30 Excel Functions Every Analyst Should Know

Prepared For
IALEIA SWOC
November Excel Workshop

November 28, 2012

By
Joseph Ryan Glover
Police Analyst.com

Contact
416-538-0659
joseph.ryan.glover@gmail.com
www.policeanalyst.com

Table of Contents

30 Excel Functions Every Analyst Should Know

Crime and intelligence analysts spend a lot of time in Excel so it's important that we know how to make Excel do as much for us as it can. The key to increased productivity in Excel is leveraging the power of functions. This workshop presents a brief overview of functions in Excel and lays out 30 Excel functions, with relevant examples, that analysts can use in their day-to-day work. The hope is that we'll move through the functions as a group, trying them on our own computers as we go, and then we'll work together on the exercises. I have always found that typing out functions and seeing the immediate results is a great way to learn Excel. Good luck and have fun.

What is a Function?

A function in Excel allows you to perform an action inside of a spreadsheet cell similar to the buttons on a scientific calculator. Like the calculator buttons, functions allow for complex manipulation of your data by employing pre-packaged functionality built into Excel.

To call a function you first type an equals sign "=" and then the function name directly into a cell. Most functions take one or more input values and the inputs are usually, but not always, values stored in other cells. This is where the practice of referring to cells by their column and row comes in handy. For example, to find the length of the string in cell A1 you can type the LEN function into cell B1 and pass in A1 as the input like this =LEN(A1). This is called cell-referencing and its important that you learn how it works in order to get the most out of functions.
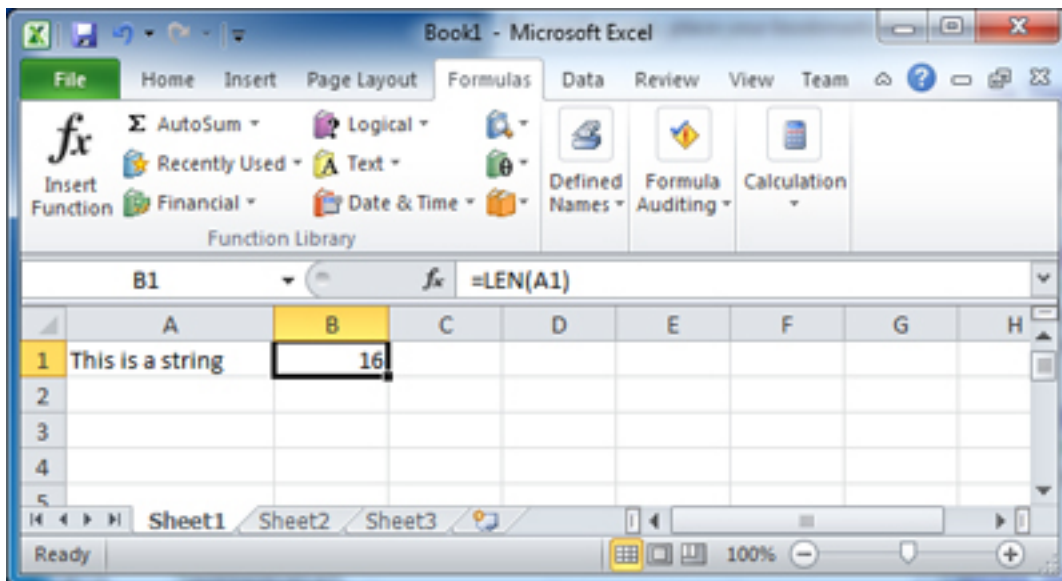
When talking about using functions I'll often refer to "calling" the function. This terminology comes out of the idea that functions are self-contained entities that you ask for a result. This is also where the idea of "passing" inputs to a function comes from; since functions are considered indepedent tools one passes information to them and accepts back the "return" or output value. Think of functions like a really capable assistant: you call them and ask them to do something, they go away and do it and return to you with the result. How they do it doesn't matter, you're just glad you didn't need to calculate the string length yourself.

A final note: somewhat confusingly, functions are also known as formulas in Excel. The names are interchangeable and throughout this document you'll find both uses. Just remember that the two terms are equal.

Text Functions

As analysts we spend a lot of time working with text in Excel and often that text comes to us in a less than ideal format. Excel's text functions are there to help us keep our sanity when our data is a mess.

