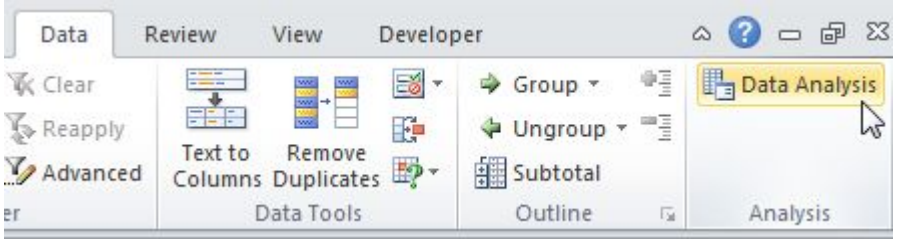
This example teaches you how to calculate the moving average of a time series in Excel. A moving avearge is used to smooth out irregularities (peaks and valleys) to easily recognize trends.

1. First, let's take a look at our time series.

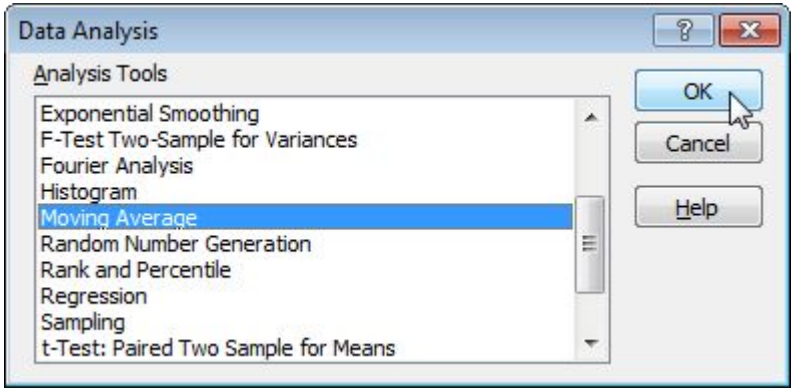


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Period	0	1	2	3	4	5	6	7	8	9	10	11	
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950	

2. On the Data tab, click Data Analysis.



3. Select Moving Average and click OK.

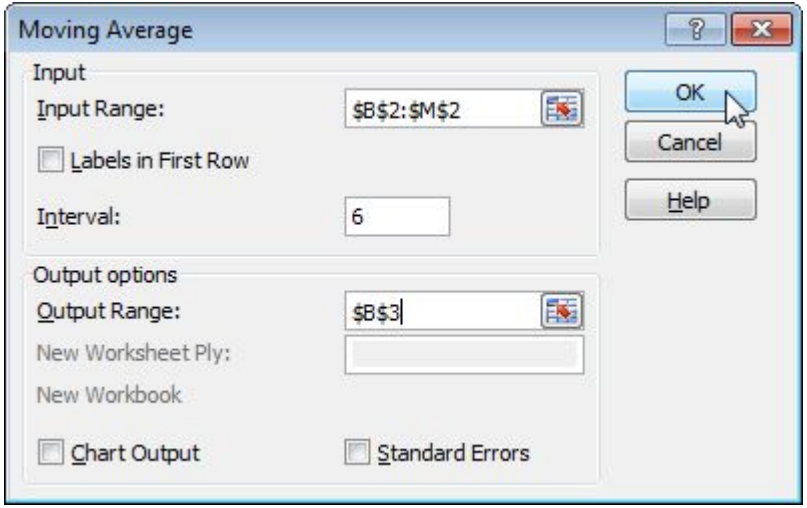


4. Click in the Input Range box and select the range B2:M2.

5. Click in the Interval box and type 6.

6. Click in the Output Range box and select cell B3.

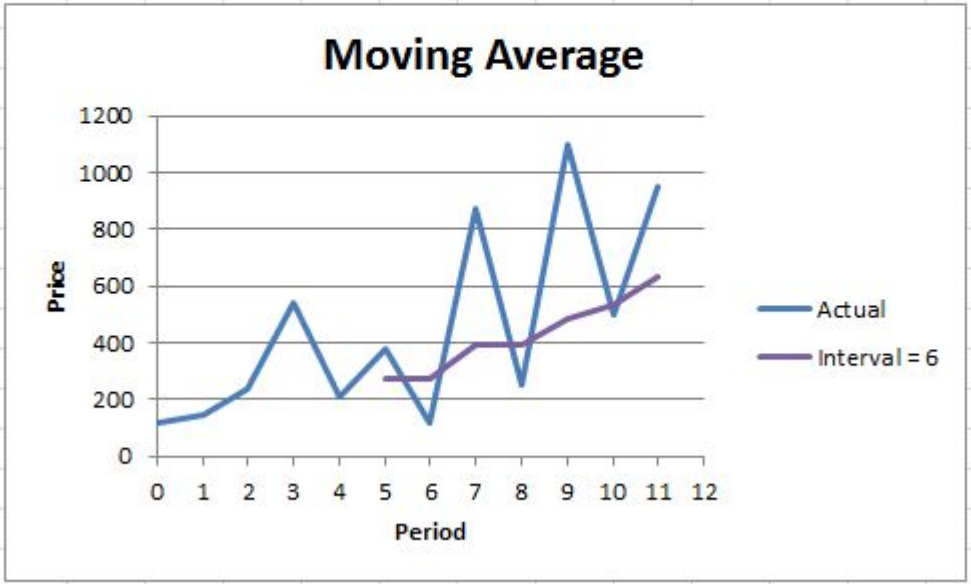
7. Click OK.



