

Advanced topics in Excel

Angelo Gargantini

Basic Computer Science AA 19/20

1. Sommario

1. Sommario.....	2
2. COUNT functions.....	4
3. Text functions	6
Join strings.....	6
Extract info	9
Separate Strings.....	10
4. Debug a formula by using the Evaluate Formula feature	13
5. IsError	15
6. Text to Columns	16
7. Remove Unwanted Characters	19
8. Excel VLOOKUP tutorial for beginners - syntax and formula examples	20
What is VLOOKUP in Excel?.....	20
Example 1 – One Sheet Using Approximate Match	20
Understanding the VLOOKUP Function	21
Deconstructing the Excel Function	22
Understanding the VLOOKUP Arguments.....	22
Table Array Rules & Caveats.....	23
Using VLOOKUP in Our Formula	24
Example 2 – Multiple Excel Sheets & Exact Match	27
Defining the Argument Values.....	28
9. Pivot Tables	31
Insert a Pivot Table.....	31
Drag fields	32
Sort	33

Change Summary Calculation.....	35
Two-dimensional Pivot Table	36
10. Find Quick Answers With Excel Goal Seek.....	39
What is Goal Seek?	39
How to Use Excel Goal Seek	39
Goal Seek Requirements & Gotchas	41
11. Paired Sample t Test	43
Motivations.....	43
How to perform paired t-test with Data analysis tool excel	43

Medical school UniMIB AA 17/18

Angelo Gargantini – University of Bergamo

2. COUNT functions

The **COUNT** function counts the number of cells that contain numbers, and counts numbers within the list of arguments. Use the **COUNT** function to get the number of entries in a number field that is in a range or array of numbers. For example, you can enter the following formula to count the numbers in the range A1:A20: **=COUNT(A1:A20)**. In this example, if five of the cells in the range contain numbers, the result is **5**.

Syntax

COUNT(value1, [value2], ...)

The COUNT function syntax has the following arguments:

- **value1** Required. The first item, cell reference, or range within which you want to count numbers.
- **value2, ...** Optional. Up to 255 additional items, cell references, or ranges within which you want to count numbers.

Note: The arguments can contain or refer to a variety of different types of data, but only numbers are counted.

Remarks

- Arguments that are numbers, dates, or a text representation of numbers (for example, a number enclosed in quotation marks, such as "1") are counted.
- Logical values and text representations of numbers that you type directly into the list of arguments are counted.
- Arguments that are error values or text that cannot be translated into numbers are not counted.
- If an argument is an array or reference, only numbers in that array or reference are counted. Empty cells, logical values, text, or error values in the array or reference are not counted.
- If you want to count logical values, text, or error values, use the **COUNTA** function.
- If you want to count only numbers that meet certain criteria, use the **COUNTIF** function or the **COUNTIFS** function.

The **COUNTA** function counts the number of cells that are not empty in a range.

Syntax

COUNTA(value1, [value2], ...)

The COUNTA function syntax has the following arguments:

- **value1** Required. The first argument representing the values that you want to count.
- **value2, ...** Optional. Additional arguments representing the values that you want to count, up to a maximum of 255 arguments.

Remarks

- The **COUNTA** function counts cells containing any type of information, including error values and empty text (""). For example, if the range contains a formula that returns an empty string, the **COUNTA** function counts that value. The **COUNTA** function does not count empty cells.
- If you do not need to count logical values, text, or error values (in other words, if you want to count only cells that contain numbers), use the **COUNT** function.
- If you want to count only cells that meet certain criteria, use the **COUNTIF** function or the **COUNTIFS** function.

Use COUNTIF, one of the statistical functions, to count the number of cells that meet a criterion; for example, to count the number of times a particular city appears in a customer list.

In its simplest form, COUNTIF says:

- =COUNTIF(Where do you want to look?, What do you want to look for?)

For example:

	A	B
1	Data	Data
2	apples	32
3	oranges	54
4	peaches	75
5	apples	86

=COUNTIF(A2:A5,"apples")

Counts the number of cells with apples in cells A2 through A5. The result is 2.

=COUNTIF(A2:A5,A4)

Counts the number of cells with peaches (the value in A4) in cells A2 through A5. The result is 1.

=COUNTIF(A2:A5,A2)+COUNTIF(A2:A5,A3)

Counts the number of apples (the value in A2), and oranges (the value in A3) in cells A2 through A5. The result is 3. This formula uses COUNTIF twice to specify multiple criteria, one criteria per expression. You could also use the [COUNTIFS](#) function.

=COUNTIF(B2:B5,">55")

Counts the number of cells with a value greater than 55 in cells B2 through B5. The result is 2.

=COUNTIF(B2:B5,"<>"&B4)

Counts the number of cells with a value not equal to 75 in cells B2 through B5. The ampersand (&) merges the comparison operator for not equal to (<>) and the value in B4 to read =COUNTIF(B2:B5,"<>75"). The result is 3.

=COUNTIF(B2:B5,">=32")-COUNTIF(B2:B5,">85")

Counts the number of cells with a value greater than (>) or equal to (=) 32 and less than (<) or equal to (=) 85 in cells B2 through B5. The result is 3.

3. Text functions

<http://www.excel-easy.com/>

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

Join strings

The CONCATENATE function in Excel is designed to join different pieces of text together or combine values from several cells into one cell.

The syntax of Excel CONCATENATE is as follows:

CONCATENATE(text1, [text2], ...)

Where **text** is a text string, cell reference or formula-driven value.

Below you will find a few examples of using the CONCATENATE function in Excel.

Note. In Excel 2016, Excel 2019, Excel Online and Excel Mobile, CONCATENATE is replaced with the CONCAT function, which has exactly the same syntax. Although the CONCATENATE function is kept for backward compatibility, it is recommended to use CONCAT instead because Microsoft does not give any promises that CONCATENATE will be available in future versions of Excel.

Concatenating the values of several cells

The simplest CONCATENATE formula to combine the values of cells A1 and B1 is as follows:

=CONCATENATE (A1, B1)

Please note that the values will be knit together without any delimiter, as in row 2 in the screenshot below.

To separate the values with a space, enter " " in the second argument, as in row 3 in the screenshot below.

=CONCATENATE (A1, " ", B1)

	A	B	C	D
1	Source data	Result	Formula	
2	Project	1 Project1	=CONCATENATE(A2, B2)	
3		Project 1	=CONCATENATE(A2, " ", B2)	

To separate the concatenated values with other delimiters such as a comma, space or slash, please see [Excel CONCATENATE formulas with special characters](#).

Concatenating a text string and cell value

There is no reason for the Excel CONCATENATE function to be limited to only joining cells' values. You can also use it to concatenate various text strings to make the result more meaningful. For example:

```
=CONCATENATE(A1, " ", B1, " completed")
```

The above formula informs the user that a certain project is completed, as in row 2 in the screenshot below. Please notice that we add a space before the word " completed" to separate the concatenated text strings.

Naturally, you can add a text string in the beginning or in the middle of your Concatenate formula as well:

```
=CONCATENATE("See ", A1, " ", B1)
```

A space (" ") is added in between the combined values, so that the result displays as "Project 1" rather than "Project1".

	A	B	C	D
1	Source data	Result	Formula	
2	Project	1	Project 1 completed	=CONCATENATE(A2, " ", B2," completed")
3			See Project 1	=CONCATENATE("See ", A2, " ", B2)

Concatenating a text string and a formula-calculated value

To make the result returned by some formula more understandable for your users, you can concatenate it with a text string that explains what the value actually is.

For example, you can use the following formula to return the current date:

```
=CONCATENATE("Today is ",TEXT(TODAY()), "dd-mmm-yy")
```

	A	B
1	Result	Formula
2	Today is 14-Jul-15	=CONCATENATE("Today is ", TEXT(TODAY()), "dd-mmm-yy")

Using CONCATENATE in Excel - things to remember

To ensure that your CONCATENATE formulas always deliver the correct results, remember the following simple rules:

- Excel CONCATENATE function requires at least one "text" argument to work.
- In a single CONCATENATE formula, you can concatenate up to 255 strings, a total of 8,192 characters.
- The result of the CONCATENATE function is always a text string, even when all of the source values are numbers.
- Excel CONCATENATE does not recognize arrays. Each cell reference must be listed separately. For example, you should write =CONCATENATE(A1, A2, A3) instead of =CONCATENATE(A1:A3).

- If at least one of the CONCATENATE function's arguments is invalid, the formula returns a #VALUE! error.

"&" operator to concatenate strings in Excel

In Microsoft Excel, & operator is another way to concatenate cells. This method come in very handy in many scenarios because typing the ampersand sign (&) is much quicker than typing the word "concatenate" :)

Similarly to the CONCATENATE function, you can use "&" in Excel to combine different text strings, cell values and results returned by other functions.

Excel "&" formula examples

To see the concatenation operator in action, let's re-write the CONCATENATE formulas discussed above:

Concatenate the values in A1 and B1:

=A1&B1

Concatenate the values in A1 and B1 separated with a space:

=A1&" "&B1

Concatenate the values in A1, B1 and a text string:

=A1 & B1 & " completed"

Concatenate a string and the result of the TEXT / TODAY function:

= "Today is " & TEXT(TODAY()), "dd-mmm-yy")

As demonstrated in the screenshot below, the CONCATENATE function and "&" operator return identical results:

1	Concatinating strings using "&" operator			
2	Source data	Result	Formula	
3	Project	1 Project1	=A3 & B3	
4				
5	Paul	Smith	Paul Smith	=A5 & " " & B5
6				
7	Project	1 Project1 completed	=A7 & B7 & " completed"	
8				
9	Project	2 See Project 2	="See "& A9 & " " & B9	
10				
11		Today is 29-Jun-15	="Today is " & TEXT(TODAY()), "dd-mmm-yy")	

Excel "&" operator vs. CONCATENATE function

Many users wonder which is a more efficient way to concatenate strings in Excel - CONCATENATE function or "&" operator.

The only essential difference between CONCATENATE and "&" operator is the 255 strings limit of the Excel CONCATENATE function and no such limitations when using the ampersand. Other

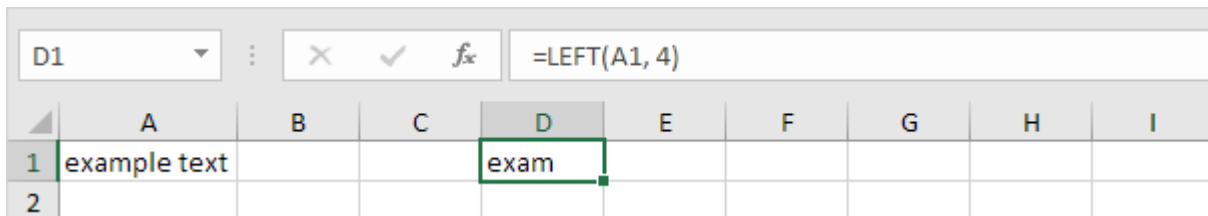
than that, there is no difference between these two concatenation methods, nor is there any speed difference between the CONCATENATE and "&" formulas.

And since 255 is a really big number and in real-life tasks someone will hardly ever need to combine that many strings, the difference boils down to the comfort and ease of use. Some users find CONCATENATE formulas easier to read, I personally prefer using the "&" method. So, simply stick to the concatenation technique that you feel more comfortable with.