The LEN Function

As we've already seen in the introduction, the LEN function is very simple; it returns the length of any string passed to it string. You may not see a lot of value in that yet but trust me it becomes valuable when it is combined with other functions.

To demonstrate how it works I'm just going to repaste the screen shot from the introduction. Try typing anything you want in cell A1 and =LEN(A1) in cell B1.
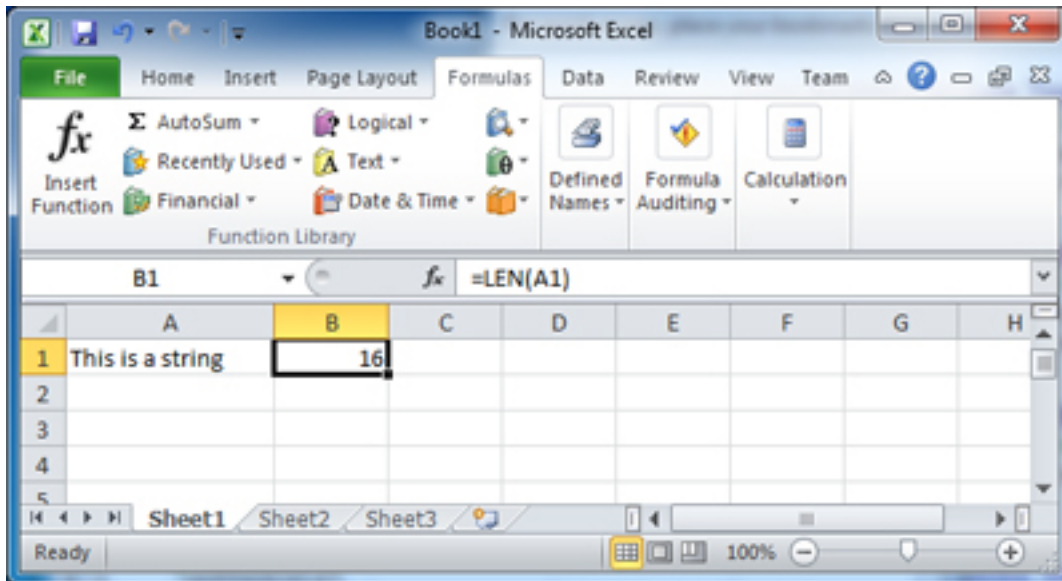
One thing to note is that LEN counts spaces when it determining the length of a string. That may seem obvious when the space is in the middle of the string but it also counts spaces on the ends of strings (see TRIM below for a fix).

The MID, LEFT and RIGHT Functions

The MID, LEFT and RIGHT functions all deal with extracting a substring from a larger string. And while LEFT and RIGHT are easier to wrap your head around they actually aren't important once you learn how to use MID. Look at the following screen shot to see an example of each of the functions in action.
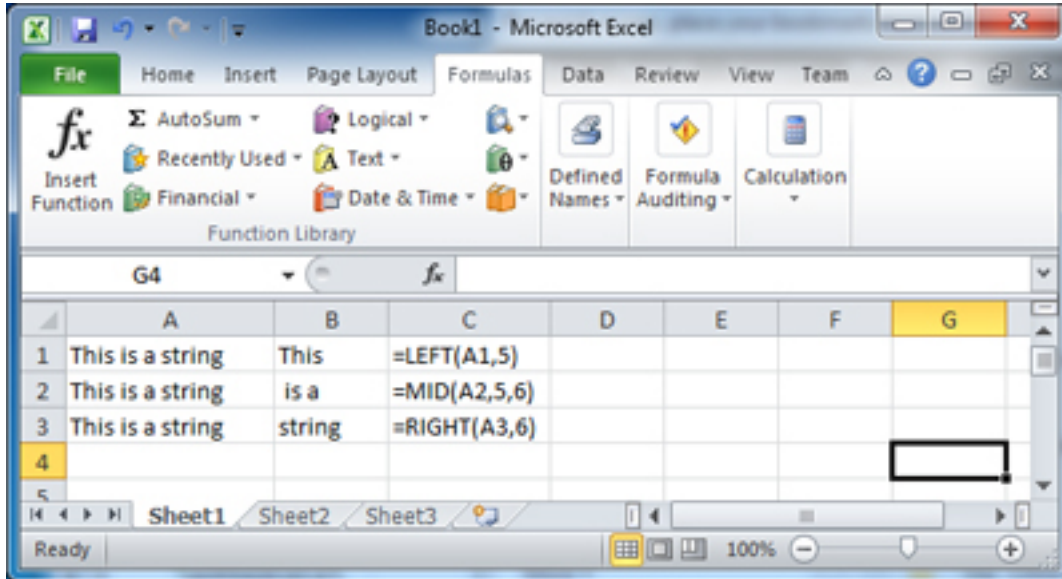
As you can see LEFT and RIGHT both take two inputs: the string that is going to be extracted from and a number that represents how many characters to extract. For LEFT we told the function that we wanted the 5 left-most characters from "This is a string" which, as shown, is the string "This " (with a trailing blank).