8. Plot a graph of these values.

COUN... =AVERAGE(B2:G2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Period	0	1	2	3	4	5	6	7	8	9	10	11	
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950	
3	Interval = 6	#N/A	#N/A	#N/A	#N/A	#N/A	=AVER	273	393	395	488	537	632	

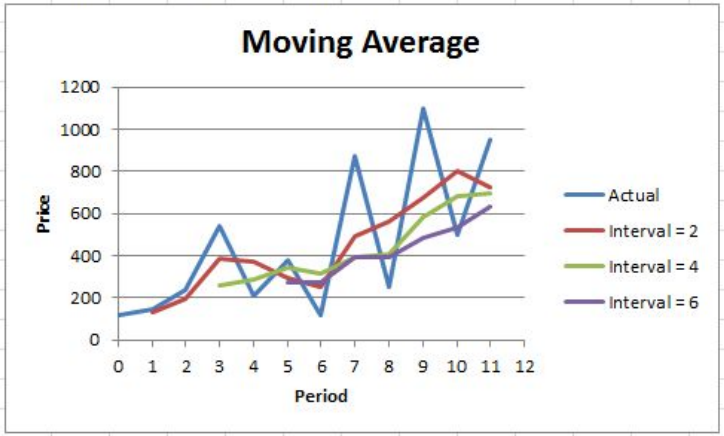


Explanation: because we set the interval to 6, the moving average is the average of the previous 5 data points and the current data point. As a result, peaks and valleys are smoothed out. The graph shows an increasing trend. Excel cannot calculate the moving average for the first 5 data points because there are not enough previous data points.

9. Repeat steps 2 to 8 for interval = 2 and interval = 4.

COUN... =AVERAGE(B2:C2)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Period	0	1	2	3	4	5	6	7	8	9	10	11	
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950	
3	Interval = 2	#N/A	=AVER	195	390	375	295	250	495	560	675	800	725	
4	Interval = 4	#N/A	#N/A	#N/A	263	285	343	313	395	405	585	680	700	
5	Interval = 6	#N/A	#N/A	#N/A	#N/A	#N/A	273	273	393	395	488	537	632	

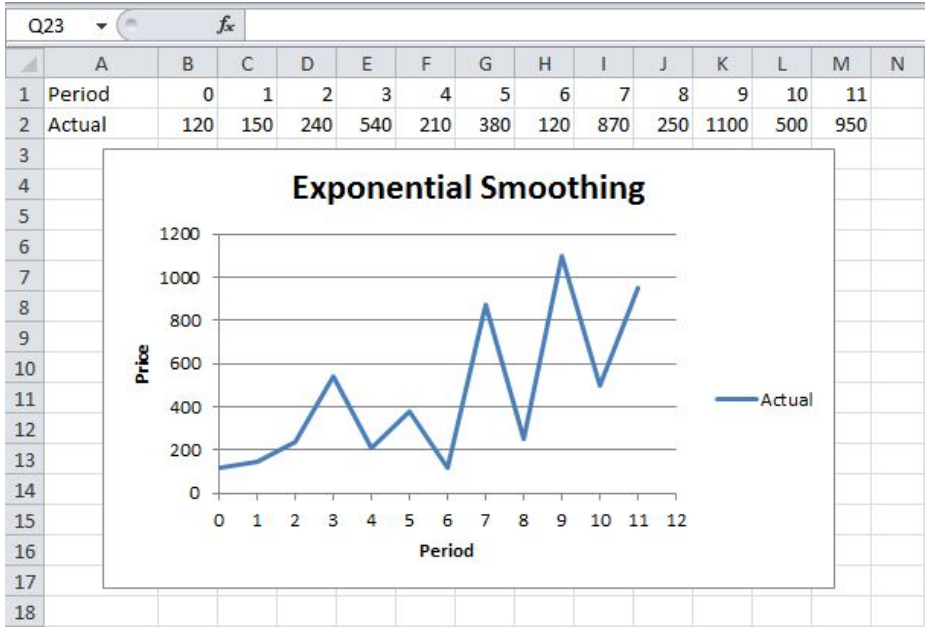


Conclusion: The larger the interval, the more the peaks and valleys are smoothed out. The smaller the interval, the closer the moving averages are to the actual data points.