Extract info

Left

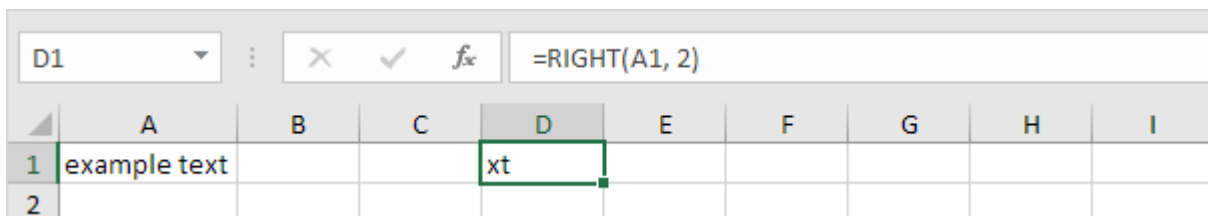
To extract the leftmost characters from a string, use the LEFT function.



	A	B	C	D	E	F	G	H	I
1	example text			exam					
2									

Right

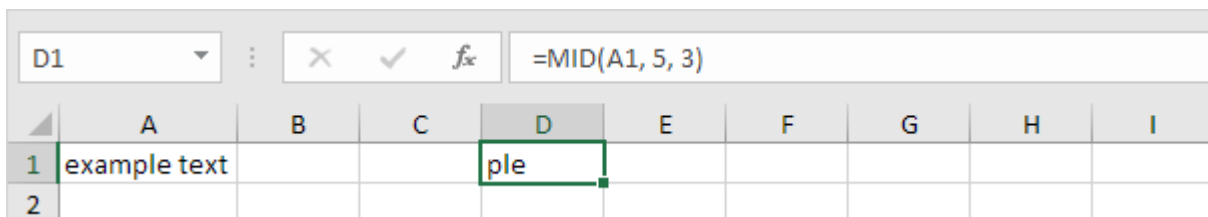
To extract the rightmost characters from a string, use the RIGHT function.



	A	B	C	D	E	F	G	H	I
1	example text			xt					
2									

Mid

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



	A	B	C	D	E	F	G	H	I
1	example text			ple					
2									

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

Len

The problem we are dealing with is that we need to tell Excel where we want to separate the string. In case of Smith, Mike the comma is at position 6 while in case of Williams, Janet the comma is at position 9.

1. To get the first name, use the formula below.

	A	B	C	D	E	F	G	H
1	Full Name	First Name	Last Name					
2	Smith, Mike	Mike						
3	Johnson, Matthew							
4	Williams, Janet							
5	Brown, Sandra							
6	Jones, Lisa							
7	Millar, Peter							
8								

Explanation: to find the position of the comma, use the FIND function (position 6). To get the length of a string, use the LEN function (11 characters). $=RIGHT(A2,LEN(A2)-FIND(",",A2)-1)$ reduces to $=RIGHT(A2,11-6-1)$. $=RIGHT(A2,4)$ extracts the 4 rightmost characters and gives the desired result (Mike).

2. To get the last name, use the following formula.

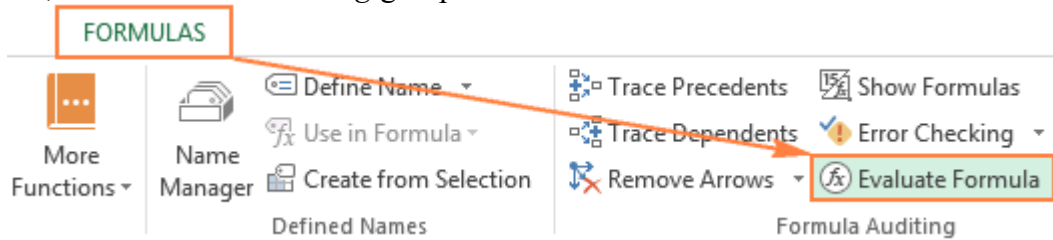
	A	B	C	D	E	F	G	H
1	Full Name	First Name	Last Name					
2	Smith, Mike	Mike	Smith					
3	Johnson, Matthew							
4	Williams, Janet							
5	Brown, Sandra							
6	Jones, Lisa							
7	Millar, Peter							
8								

Explanation: to find the position of the comma, use the FIND function (position 6). $=LEFT(A2,FIND(",", A2)-1)$ reduces to $=LEFT(A2,6-1)$. $=LEFT(A2,5)$ extracts the 5 leftmost characters and gives the desired result (Smith).

3. Select the range B2:C2 and drag it down.

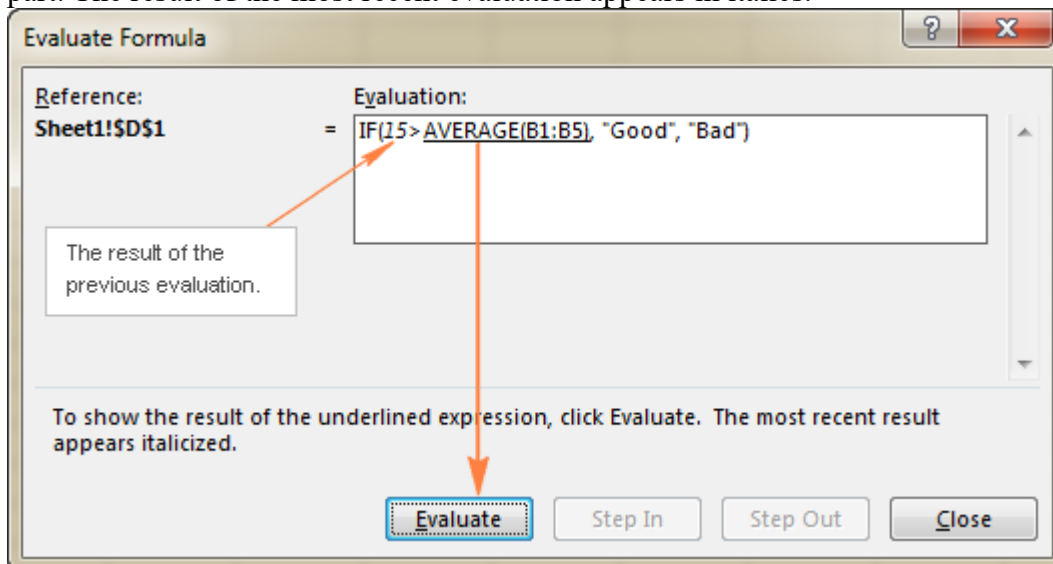
4. Debug a formula by using the Evaluate Formula feature

A way to evaluate formulas in Excel is the **Evaluate Formula** option that resides on the *Formulas* tab, in the *Formula Auditing* group.



As soon as you click this button, the *Evaluate Formula* dialog box will pop up, where you can inspect each part of your formula in the order the formula is calculated.

All you need to do is click the **Evaluate** button and examine the value of the underlined formula part. The result of the most recent evaluation appears in italics.



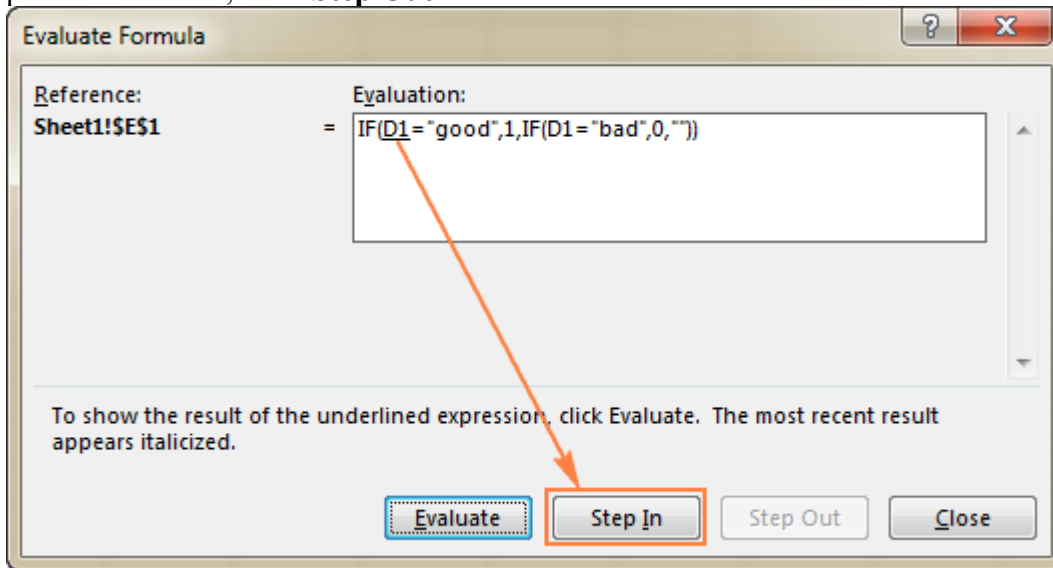
Continue clicking the **Evaluate** button until each part of your formula has been tested.

To end the evaluation, click the **Close** button.

To start the formula evaluation from the beginning, click **Restart**.

If the underlined part of the formula is a reference to a cell containing another formula, click the **Step In** button to have that other formula displayed in the *Evaluation* box. To get back to the

previous formula, click **Step Out**.

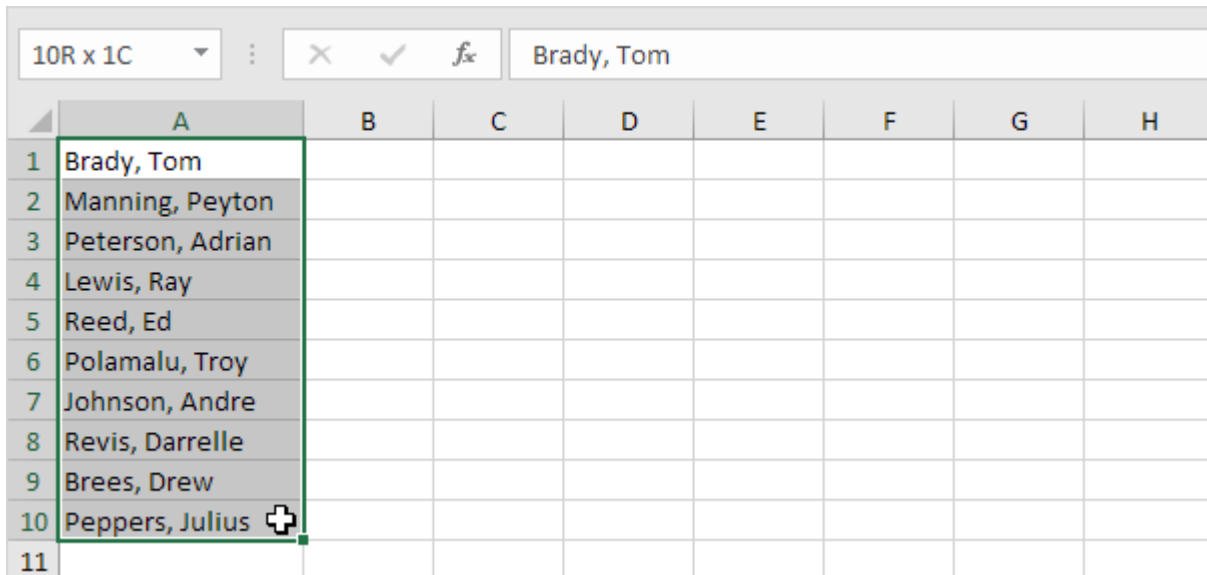


Note. The *Step In* button is not available for a cell reference pointing to another formula in a different workbook. Also, it is not available for a cell reference that appears in the formula for the second time (like the second instance of D1 in the screenshot above).

6. Text to Columns

To separate the contents of one **Excel cell** into **separate columns**, you can use the 'Convert Text to Columns Wizard'. For example, when you want to separate a list of full names into last and first names.