Similarily, the RIGHT function wants to know how many characters in from the right we want to extract and in this case 6 characters is enough to extract the word "string".

Finally, MID allows you to extract text from anywhere within the string as long as you also provide a starting position as an input. In the example we told the function to use a starting position of 5, that is 5 characters in from the left, and extract the next 6 characters which resulted in the output " is a " (with two leading blanks and a trailing blank).
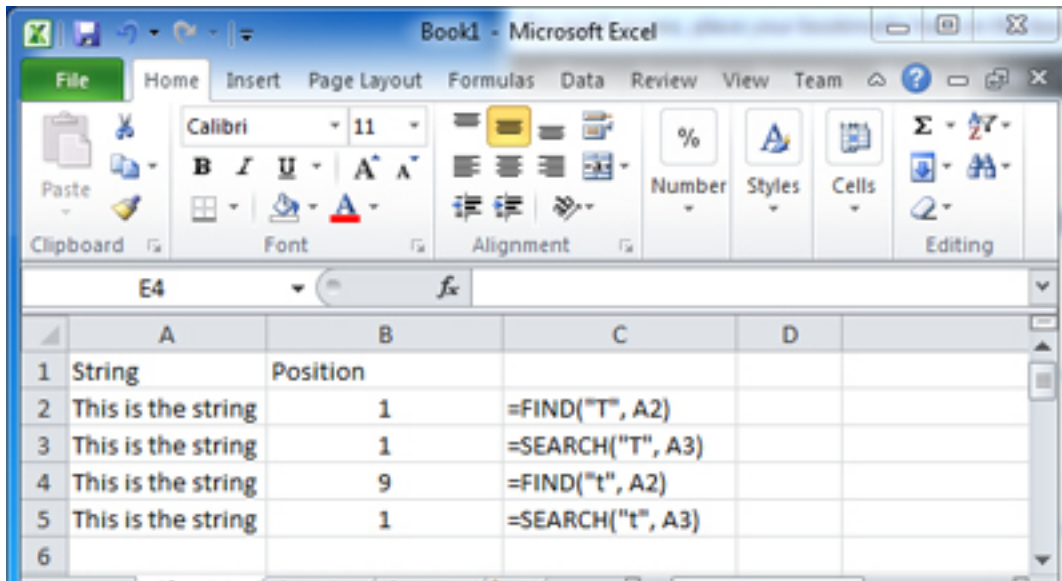
The reason I say that LEFT and RIGHT are unnecessary once you learn MID is that it is possible to duplicate their functionality simply by selecting the proper starting position. For example, we can specify 1 as the beginning point for MID and get the same result as LEFT. Can you figure out how to duplicate RIGHT functionality using MID? (hint: you'll need to use LEN)

One shortcoming of these functions, when they are used in isolation, is that you need to know beforehand exactly how you want to split out the string. The functions SEARCH and FIND discussed below can help with this problem.

The SEARCH and FIND Functions

The SEARCH and FIND functions are both used to locate the position of a substring inside of a larger string. MID, LEFT and RIGHT all require that we specify where and how far we want the function to extract from but if we don't know the precise location of the substring, for example if it changes from cell to cell, we can use SEARCH or FIND to find the location.

There are two fundamental differences between the functions: the first is FIND is case sensitive while SEARCH is not and the second is that SEARCH allows for the use of wild cards such as ? and * while FIND does not. Have a look at the screen shot below that demonstrates the case sensitivity and insensitivity of FIND and SEARCH.


The interesting examples occur in lines 4 and 5. Notice how we are using FIND to locate a lower case "t" and FIND locates it at position 9. SEARCH, however, locates a lower case "t" in position one, which is actually an upper case "T", but since SEARCH ignores case, the answer is correct.