5. Exponential Smoothing

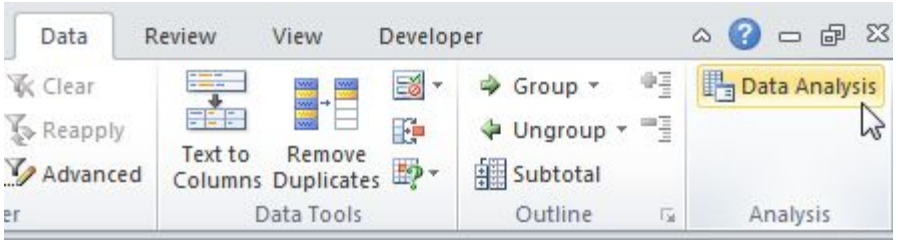
This example teaches you how to apply exponential smoothing to a time series in Excel. Exponential smoothing is used to smooth out irregularities (peaks and valleys) to easily recognize trends.

1. First, let's take a look at our time series.

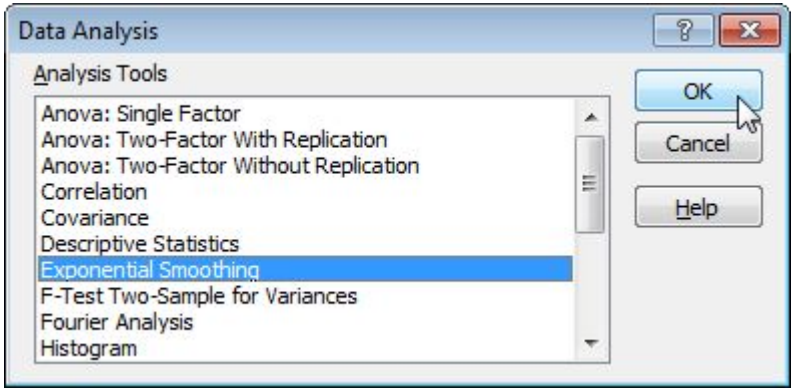


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Period	0	1	2	3	4	5	6	7	8	9	10	11
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950