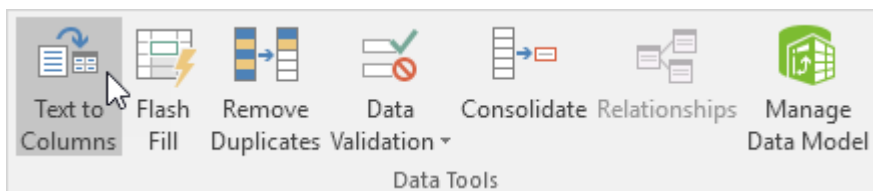
1. Select the range with full names.



The screenshot shows an Excel spreadsheet with a range of names selected in column A. The names are: Brady, Tom; Manning, Peyton; Peterson, Adrian; Lewis, Ray; Reed, Ed; Polamalu, Troy; Johnson, Andre; Revis, Darrelle; Brees, Drew; Peppers, Julius. The selection is highlighted with a green border. The formula bar at the top shows the active cell contains 'Brady, Tom'.

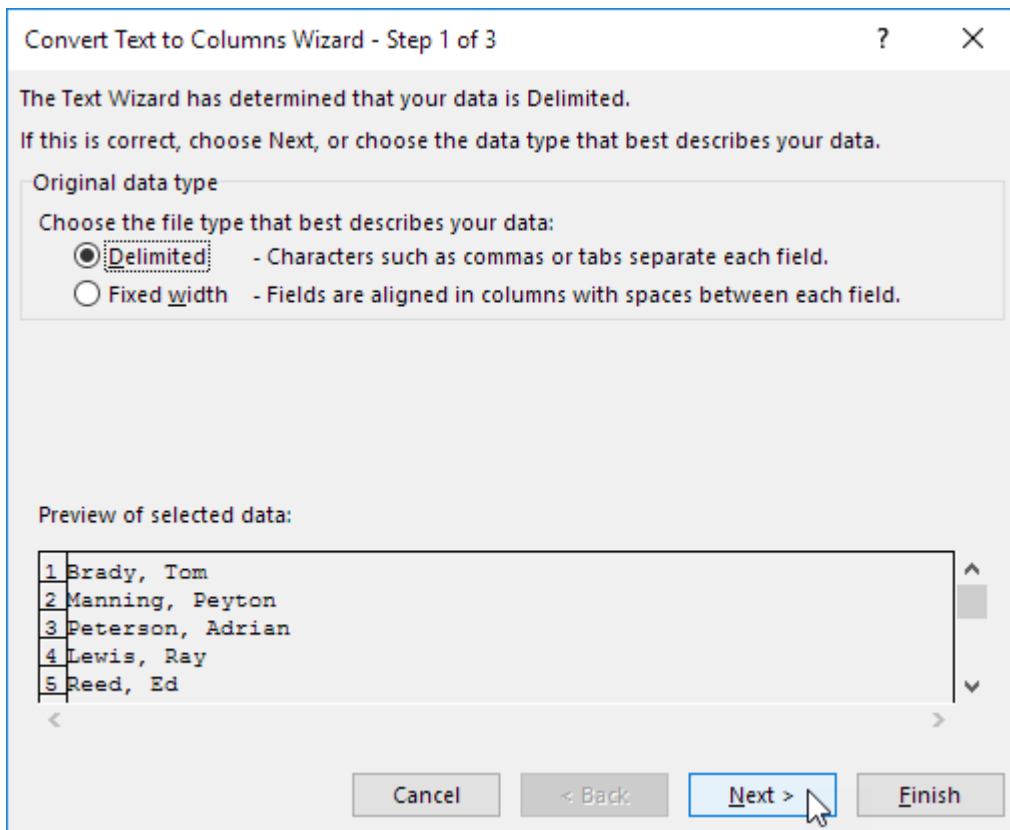
	A	B	C	D	E	F	G	H
1	Brady, Tom							
2	Manning, Peyton							
3	Peterson, Adrian							
4	Lewis, Ray							
5	Reed, Ed							
6	Polamalu, Troy							
7	Johnson, Andre							
8	Revis, Darrelle							
9	Brees, Drew							
10	Peppers, Julius							
11								

2. On the Data tab, in the Data Tools group, click Text to Columns.



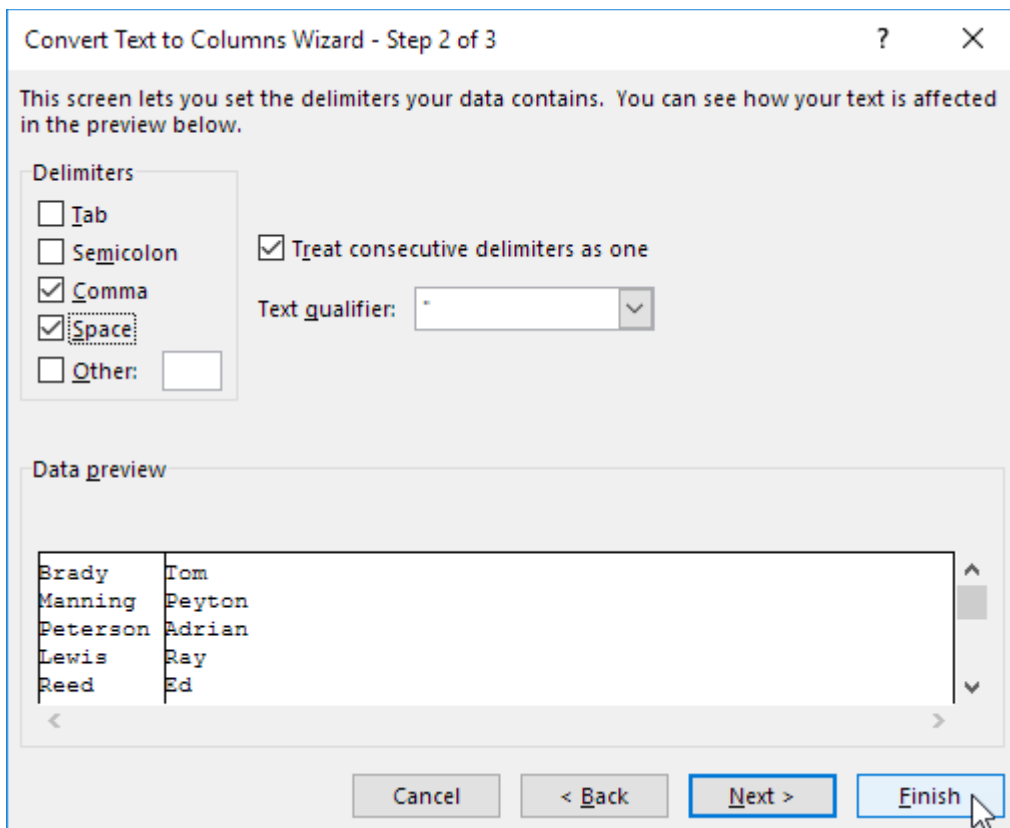
The following dialog box appears.

3. Choose Delimited and click Next.



4. Clear all the check boxes under Delimiters except for the Comma and Space check box.

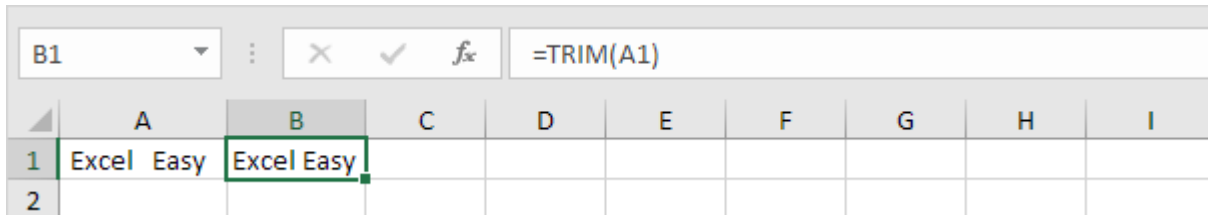
5. Click Finish.



7. Remove Unwanted Characters

The TRIM function returns a string with extra **spaces**, starting spaces and ending spaces removed. The CLEAN function **removes nonprintable characters** from a string.

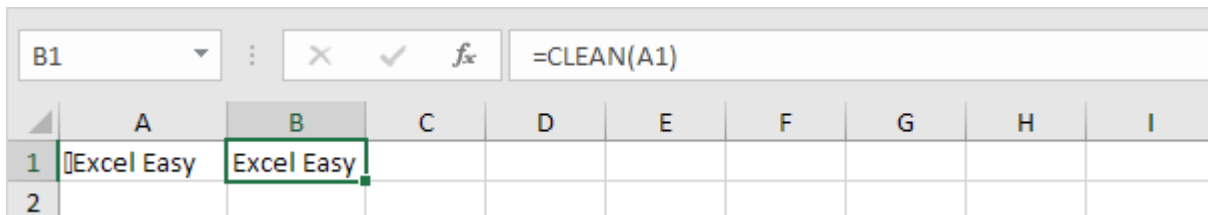
1. For example, cell A1 below contains the string " Excel Easy "



	A	B	C	D	E	F	G	H	I
1	Excel Easy	Excel Easy							
2									

Note: the TRIM function returns the string "Excel Easy".

2. For example, cell A1 below contains a nonprintable character.



	A	B	C	D	E	F	G	H	I
1	Excel Easy	Excel Easy							
2									

Note: the CLEAN function removes this nonprintable character. Text imported from other applications may contain nonprintable characters.

8. Excel VLOOKUP tutorial for beginners - syntax and formula examples

Last Updated on 28-Sep-2019 by [Anne Hennegar](#)

Spreadsheets are known for having a number of formulas and functions. One that continually intrigues users is VLOOKUP. Excel does include additional information. But sometimes it's best to learn how to use VLOOKUP by following along with example spreadsheets.

```
VLOOKUP(lookup_value,table_array,col_index_num,[range_lookup])
```

It looked too difficult and I couldn't see the immediate value. But sometimes pushing through difficulties can make things easier in the long run.

When it comes to learning Microsoft Excel formulas and functions, I like to start with an easy example. This training will provide two examples using different arguments and lookup values.

What is VLOOKUP in Excel?

VLOOKUP is an Excel function that allows you to search and retrieve a cell's content from one column and use it in another location. As you might guess, the "V" stands for **vertical** and relies on looking up data from the leftmost column of a **lookup table**.

This lookup column could be on the same sheet you're viewing or another within your workbook. The function requires a **common field** or **key** and four **arguments**. The function allows you to specify whether to use an **exact match** or **approximate match**.

Let's put these terms in context and give a real example of where and how it might be used. When I do website analysis, I use a number of tools. I export data from each service into a .CSV file or Excel .XLSX file. Each tool has its own file that I use to create a sheet within a workbook.

The problem is I have five worksheets in this workbook. I don't want to be switching between each sheet because that's not efficient. Moreover, I want to do some of my own calculations using other Excel formulas.

The solution is to find a common denominator or key between these worksheets. In this case, all the data files from my analysis tools have a field for the page URL. Using this common key, I can create a new spreadsheet and pull only the needed columns from each tool using VLOOKUP. Excel does all the heavy lifting. This allows me to concentrate on one worksheet with just the data I want.

Example 1 – One Sheet Using Approximate Match

Periodically, I've volunteered to work on local elections. I often get assigned the data analysis of the voter registration file. These files are massive.

One column contains the voter's birth date. However, I didn't want the voter's birth date to show on any of the final distributed files. But, I did want to do some age analysis.