To test wild cards in SEARCH try typing =SEARCH("t?e", A2) into cell B2. You should get the value 9 which is the position at which the word "the" begins. Now try =SEARCH("st*g",A2) and you should get position 13 where the word string begins. The difference between the two wild cards ? and * is that ? represents any single character while * represents one or more characters.

The UPPER, LOWER and PROPER Functions

Continuing in the vein of messy data, or at least data that isn't formatted the way we'd like, are three functions for changing the case of strings. UPPER and LOWER do exactly what their names imply: convert strings to all upper case or lower case but PROPER is a bit more of a mystery. It capitalizes the first letter of every word in a string sort of like a title case.

The following screen shot demonstrates how to use all three functions. They are pretty self explanatory.
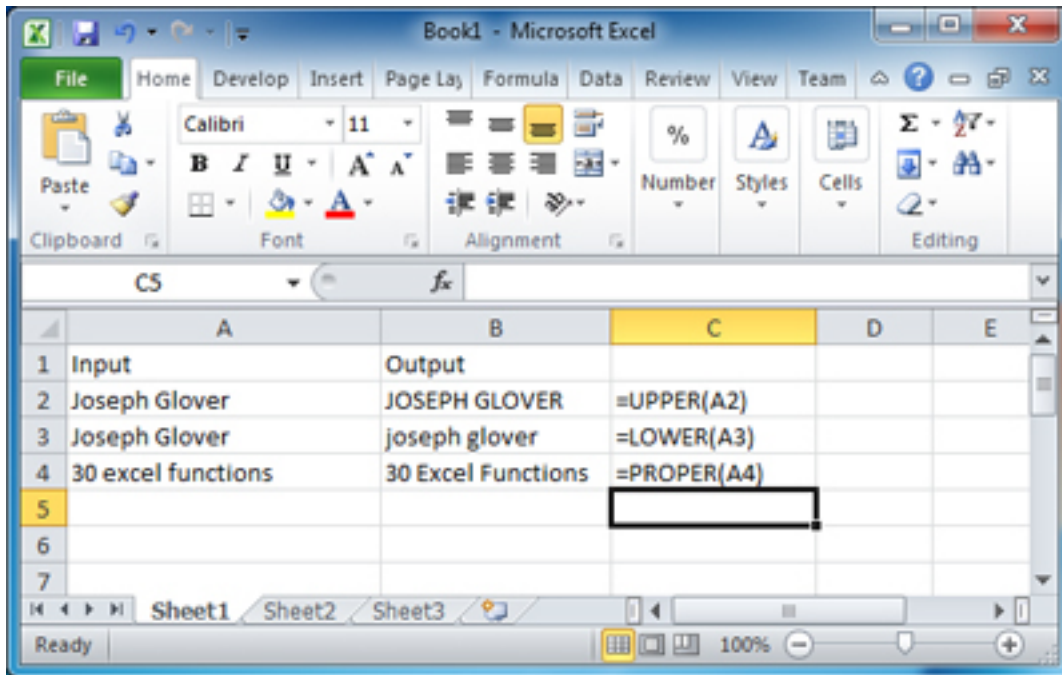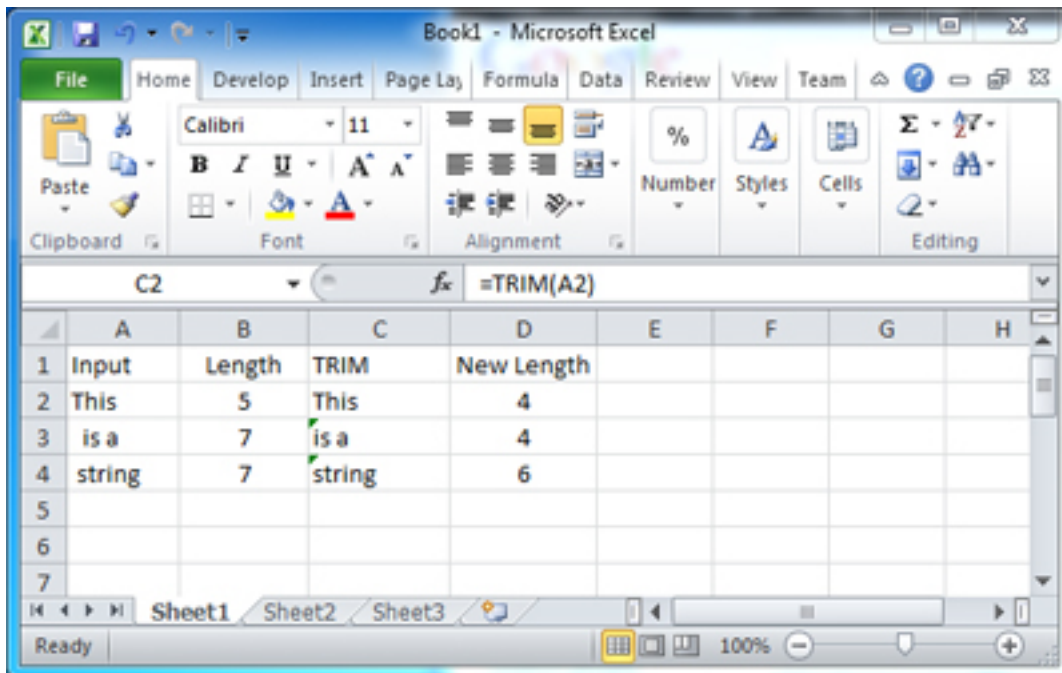