2. On the Data tab, click Data Analysis.



3. Select Exponential Smoothing and click OK.

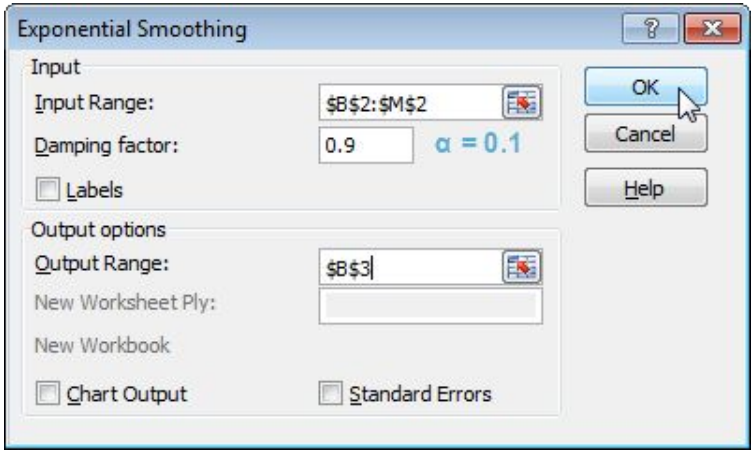


4. Click in the Input Range box and select the range B2:M2.

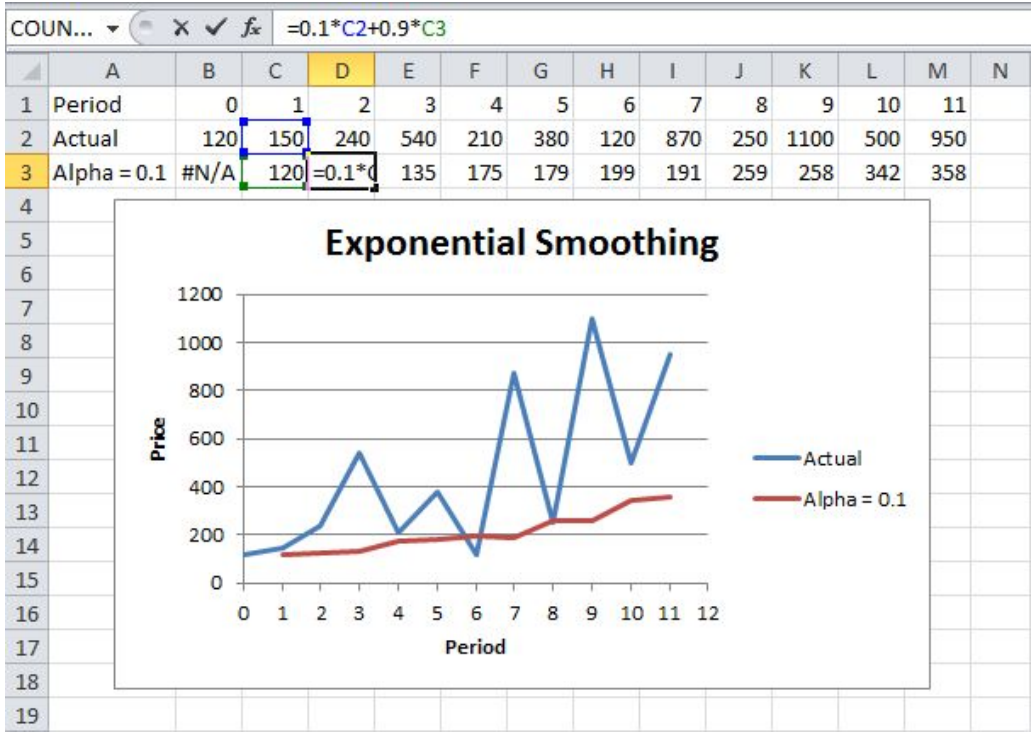
5. Click in the Damping factor box and type 0.9. Literature often talks about the smoothing constant α (alpha). The value (1- α) is called the damping factor.

6. Click in the Output Range box and select cell B3.

7. Click OK.



8. Plot a graph of these values.

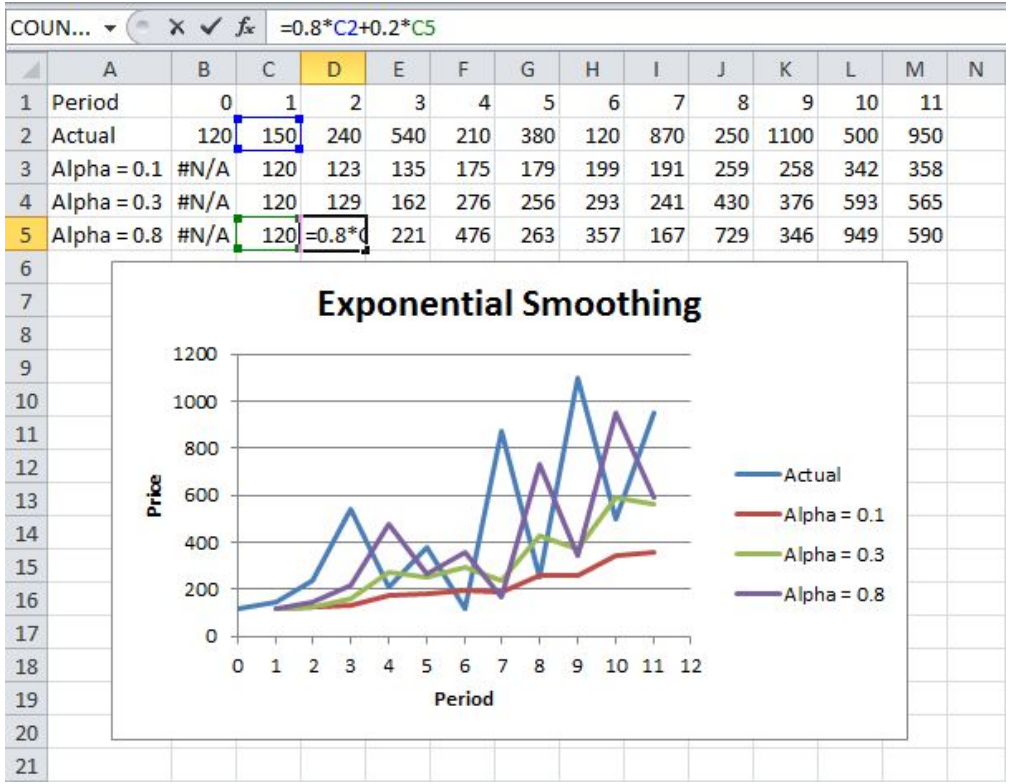


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Period	0	1	2	3	4	5	6	7	8	9	10	11
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950
3	Alpha = 0.1	#N/A	120	=0.1*C	135	175	179	199	191	259	258	342	358

Explanation: because we set alpha to 0.1, the previous data point is given a relatively small weight while the previous smoothed value is given a large weight (i.e. 0.9). As a result, peaks and valleys are smoothed out. The graph shows an increasing trend. Excel cannot calculate the smoothed value for the first data point because there is no previous data point. The smoothed value for the second data point equals the previous data point.

9. Repeat steps 2 to 8 for alpha = 0.3 and alpha = 0.8.

Conclusion: The smaller alpha (larger the damping factor), the more the peaks and valleys are smoothed out. The larger alpha (smaller the damping factor), the closer the smoothed values are to the actual data points.