Instead, I decided to create a segment based on age ranges and VLOOKUP. Excel would do a vertical lookup that returns the value from one column to the desired cell. Rather than showing a voter was 28, I would define them as “Young”.

	A	B	C	D	E	F	G	H	I
1	First	Last	DOB	Age	Segment			Age	Label
2	Sophia	Collins	9/23/1978	39	Mature			18	New
3	Evelyn	Bennett	2/14/1967	51	Mature			21	Young
4	Geneva	Allen	6/6/1990	28	Young			39	Mature
5	Silvia	Walsh	5/30/1983	35	Young			60	Senior
6	Katrina	Rodriquez	12/23/1998	19	New				
7	Sandy	Day	7/24/1973	45	Mature				
8	Dallas	Roberts	6/12/1952	66	Senior				
9	Jamie	Harrington	8/22/1958	59	Mature				
10	Jermaine	Shelton	10/23/1977	40	Mature				
11	Freddie	Baker	5/12/1968	50	Mature				
12									
13									
14									

Using VLOOKUP to determine Segments

Understanding the VLOOKUP Function

Let's refer to the screenshot above with my first fictitious voter, Sophia Collins. If you scan across to Column D (Age) you'll see she is 39 years old and in the “Mature” segment. The value of “Mature” in Column E was dynamically pulled in using Excel's VLOOKUP function.

The small table in columns H and I with blue headings is the **lookup table**. Microsoft refers to this lookup table as a **table_array**. This is where I've defined my 4 age segments.

The way “Segment” works is if a voter is under 21, they are “New”. From 21-38, they are “Young”. From 39-59, they are “Mature”. And if they are 60 or older they are “Senior”.

In the case of Sophia, Excel would take her age of 39 from cell D2 and return the closest match from Column H. Both these columns contain age data, which is our common key.

When Excel found a match it would go to Column I and get the Label. The returned value was then copied to cell E2, the Segment. It might help to think of columns as being vertical. After all, this is a vertical lookup.

You might notice that the lookup table doesn't list every age. It doesn't have to because I'm using an **approximate match**. I'm telling Excel to find me the closest age.

For example, the next voter Evelyn Bennett is 51, but there is no value for 51 in Column H. In this case, 51 falls between 39 and 59 so she is also labeled “Mature”.

As we stated, to use the VLOOKUP function there needs to be a common key. In this case, it’s age. Both columns D and H contain ages. The column headings and cell contents can be different and don’t have to be an exact match.

Deconstructing the Excel Function

Let’s peel away some of the mystery and display how VLOOKUP shows in the Excel formula bar. In this illustration, I’ve clicked cell D2.

	A	B	C	D	E	F	G	H	I	J
1	First	Last	DOB	Age	Segment			Age	Label	
2	Sophia	Collins	9/23/1978	39	(TRUE)			18	New	
3	Evelyn	Bennett	2/14/1967	51	Mature			21	Young	
4	Geneva	Allen	6/6/1990	28	Young			39	Mature	
5	Silvia	Walsh	5/30/1983	35	Young			60	Senior	
6	Katrina	Rodriquez	12/23/1998	19	New					
7	Sandy	Day	7/24/1973	45	Mature					
8	Dallas	Roberts	6/12/1952	66	Senior					
9	Jamie	Harrington	8/22/1958	59	Mature					
10	Jermaine	Shelton	10/23/1977	40	Mature					
11	Freddie	Baker	5/12/1968	50	Mature					
12										

Showing VLOOKUP in formula bar

[A] – This represents the VLOOKUP formula for Cell D2

=VLOOKUP(D2,\$H\$2:\$I\$5,2,TRUE)

[B] – Cell D2 is our first argument called **Lookup_value**.

[C] – The cell range \$H\$2:\$I\$5 is our **Table_array** and the second argument.

[D] – 2 is the **Col_index_num** from our Table_array and the third argument.

[E] – TRUE is the **Range_lookup** and the fourth argument.

The good news is the **Functions Arguments** dialog guides you through these elements so you don’t need to type that long string in Excel’s formula bar.

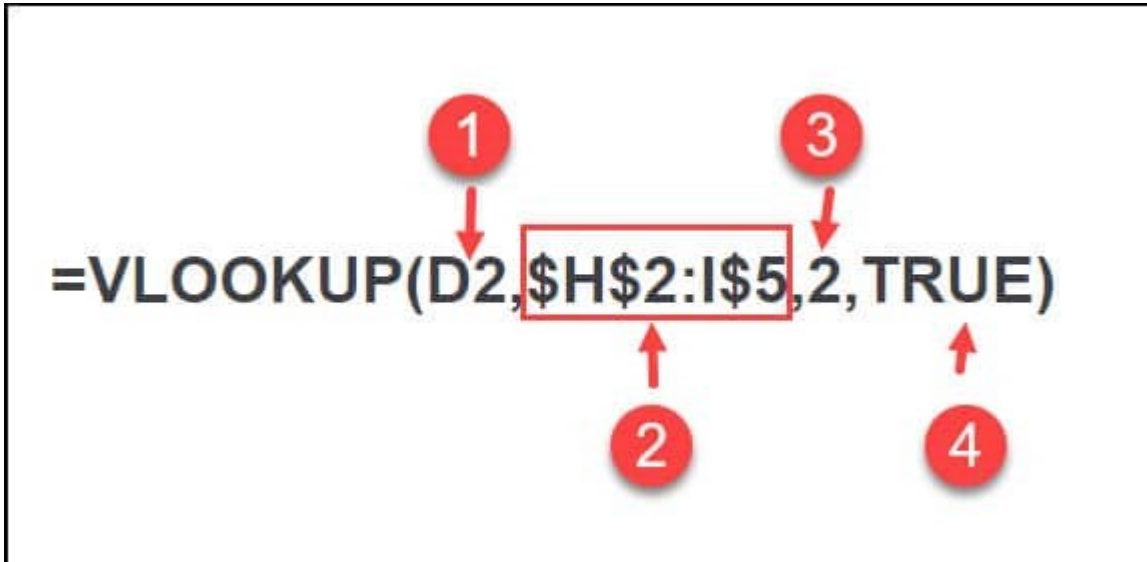
Understanding the VLOOKUP Arguments

The term “argument” isn’t as complicated or negative as it sounds. If you’re familiar with the Excel formula bar, an argument is what goes in between the parentheses (). It provides some sort of input value for an Excel function.

Some Excel functions have required arguments and others don’t need arguments. For example, to compute the voter’s age I also used the TODAY function =TODAY() which doesn’t use any arguments. Some common argument examples include:

- cell range
- true/false logical value
- number

Using the formula from cell D2, here's how these four arguments work.



1. Lookup_value – Think of this field as your starting point. In this example, I want to look up Sophia's Age from cell D2.

2. Table_array – This is the cell range for your lookup table. This range lookup can be on your existing worksheet or another worksheet. In this example, I have a small table with the age groupings and corresponding label.

3. Col_index_num – This is column number on your lookup table that has the information you need. In our example, we want column 2 from the Label column. This will become our Segment name.

When we count, we're counting the columns on the lookup table. Even though the "Label" column is Column I or the 9th column, it's the 2nd column on the lookup table. Some people call this a **Column Index**.

4. Range-lookup – this field defines how close a match should exist between your Lookup_value (D2) and the value in the leftmost column on our lookup table. In our case, we want an approximate match so we'll use "TRUE".

Table Array Rules & Caveats

There are several rules to remember about this table array.

- Rule 1 – The **left column must contain the values being referenced**. Leftmost doesn't mean it has to be in Column A. It's just the leftmost or first column on the table_array. For example, on the lookup table above, the leftmost column is H.

- Rule 2 – **You can't have duplicate values** in the leftmost column of the lookup range. I couldn't have two entries with the value "39" with one being "Mature" and another "39" for "Go Getter". Excel would complain.
- Rule 3 – **When referring to the lookup table, you want absolute cell references when you copy the VLOOKUP formula to other cells.**

For example, if I want to use the same formula in cells E3 through E11, I don't want my lookup cell references shifting each time I move down to the next cell. I need the cell references to be constant. This is called an **absolute cell reference**.

After you define your lookup range of cells, you can press **F4**. This will cycle through absolute and relative cell references. You want to select the option that includes a \$ before your Column and Row. You can get around this if you know how to use [Excel name ranges](#).

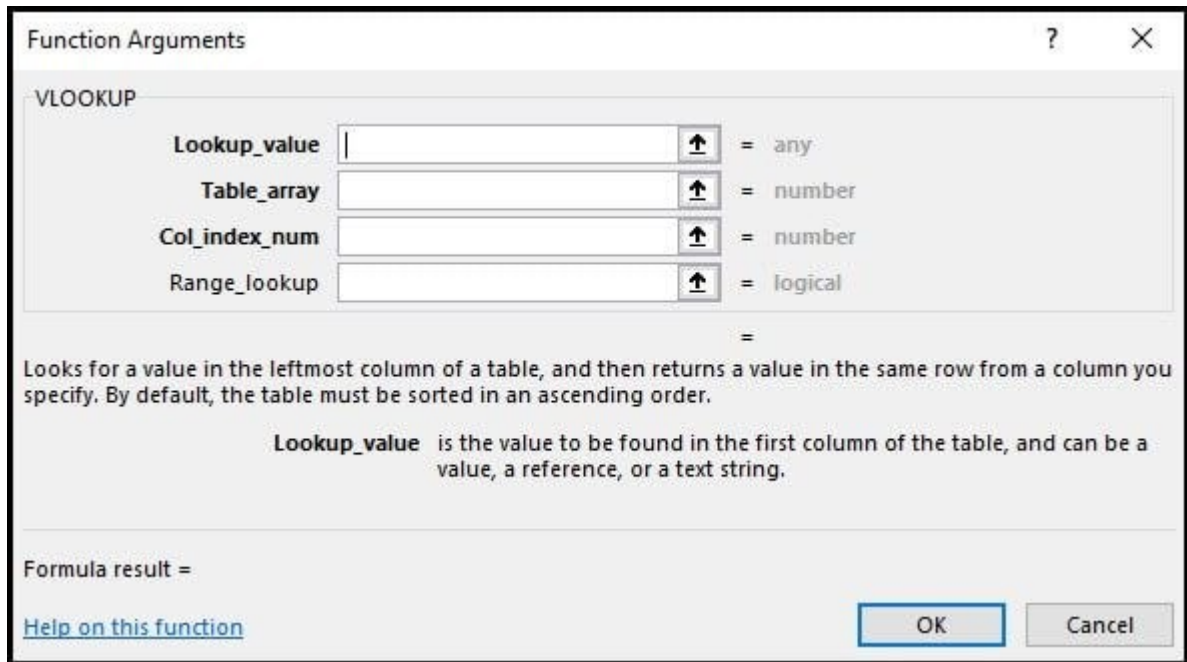
Using VLOOKUP in Our Formula

1. Add in the column where you'll enter the formula. In my case, I added Column E – Segment.
2. Click cell E2.
3. Click your **Formulas** tab from the top menu.
4. Click the **Insert Function** button.
5. From the **Insert Function** dialog, type "vlookup" in the **Search for a function** textbox. You may also select it from the **Lookup & Reference** category.
6. Click **Go**.

The screenshot shows the Microsoft Excel interface with the 'Formulas' ribbon selected. The 'Insert Function' button is highlighted with a red box. Below the ribbon, the spreadsheet data is visible, including columns for 'First', 'Last', 'DOB', 'Age', and 'Segment'. A table on the right side of the spreadsheet lists age ranges and labels: 18 New, 21 Young, 39 Mature, and 60 Senior. The 'Insert Function' dialog box is open, with 'vlookup' entered in the search field. The 'VLOOKUP' function is selected in the list of functions. The dialog box also provides a brief description of the VLOOKUP function: 'VLOOKUP(lookup_value,table_array,col_index_num,range_lookup) Looks for a value in the leftmost column of a table, and then returns a value in the same row from a column you specify. By default, the table must be sorted in an ascending order.'