This is usually where I would ask you to try something but these functions are pretty self-explanatory. Remember them though because you'll need to use them for the exercises later.

The TRIM Function

TRIM is a handy function that is used to clean up text that has leading and trailing white space. Often when you get a dump of data from a database the text you're dealing with is padded with blanks and if you don't deal with them they can really throw a wrench into your spreadsheet.

The biggest problem is that Excel treats a white space as a legitimate character and if we're going to use something like a pivot table to count the number of unqiue entries in a list the presence of leading or trailing whitespace will mess up the counts.

For example, in the screen shot below are the strings that were exracted using MID, LEFT and RIGHT a few pages back and the corresponding lengths (done with the LEN function) in column B. In cell C2 is the function =TRIM(A2) and and in column D the length of the trimmed values.


While there isn't much visual difference between column A and C the new length column in D tells the story. TRIM has cleaned off all the leading and trailing white space and now the values in the cells can be compared and tabulated.

The CLEAN Function

The CLEAN function is like TRIM but on steroids. Whereas TRIM removes leading and trailing whitespace characters CLEAN will remove any and all so-called non-printable characters. Now, you may not know what a non-printable character is but I'm sure that you've seen some because they often show up in text that is copied and pasted into Excel from places such as Word or websites. While they can take all kinds of forms the one people are most familiar with are the little squares with the question mark in them.

The CLEAN function is used to remove these characters and it's as simple as putting =CLEAN(A2) in cell B2 to clean the string in A2. See the screen shot below for the details.
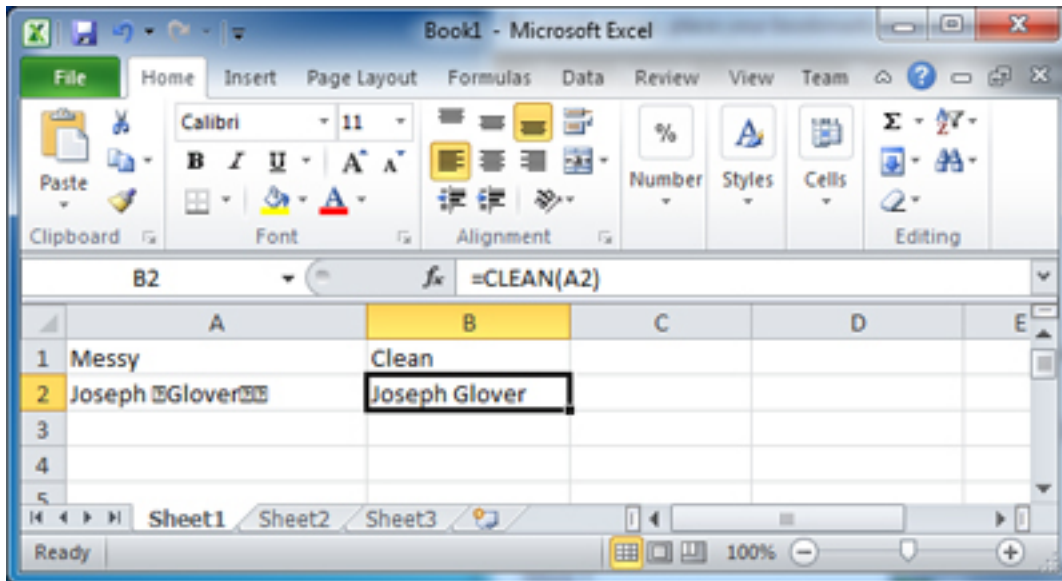
Why would you want to do this? Because Excel functions don't like non-printable characters and they can cause all kinds of weirdness when you use them as input to other string functions.