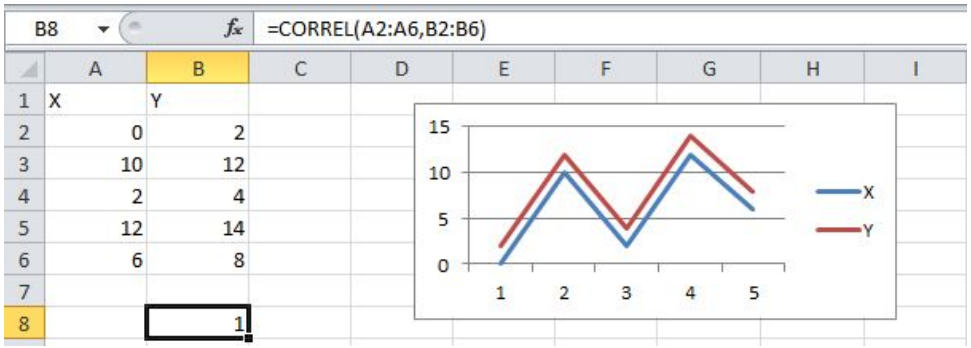


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Period	0	1	2	3	4	5	6	7	8	9	10	11
2	Actual	120	150	240	540	210	380	120	870	250	1100	500	950
3	Alpha = 0.1	#N/A	120	123	135	175	179	199	191	259	258	342	358
4	Alpha = 0.3	#N/A	120	129	162	276	256	293	241	430	376	593	565
5	Alpha = 0.8	#N/A	120	=0.8*C	221	476	263	357	167	729	346	949	590

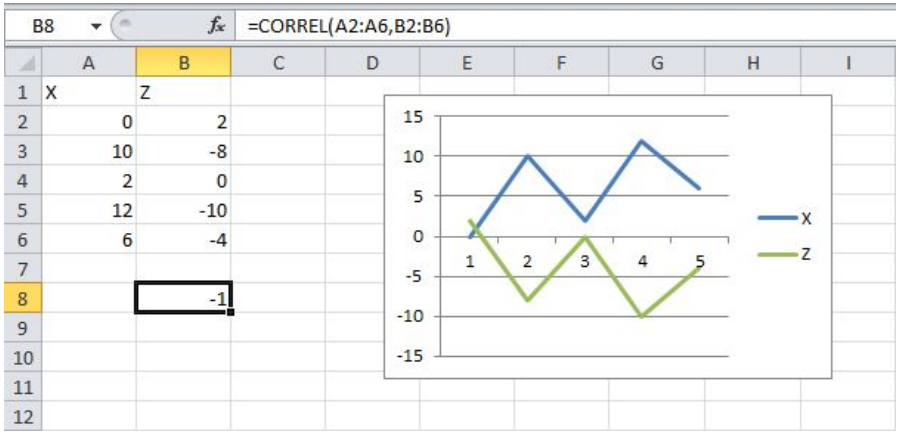
6. Correlation

The correlation coefficient (a value between -1 and +1) tells you how strongly two variables are related to each other. We can use the CORREL function or the Analysis Toolpak add-in in Excel to find the correlation coefficient between two variables.

- A correlation coefficient of +1 indicates a perfect positive correlation. As variable X increases, variable Y increases. As variable X decreases, variable Y decreases.



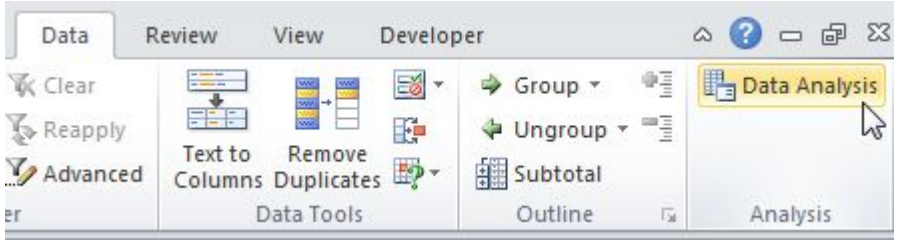
- A correlation coefficient of -1 indicates a perfect negative correlation. As variable X increases, variable Z decreases. As variable X decreases, variable Z increases.