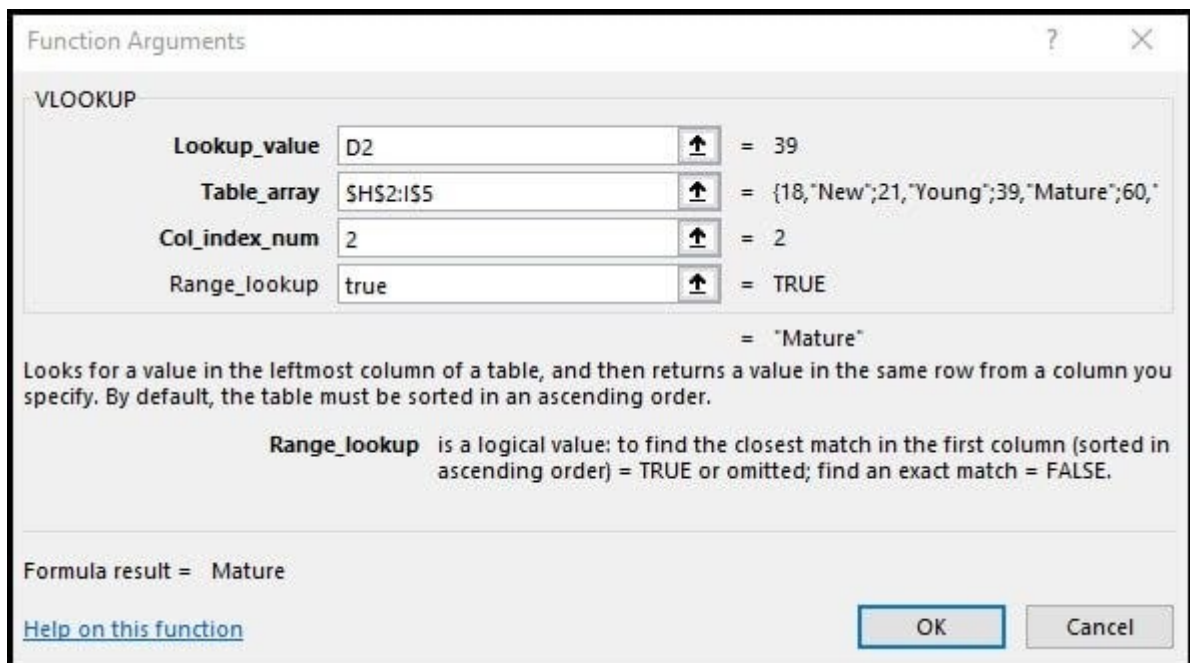
Finding the VLOOKUP function

7. Click **OK**. The **Function Arguments** dialog will appear with text boxes for the required arguments.



VLOOKUP Required arguments

8. In **Lookup_value** type **D2**. Or, you can click the cell.
9. In **Table_array** type **\$H\$2:\$I\$5**. Note the \$ signs.
10. In **Col_index_num** type **2**.
11. In **Range_lookup** type **true**.
12. Your Function Arguments dialog should look like the following. Notice in the lower left, you can see the **Formula result**.



Completed VLOOKUP Arguments

13. Click **OK**. You should now see "Mature" in cell E2.
14. Click cell E2.
15. Click the small green square (fill handle) in the cell's lower right corner to copy the VLOOKUP formula down the column.

=VLOOKUP(D2,\$H\$2:I\$5,2,TRUE)

C	D	E	F
B	Age	Segment	
9/23/1978	39	Mature	
2/14/1967	51		
6/6/1990	28		
5/30/1983	33		

Click square to autofill cell

Example 2 – Multiple Excel Sheets & Exact Match

The second scenario also dealt with that same election file. This time there was an extra worksheet for political parties. The voter's party was listed as an alphanumeric value called "Pcode" and not the political party.

This coding wasn't intuitive. For example, "D" was for "American Independent Party", but some thought it meant "Democratic Party". Another difference was we needed an exact match for Party.

	A	B	C	D	E	F
1	First	Last	Pcode			
2	Smith	Fred	A			
3	Robbins	Terry	1			
4	O'Neill	Susan	B			
5	Parker	Scott	D			
6	Perkins	Ralph	D			
7	Talbot	Angie	7			
8						
9						

Non-intuitive column

needing lookup value

Again, the way to solve this problem was to use the worksheet with the **Pcode** and translation and have Excel use the VLOOKUP function for the **Party name**. I could then add a column called "Political Party" to my original worksheet to show the information from the lookup table.

Using the Starting VLOOKUP Example File

1. Download the [starting Excel sample file](#). The file link is at the bottom of this tutorial.
2. Review the **Example 2 – Voters** worksheet. It has voter first and last names, but only a PCODE.

- Review the **Example 2 -Party Codes** worksheet. It has a listing of party codes and political names. Each of the Party Codes and Names are unique. You'll also note that Column A is in **sorted in ascending order**.

	A	B
1	PARTY CODE	NAME
20	A	Democratic
21	B	Republican
22	C	Decline to State
23	D	American Independent
24	E	Citizen Party
25	F	Communist
26	G	Conservative
27	H	Environmental
28	I	Ind. Progressive
29	J	Liberal

Worksheet with Translated Code Names

- Add your new column on the **Voters** worksheet that will display the info pulled from the Lookup table on the **Party Codes** worksheet. In my example, I added a column called Political Party in Column D. This is where I will insert the Excel function.

	A	B	C	D	E
1	First	Last	Pcode	Political Party	
2	Smith	Fred	A		
3	Robbins	Terry	1		
4	O'Neill	Susan	B		
5	Parker	Scott	D		
6	Perkins	Ralph	D		
7	Talbot	Angie	7		
8					
9					

Add a new Excel column for lookup values

- Place your cursor in the first blank cell in that column. In my example, this is cell D2.
- Follow the Steps 3-7 from Example 1.

Defining the Argument Values

After you click **OK**, Excel's **Function Arguments** dialog appears and allows you to define the four values. You'll see that your starting cell and the formula bar show the beginning part of the function

=VLOOKUP(). The Function Arguments dialog adds the needed data elements that will display between ().

For illustration purposes, I have overlaid the Party Codes worksheet on top to show the relationships.

The screenshot shows the Excel interface with the VLOOKUP function arguments dialog box open. The dialog box has four numbered callouts (1-4) pointing to the fields: 1. Lookup_value, 2. Table_array, 3. Col_index_num, and 4. Range_lookup. Red arrows connect these callouts to the corresponding data in the worksheet: 'A' in cell C2, the range B20:C35, the number '1' in cell C3, and the range B20:C35. A green arrow points from the 'Table_array' field to the 'PARTY CODE' header in the Party Codes worksheet.

First	Last	Pcode	Political Party
Smith	Fred	A	=VLOOKUP()
Robbins	Terry	1	
O'Neill	Susan	B	
Parker	Scott	D	
Perkins	Ralph	D	
Talbot	Angie	7	

PARTY CODE	NAME
A	Democratic
B	Republican
C	Decline to State
D	American Independent
E	Citizen Party
F	Communist
G	Conservative
H	Environmentalist
I	Ind. Progressive
J	Liberal
K	Peace & Freedom
L	Prohibition
M	New Economy
N	Socialist
O	Socialist Labor
P	Pot Party

Mapping the VLOOKUP Function Arguments

After entering the required arguments, my dialog looks like the example below.

The screenshot shows an Excel spreadsheet with a VLOOKUP formula in cell D2. The formula bar displays `=VLOOKUP(C2,'Party Codes'!A2:B45,2,False)`. The spreadsheet data is as follows:

	A	B	C	D	E	F	G	H	I	J	K
1	First	Last	Pcode	Political Party							
2	Smith	Fred	A	=VLOOKUP(C2,'Party Codes'!A2:B45,2,False)							
3	Robbins	Terry	1								
4	O'Neill	Susan	B								
5	Parker	Scott	D								
6	Perkins	Ralph	D								
7	Talbot	Angie	7								

The Function Arguments dialog box for VLOOKUP is open, showing the following arguments:

- Lookup_value: C2 = "A"
- Table_array: 'Party Codes'!A2:B45 = {1,"Green";2,"Reform";3,"Whig";4,"Islar"
- Col_index_num: 2 = 2
- Range_lookup: False = FALSE

The dialog box also includes a description of the function and a 'Formula result' field showing 'Democratic'.

Mapped Values from Both Worksheets

You can see in the red outlined formula bar above, I now have more information based on my entries in the **Function Arguments dialog box**. You might also note that when I clicked the Party Codes worksheet to add in my Table_array, Excel prepended the tab name before the cell range. However, I need to go back and enter my \$ signs.

The other item of interest is that when you build these functions, Excel displays the result in the **Formula result** = text line. This is great feedback which can show if your function is on target. In our example, we can see Excel looked up the Pcode of "A" and returned the Political Party "Democratic".

VLOOKUP is a powerful Excel function that can leverage spreadsheet data from other sources. There are many ways you can benefit from this function. In this example, I used a 1:1 code translation, but you could also use it for group assignments. For example, you could assign state codes to a region such as CT, VT, and MA to a region called "New England".

One important note about using Excel functions and formulas is you want to be careful when deleting columns. For example, in the final file I distributed, I omitted the Age column. So, after I completed my VLOOKUP and got my segments, I copied the cell values to a new Excel worksheet. If I had just deleted Column D, my Excel formula would've returned an error.

If you're trying to do a horizontal lookup, you'll be happy to learn that [Excel has a HLOOKUP function](#). I haven't done an HLOOKUP tutorial yet. If this interests you, let me know. However, Microsoft has released [a new lookup function called XLOOKUP](#) which is very versatile.

9. 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.

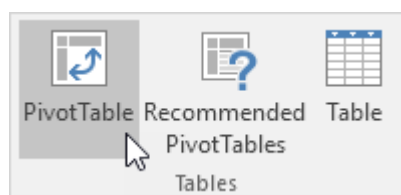
Our data set consists of 213 records and 6 fields. Order ID, Product, Category, Amount, Date and Country.