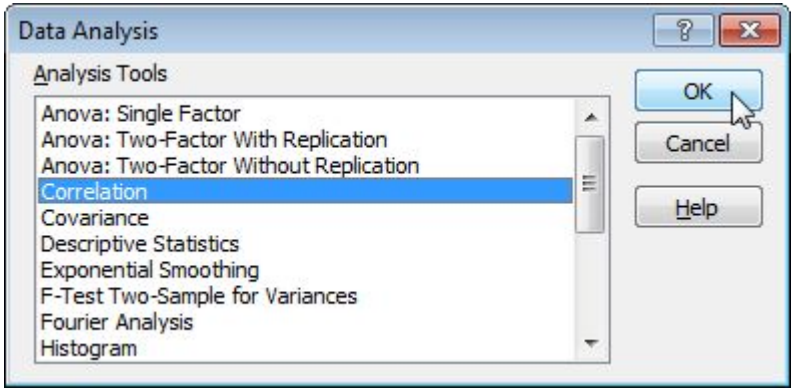
- A correlation coefficient near 0 indicates no correlation.

To use the Analysis Toolpak add-in in Excel to quickly generate correlation coefficients between multiple variables, execute the following steps.

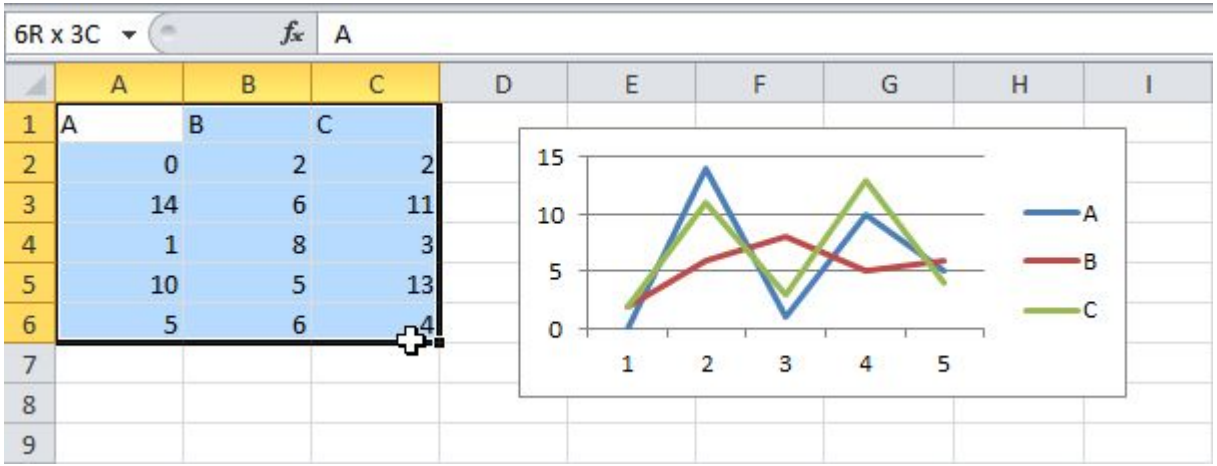
1. On the Data tab, click Data Analysis.



2. Select Correlation and click OK.



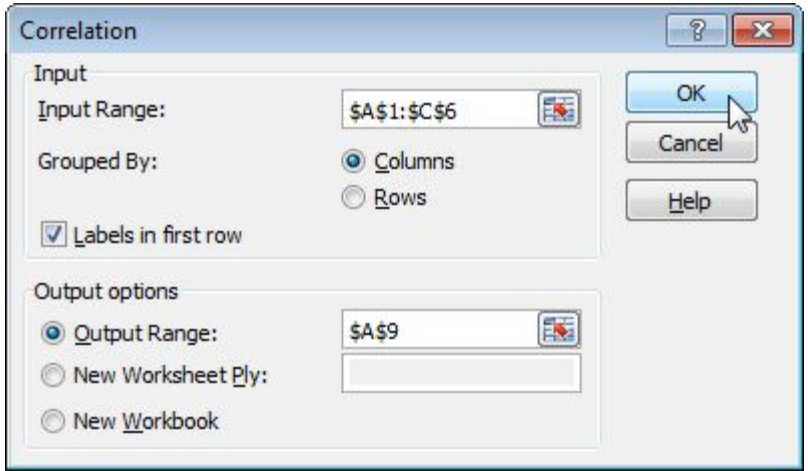
3. For example, select the range A1:C6 as the Input Range.



4. Check Labels in first row.

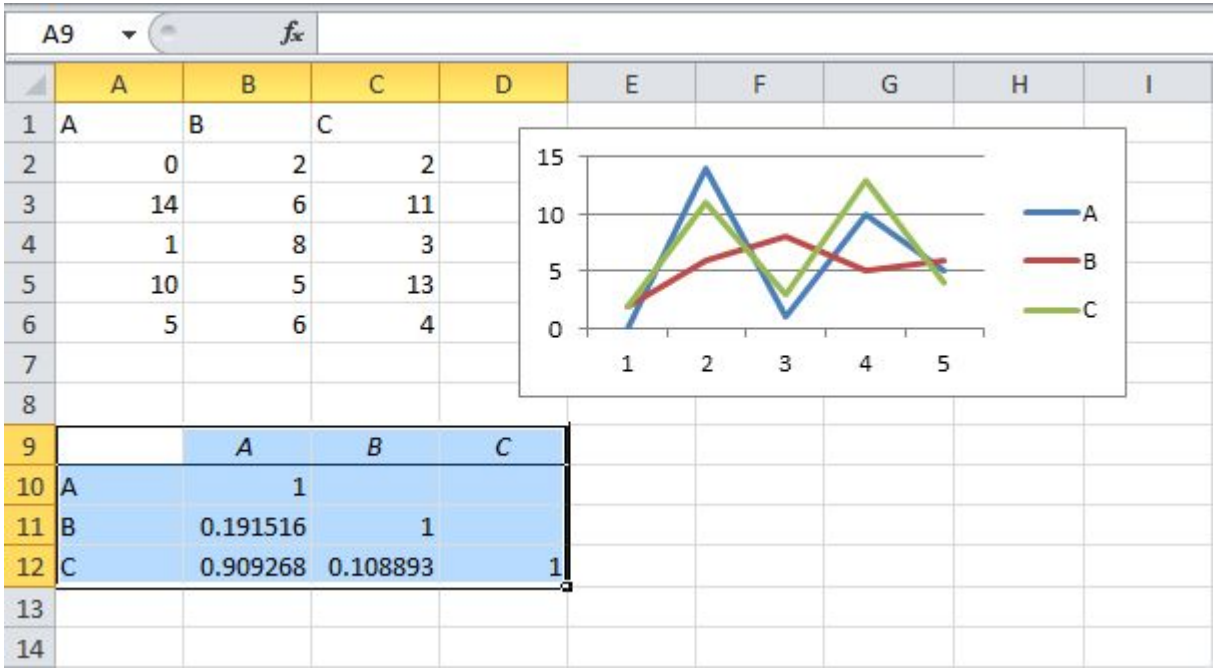
5. Select cell A9 as the Output Range.

6. Click OK.



Result.

Conclusion: variables A and C are positively correlated (0.91). Variables A and B are not correlated (0.19). Variables B and C are also not correlated (0.11) . You can verify these conclusions by looking at the graph.



7. Regression

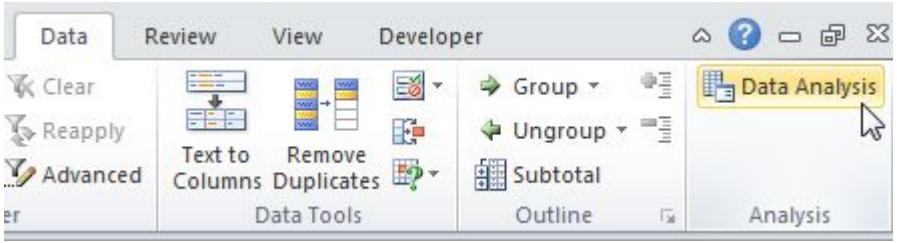
This example teaches you how to perform a regression analysis in Excel and how to interpret the Summary Output.

Below you can find our data. The big question is: is there a relation between Quantity Sold (Output) and Price and Advertising (Input). In other words: can we predict Quantity Sold if we know Price and Advertising?

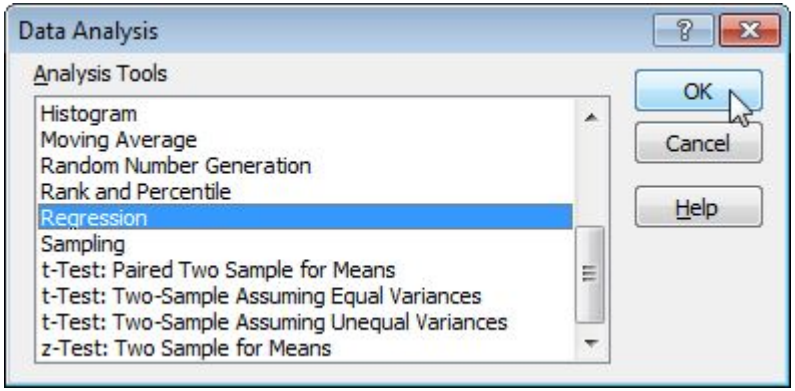
G14

	A	B	C	D
1	Quantity Sold	Price	Advertising	
2	8500	$2	$2,800	
3	4700	$5	$200	
4	5800	$3	$400	
5	7400	$2	$500	
6	6200	$5	$3,200	
7	7300	$3	$1,800	
8	5600	$4	$900	
9				
10				