	A	B	C	D	E	F	G	H
1	Order ID	Product	Category	Amount	Date	Country		
2	1	Carrots	Vegetables	\$4,270	1/6/2016	United States		
3	2	Broccoli	Vegetables	\$8,239	1/7/2016	United Kingdom		
4	3	Banana	Fruit	\$617	1/8/2016	United States		
5	4	Banana	Fruit	\$8,384	1/10/2016	Canada		
6	5	Beans	Vegetables	\$2,626	1/10/2016	Germany		
7	6	Orange	Fruit	\$3,610	1/11/2016	United States		
8	7	Broccoli	Vegetables	\$9,062	1/11/2016	Australia		
9	8	Banana	Fruit	\$6,906	1/16/2016	New Zealand		
10	9	Apple	Fruit	\$2,417	1/16/2016	France		
11	10	Apple	Fruit	\$7,421	1/16/2016	Canada		

Insert a Pivot Table

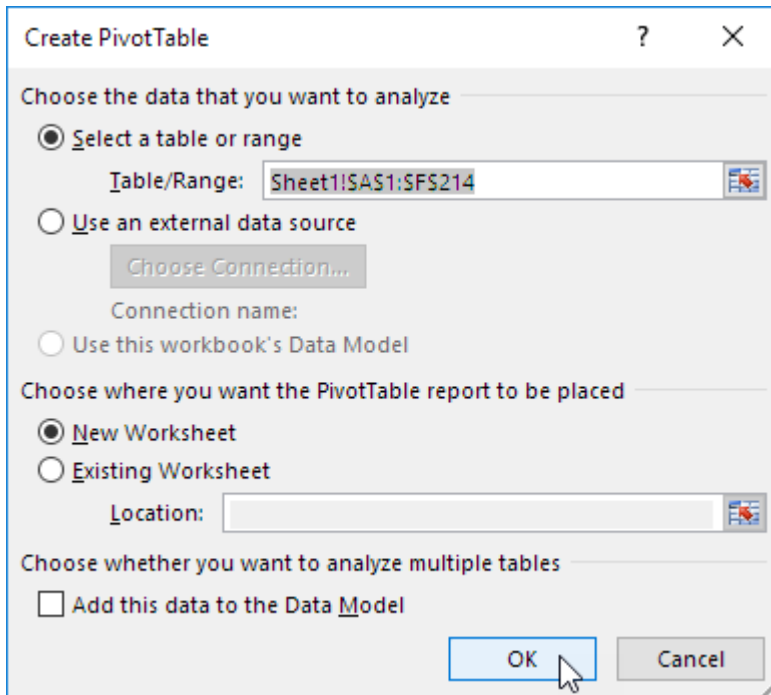
To insert a **pivot table**, execute the following steps.

1. Click any single cell inside the data set.
2. On the Insert tab, in the Tables group, 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.

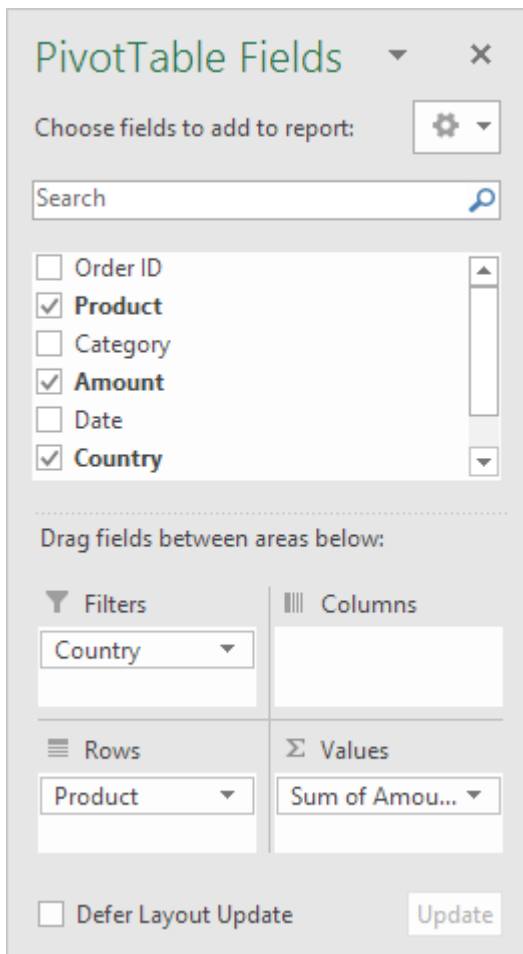
3. Click OK.



Drag fields

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

1. Product field to the Rows area.
2. Amount field to the Values area.
3. Country field to the Filters area.



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

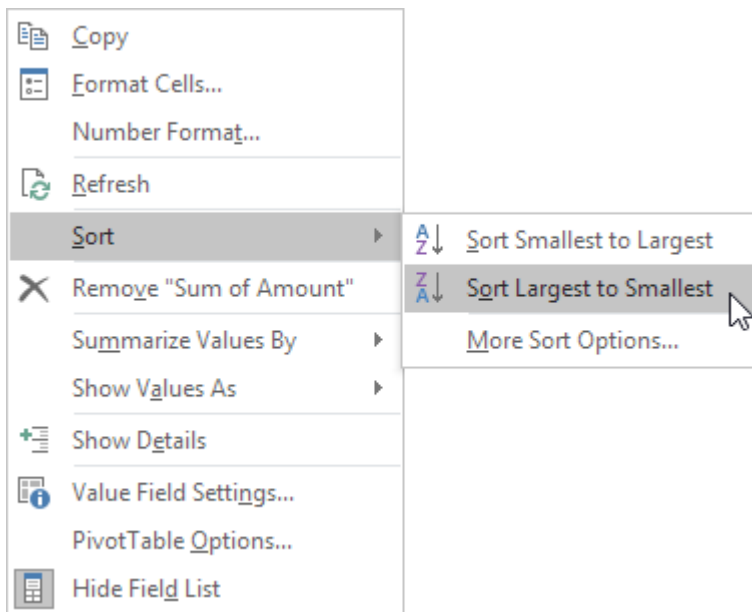
	A	B	C
1	Country	(All)	
2			
3	Row Labels	Sum of Amount	
4	Apple	191257	
5	Banana	340295	
6	Beans	57281	
7	Broccoli	142439	
8	Carrots	136945	
9	Mango	57079	
10	Orange	104438	
11	Grand Total	1029734	
12			

Sort

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

1. Click any cell inside the Sum of Amount column.

2. Right click and click on Sort, Sort Largest to Smallest.



Result.

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

Filter

Because we added the Country field to the Filters 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.

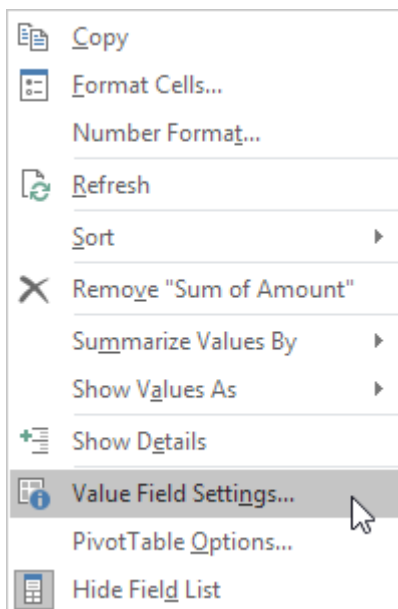
	A	B	C
1	Country	France	
2			
3	Row Labels	Sum of Amount	
4	Apple	80193	
5	Banana	36094	
6	Carrots	9104	
7	Mango	7388	
8	Broccoli	5341	
9	Orange	2256	
10	Beans	680	
11	Grand Total	141056	
12			

Note: you can use the standard filter (triangle next to Row Labels) to only show the amounts of specific products.

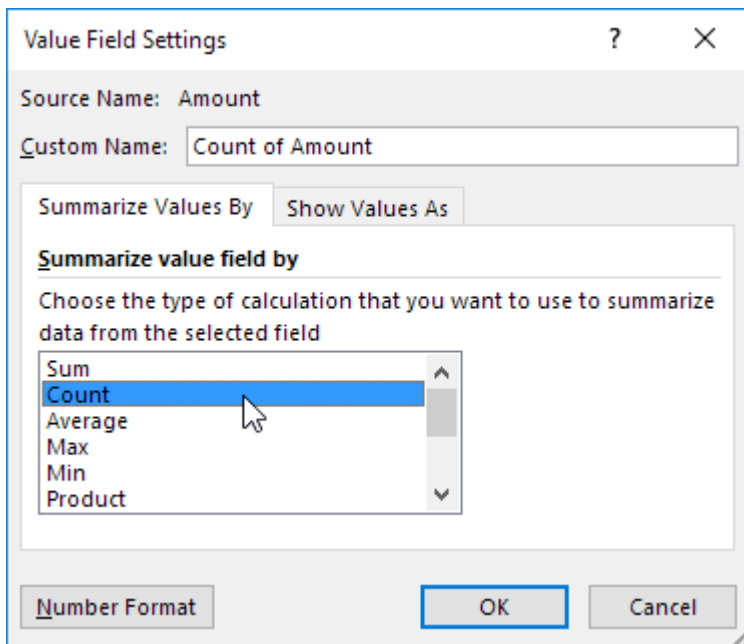
Change Summary Calculation

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 Sum of Amount 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.

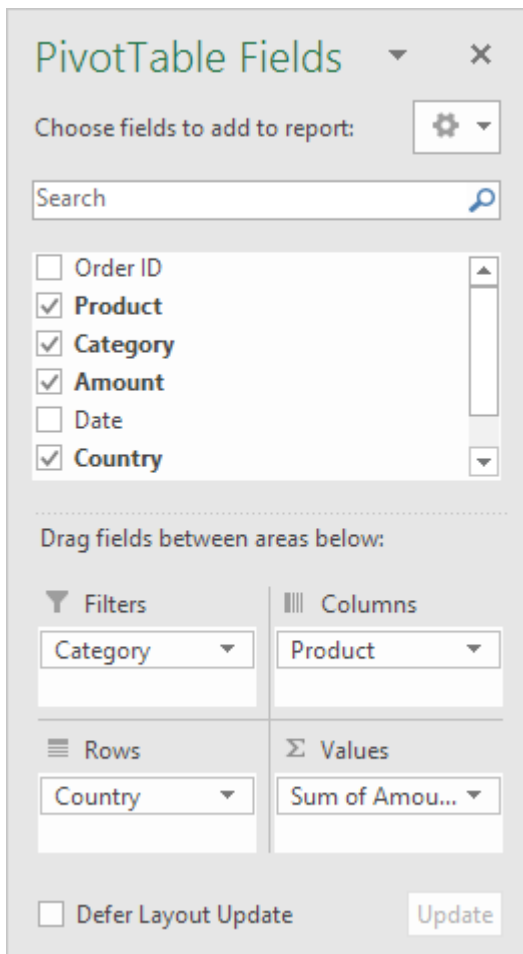
Result. 16 out of the 28 orders to France were 'Apple' orders.

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

Two-dimensional Pivot Table

If you drag a field to the Rows area and Columns area, you can create a two-dimensional pivot table. First, insert a pivot table. Next, to get the total amount exported to each country, of each product, drag the following fields to the different areas.

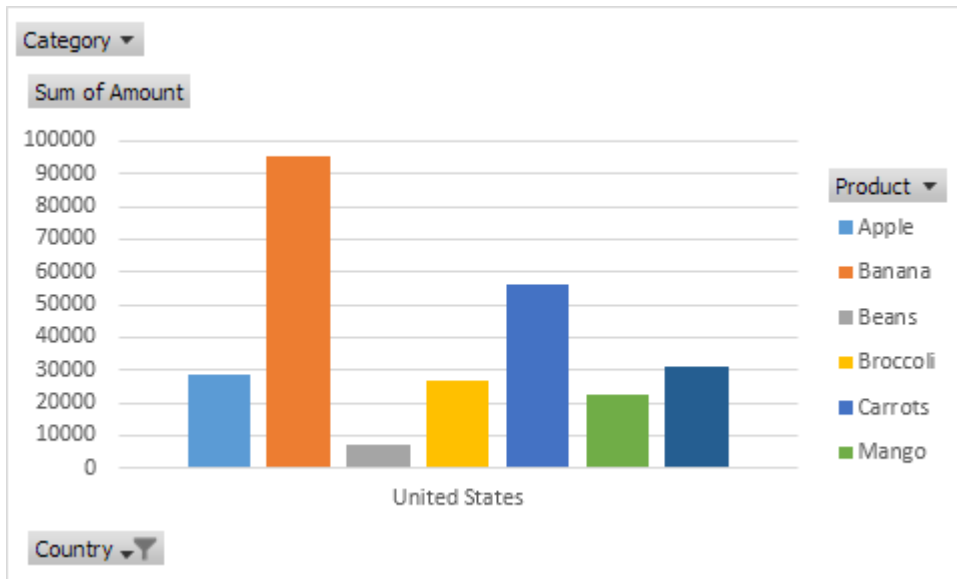
1. Country field to the Rows area.
2. Product field to the Columns area.
3. Amount field to the Values area.
4. Category field to the Filters area.