1. On the Data tab, click Data Analysis.



2. Select Regression and click OK.



3. Select the Y Range (A1:A8). This is the predictor variable (also called dependent variable).

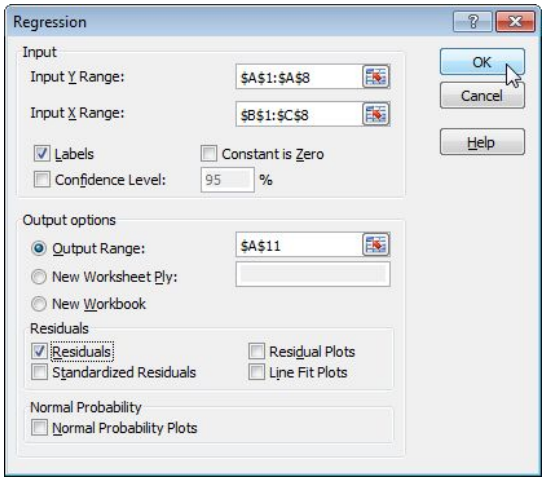
4. Select the X Range(B1:C8). These are the explanatory variables (also called independent variables). These columns must be adjacent to each other.

5. Check Labels.

6. Select an Output Range.

7. Check Residuals.

8. Click OK.



Excel produces the following Summary Output (rounded to 3 decimal places).

R Square

R Square equals 0.962, which is a very good fit. 96% of the variation in Quantity Sold is explained by the independent variables Price and Advertising. The closer to 1, the better the regression line (read on) fits the data.

11	SUMMARY OUTPUT	
12		
13	*Regression Statistics*	
14	Multiple R	0.981
15	R Square	0.962
16	Adjusted R Square	0.943
17	Standard Error	310.524
18	Observations	7
19		

Significance F and P-values

To check if your results are reliable (statistically significant), look at Significance F (0.001). If this value is less than 0.05, you're OK. If Significance F is greater than 0.05, it's probably better to stop using this set of independent variables. Delete a variable with a high P-value (greater than 0.05) and rerun the regression until Significance F drops below 0.05.

Most or all P-values should be below below 0.05. In our example this is the case. (0.000, 0.001 and 0.005).

20	ANOVA					
21		*df*	*SS*	*MS*	*F*	*Significance F*
22	Regression	2	9694299.568	4847149.784	50.269	0.001
23	Residual	4	385700.432	96425.108		
24	Total	6	10080000.000			
25						
26		*Coefficients*	*Std Error*	*t Stat*	*P-values*	*Lower 95%*	*Upper 95%*
27	Intercept	8536.214	386.912	22.062	0.000	7461.975	9610.453
28	Price	-835.722	99.653	-8.386	0.001	-1112.404	-559.041
29	Advertising	0.592	0.104	5.676	0.005	0.303	0.882
30							

Coefficients

The regression line is: y = Quantity Sold = 8536.214 -835.722 * Price + 0.592 * Advertising. In other words, for each unit increase in price, Quantity Sold decreases with 835.722 units. For each unit increase in Advertising, Quantity Sold increases with 0.592 units. This is valuable information.

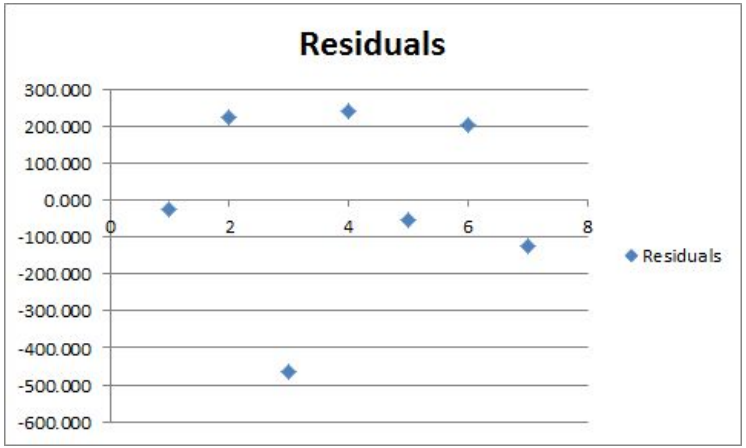
You can also use these coefficients to do a forecast. For example, if price equals $4 and Advertising equals $3000, you might be able to achieve a Quantity Sold of 8536.214 -835.722 * 4 + 0.592 * 3000 = 6970.

Residuals

The residuals show you how far away the actual data points are fom the predicted data points (using the equation). For example, the first data point equals 8500. Using the equation, the predicted data point equals 8536.214 -835.722 * 2 + 0.592 * 2800 = 8523.009, giving a residual of 8500 - 8523.009 = -23.009.

33	RESIDUAL OUTPUT		
34			
35	*Observation*	*Predicted Quantity Sold*	*Residuals*
36	1	8523.009	-23.009
37	2	4476.048	223.952
38	3	6265.938	-465.938
39	4	7160.883	239.117
40	5	6252.733	-52.733
41	6	7095.058	204.942
42	7	5726.330	-126.330
43			

You can also create a scatter plot of these residuals.



End

This is the end of Data Analytics using Excel