Below you can find the two-dimensional pivot table.

	A	B	C	D	E	F	G	H	I	J
1	Category	(All)								
2										
3	Sum of Amount	Column								
4	Row Labels	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										

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.



10. Find Quick Answers With Excel Goal Seek

Last Updated on 30-Mar-2017 by [AnneH](#)

The days after elections people often ask “**what-if**” questions. Some races are close and the vote count goes on for days. In other cases, people want to know how many votes would’ve been needed to win. **Excel Goal Seek** is good tool for answering these and other forecasting questions.

What is Goal Seek?

Goal Seek is a built-in Excel tool that allows you to see how one data item in a formula impacts another. You might look at these as “cause and effect” scenarios. For example, you might be looking at your local election results and see:

	Votes	% of Votes
YES	4478	63.90 *
NO	2530	36.10
Total	7008	100

* Needs approval from 2/3 of the voters

In our example, the YES votes are a majority, but shy of the required 2/3 approval to win the election. People quickly realize they were close, but which item do they change to find out how close. What would’ve made a difference?

Using Goal Seek we can change the value of various cells to see how the results change. This would allow you to answer these types of questions.

- *How many “NO” voters needed to be converted to YES to win the election?*
- *How many more votes were needed by the YES team to win the election?*
- *If 500 more people voted could the YES team have won?*

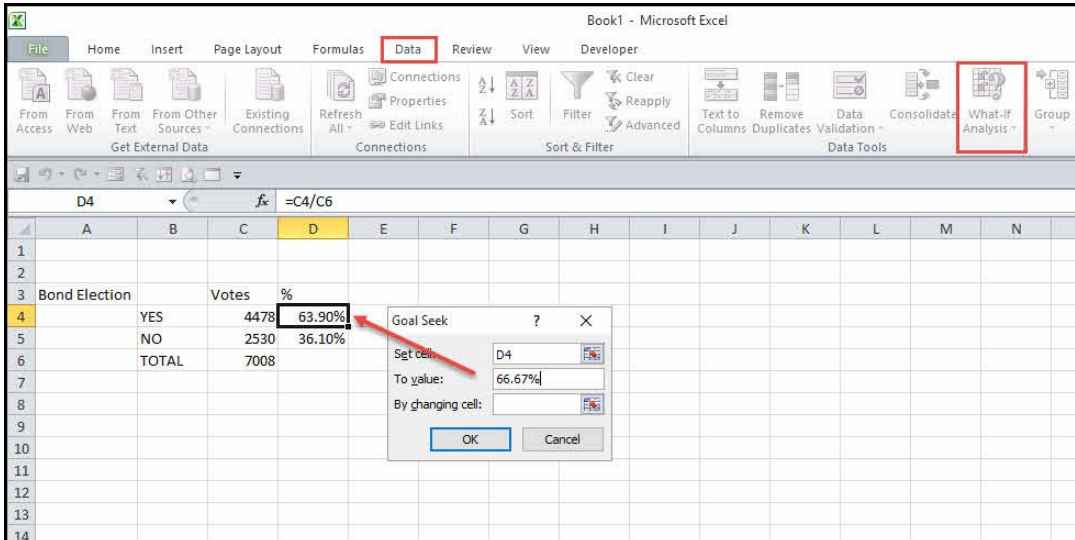
In each of these questions, the goal is to change a data value to see if the YES percentage went over that two-thirds mark or 66.67%. Rather than haphazardly changing cell values to see the results, Goal Seek can find the answers.

How to Use Excel Goal Seek

1. Create a spreadsheet in Excel that has your data.

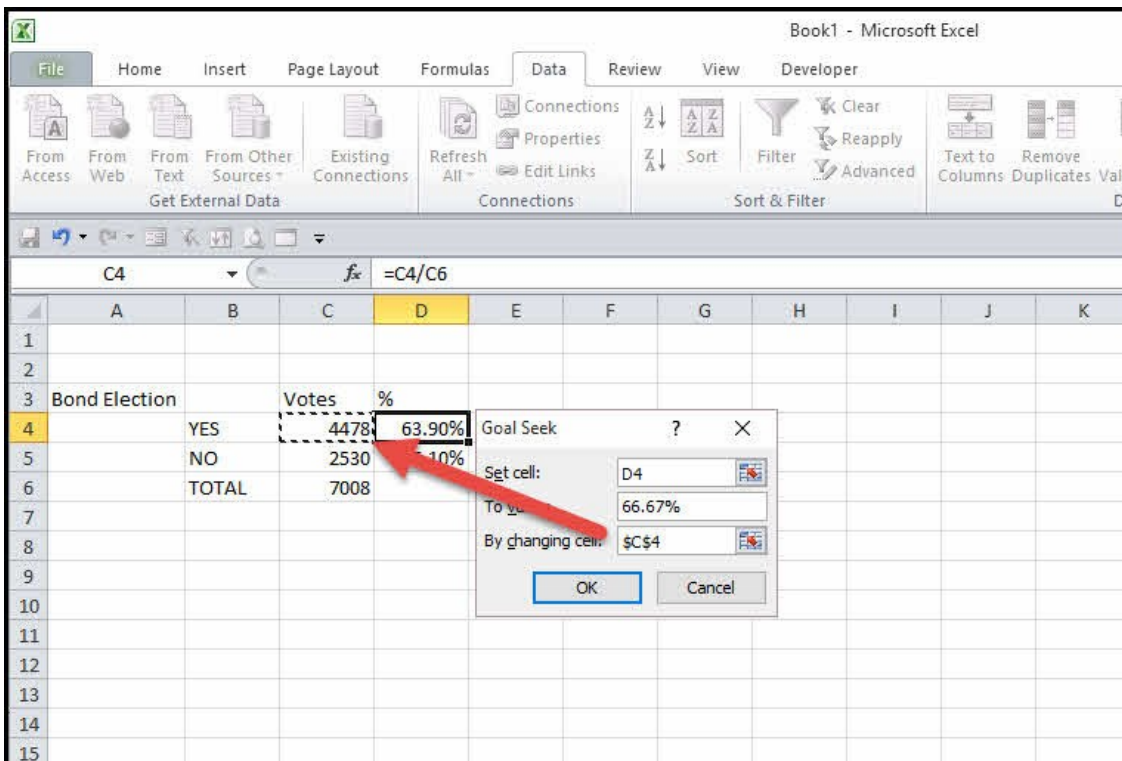
	A	B	C	D	E
1					
2					
3	Bond Election		Votes	%	
4		YES	4478	63.90%	
5		NO	2530	36.10%	
6		TOTAL	7008		
7					
8					

- Click the cell you want to change. This is called the “**Set cell**”. In my example, this will be D4.
- From the **Data** tab, select the **What if Analysis...** button
- Select **Goal seek..** from the drop down menu
- In the Goal Seek dialog, enter the new “what if” amount in the **To value** text box. (Remember to add the percentage sign if you have one.)

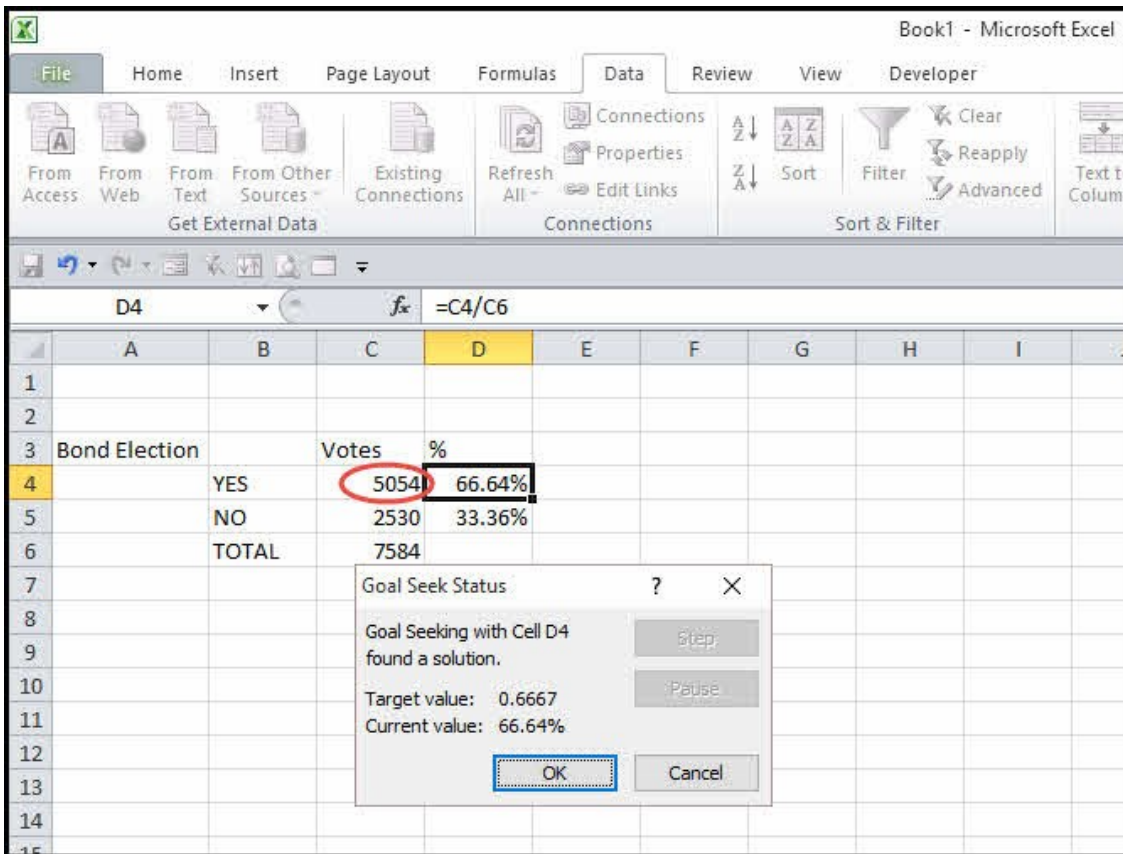


In this example, we’re asking Excel to replace the contents of cell D4 which is 63.90% with 66.67%. This is the percentage needed to win the election.

- We also need to tell Excel which cell to change. Since we wanted to know the number of YES votes, we’ll click C4.



- Click **OK**. Excel will overwrite the previous cell value with the new one.



8. If you wish to accept the new value, click **OK**.

Goal Seek Requirements & Gotchas

I can hear some readers yelling “wait.... don’t accept the value”. Actually, people may be yelling for two reasons.

If you look closely in the Goal Seek Status box, you’ll see the **Target value:** reads: 0.6667 came from our 66.67%. But the **Current value:** shows as 66.64%. Excel rounded down and the revised YES value wouldn’t win the election since 66.64% is less than 66.67%.

Another group may be asking why Excel didn’t just take 66.67% of the TOTAL vote count 7008. And that leads us to a requirement of Goal Seek – formulas. **The Set cell in Step 2 must contain a formula.** In our example, the initial 63.90% was a formula using the YES value (C4) divided by TOTAL value (C6). So, we met that requirement.

The screenshot shows the Microsoft Excel interface with the Data tab selected. The formula bar shows $=C4/C6$ for cell D4. The spreadsheet data is as follows:

	A	B	C	D	E
1					
2					
3	Bond Election		Votes	%	
4		YES	4478	63.90%	
5		NO	2530	36.10%	
6		TOTAL	7008		
7					

The Goal Seek Status dialog box is open, showing:

- Goal Seeking with Cell D4 found a solution.
- Target value: 0.6667
- Current value: 63.90%

The OK button is highlighted with a blue border.

If I just typed the value 63.90% and tried to use Goal Seek, I would see an Excel error stating “**Cell must contain a formula**”.

Just as I used a formula in my set cell, I also used a formula for TOTAL in B6. This was the Sum of my YES and NO votes. Since this TOTAL cell was a formula, my TOTAL count automatically adjusted when Goal Seek changed the YES cell value. If I typed “7008” in cell C6, only the YES cell value would have changed.

The other requirement is **the cell you change in Step 6 can’t contain a formula**. It must be a typed value.

As this example shows, Goal Seek is a nice tool that can quickly find the answers to different situations. Although this was a simple example, you can use the same tool for complicated spreadsheets containing many variables and formulas. Just remember, you may not always like the results.

11. Paired Sample t Test

Charles Zaiontz

Motivations

In **paired sample** hypothesis testing, a sample from the population is chosen and two measurements for each element in the sample are taken. Each set of measurements is considered a sample. Unlike the hypothesis testing studied so far, the two samples are not independent of one another. Paired samples are also called **matched samples** or **repeated measures**.

For example, if you want to determine whether drinking a glass of wine or drinking a glass of beer has the same or different impact on memory, one approach is to take a sample of say 40 people, and have half of them drink a glass of wine and the other half drink a glass of beer, and then give each of the 40 people a memory test and compare results. This is the approach with **independent** samples.

Another approach is to take a sample of 20 people and have each person drink a glass of wine and take a memory test, and then have the same people drink a glass of beer and again take a memory test; finally we compare the results. This is the approach used with **paired** samples.

The advantage of this second approach is the sample can be smaller. Also since the sampled subjects are the same for beer and wine there is less chance that some external factor (**confounding variable**) will influence the result. The problem with this approach is that it is possible that the results of the second memory test will be lower simply because the person has imbibed more alcohol. This can be corrected by sufficiently separating the tests, e.g. by conducting the test with beer a day after the test with wine.

It is also possible that the order in which people take the tests influences the result (e.g. the subjects learn something on the first test that helps them on the second test, or perhaps taking the test the second time introduces a degree of boredom that lowers the score). One way to address these **order effects** is to have half the people drink wine on day 1 and beer on day 2, while for the other half the order is reversed (called **counterbalancing**).

The following table summarizes the advantages of paired samples versus independent samples:

Paired Samples	Independent Samples
Need fewer participants	Fewer problems with fatigue or practice effects
Greater control over confounding variables	Participants are less likely to figure out the purpose of the study

Figure 1 – Comparison of independent and paired samples

Obviously not all experiments can use the paired sample design. E.g. if you are testing differences between men and women, then independent samples will be necessary.

How to perform paired t-test with Data analysis tool excel

Excel provides a data analysis tool that takes care of just about everything for matched samples. It's called t-test: Paired Two Sample for Means. In this section, I use it on the weight-loss data.

Null hypothesis: there is not loss of weight

Activate the data analysis tool

If you need to develop complex statistical or engineering analyses, you can save steps and time by using the Analysis ToolPak. You provide the data and parameters for each analysis, and the tool uses the appropriate statistical or engineering macro functions to calculate and display the results in an output table. Some tools generate charts in addition to output tables.

The data analysis functions can be used on only one worksheet at a time. When you perform data analysis on grouped worksheets, results will appear on the first worksheet and empty formatted tables will appear on the remaining worksheets. To perform data analysis on the remainder of the worksheets, recalculate the analysis tool for each worksheet.

1. Click the **File** tab, click **Options**, and then click the **Add-Ins** category.
2. In the **Manage** box, select **Excel Add-ins** and then click **Go**.

If you're using Excel for Mac, in the file menu go to **Tools > Excel Add-ins**.

3. In the **Add-Ins** box, check the **Analysis ToolPak** check box, and then click **OK**.
 - If **Analysis ToolPak** is not listed in the **Add-Ins available** box, click **Browse** to locate it.
 - If you are prompted that the Analysis ToolPak is not currently installed on your computer, click **Yes** to install it.

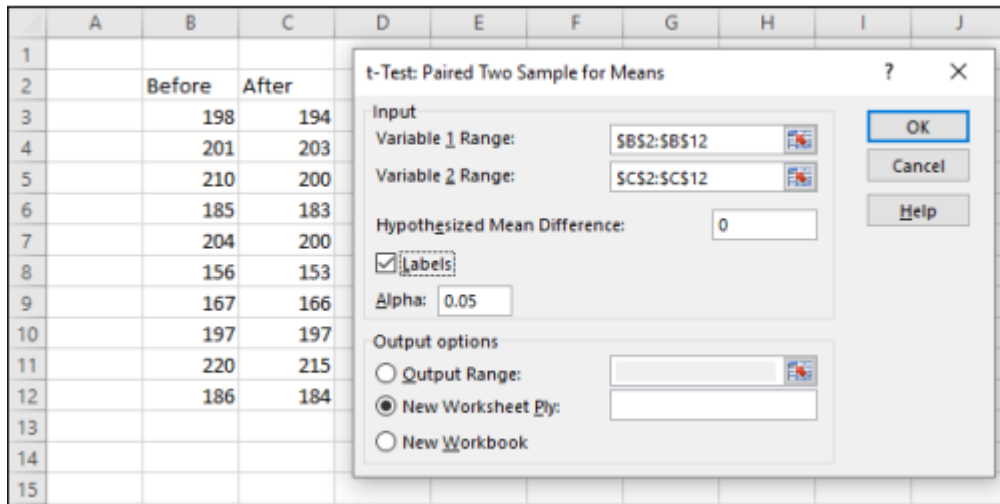
How to perform the t-test

Here are the steps to follow:

1. Enter the data for each sample into a separate data array.
For this example, the data in the Before sample are in column B and the data for the After sample are in column C.
2. Select **Data | Data Analysis** to open the Data Analysis dialog box.
3. In the Data Analysis dialog box, scroll down the Analysis Tools list, and select **t-Test: Paired Two Sample for Means**.

4. 4. Click OK to open this tool's dialog box.

This is the dialog box



5. In the Variable 1 Range box, enter the cell range that holds the data for one of the samples. For the example, the Before data are in \$B\$2:\$B\$12, including the heading.

(Note the \$ signs for absolute referencing.)

6. In the Variable 2 Range box, enter the cell range that holds the data for the other sample.

The After data are in \$C\$2:\$C\$12, including the heading.

7. In the Hypothesized Mean Difference box, type the difference between μ_1 and μ_2 that H_0 specifies.

In this example, that difference is 0.

8. If the cell ranges include column headings, select the Labels check box.

I included the headings in the ranges, so I selected the check box.

9. The Alpha box has 0.05 as a default. Change that value if you want to use a different α .

10. In the Output Options, select a radio button to indicate where you want the results. I selected New Worksheet Ply to put the results on a new page in the work sheet.

11. Click OK.

Because I selected New Worksheet Ply, a newly created page opens with the results.

	A	B	C
1	t-Test: Paired Two Sample for Means		
2			
3		<i>Before</i>	<i>After</i>
4	Mean	192.4	189.5
5	Variance	377.6	342.9444444
6	Observations	10	10
7	Pearson Correlation	0.986507688	
8	Hypothesized Mean Difference	0	
9	df	9	
10	t Stat	2.824139508	
11	P(T<=t) one-tail	0.009955992	
12	t Critical one-tail	1.833112933	
13	P(T<=t) two-tail	0.019911984	
14	t Critical two-tail	2.262157163	
15			

Pearson Correlation Coefficient

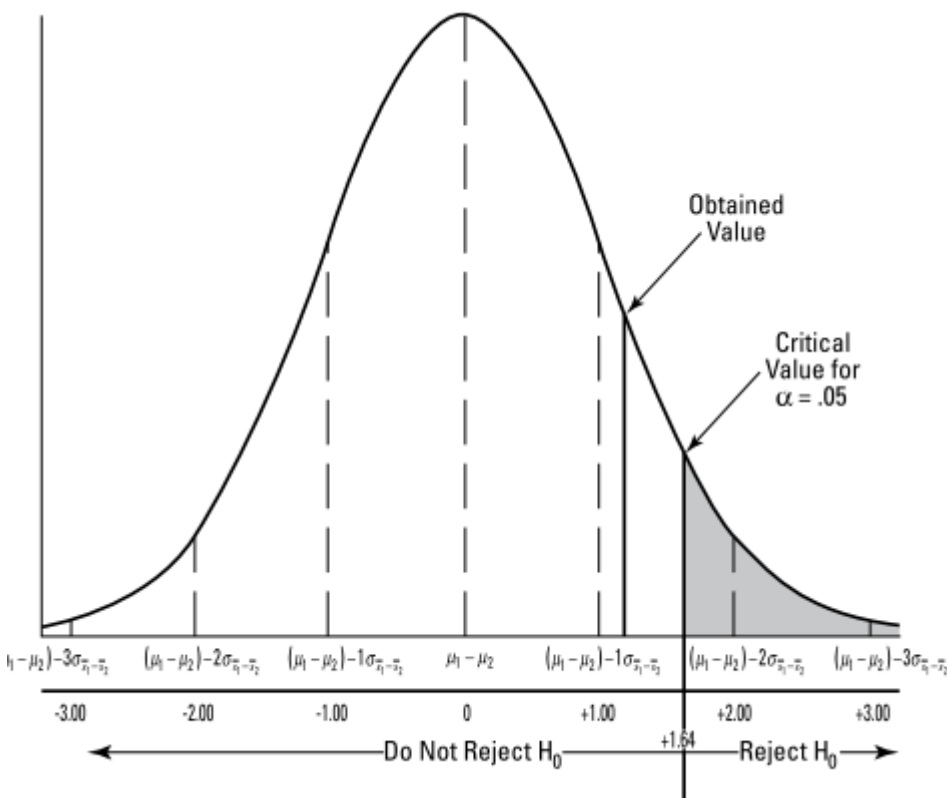
The only item that's new is the number in cell B7, the Pearson Correlation Coefficient. This is a number between -1 and $+1$ that indicates the strength of the relationship between the data in the first sample and the data in the second. If the number is close to -1 , high scores in the first sample are associated with low scores in the second, and low scores in the first are associated with high scores in the second. If the number is close to zero, scores in the first sample are unrelated to scores in the second. Because the two samples consist of scores on the same people, you expect a high value.

Cell B8 shows the H_0 -specified difference between the population means, and B9 shows the degrees of freedom.

The remaining rows provide t -related information. The calculated value of the test statistic is in B10. Cell B11 gives the proportion of area the positive value of the test statistic cuts off in the upper tail of the t -distribution with the indicated df (gdl in italiano).

Cell B12 gives the critical value for a one-tailed test: That's the value that cuts off the proportion of the area in the upper tail equal to α .

Cell B13 doubles the proportion in B11. This cell holds the proportion of area from B11 added to the proportion of area that the negative value of the test statistic cuts off in the lower tail. Cell B13 shows the critical value for a two-tailed test: That's the positive value that cuts off $\alpha/2$ in the upper tail. The corresponding negative value (not shown) cuts off $\alpha/2$ in the lower tail.



In brief:

If $p > \alpha$, we reject the H_0

If $p < \alpha$ we cannot reject H_0