One note: CLEAN is a very literal function, it just removes the non-printable character, it doesn't replace it with anything so if the non-printable character was seperating two words those words now run together.

The SUBSTITUTE Function

SUBSTITUTE is the string function for performing replacements. I know, I know, most of the time you could use Find and Replace but SUBSTITUTE really shines when you combine it with other functions to decide when and what to replace (but we won't get into that now). To demonstrate how SUBSTITUTE works have a look at the screen shot below where I made a fat-finger typo.
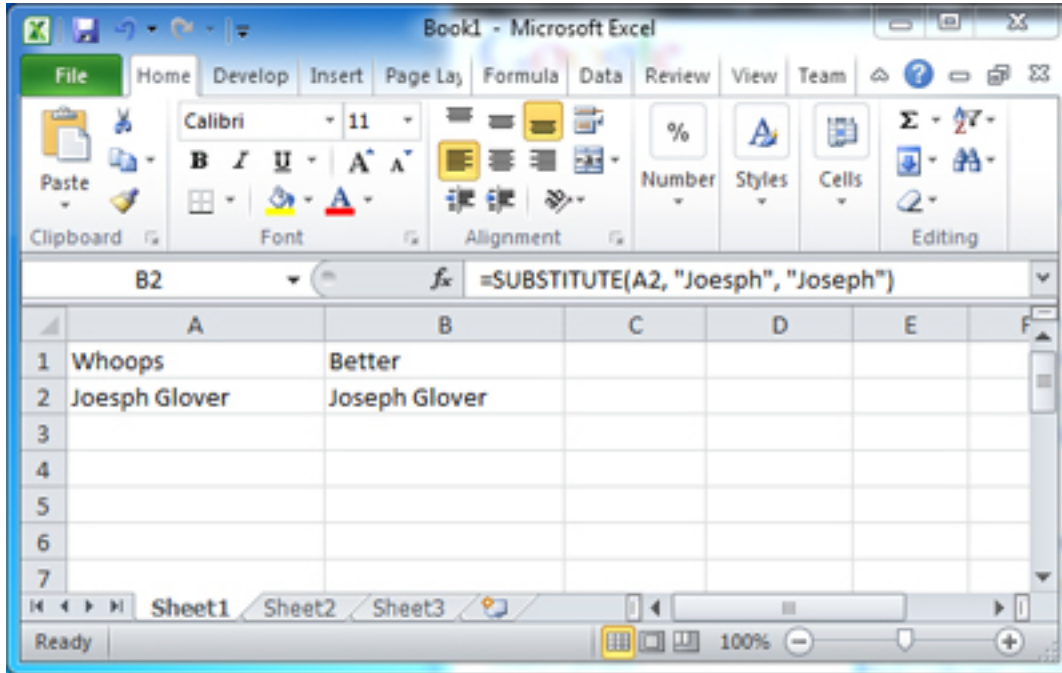

Note that I in cell A2 wrote "Joesph" instead of "Joseph". Whoops. Obviously, when there is only replacement, I could just go in and fix it but that wouldn't demonstrate SUBSTITUTE. To correct the problem I will put in cell B2 =SUBSTITUTE(A2, "Joesph", "Joseph") and as the screen shot reveals, everything is fixed.

The inputs to the SUBSTITUTE function are: the string to be substituted, the string to be replaced, and the string that will serve as the replacement.

The CONCATENATE Function

To concatenate something means to stick it together and in Excel the CONCATENATE function is used to connect together several smaller strings into one large one. CONCATENATE is an interesting function because it takes an open-ended number of inputs. The function needs at least one input (although it doesn't make much sense to concatenate one thing) but it is capable of piecing together up to 30 strings.

In the following screen shot there are two columns, first name and last name, and I want to connect the two columns into one. To do that I type in cell C2 =CONCATENATE(A2,B2) as the screen shot reveals.
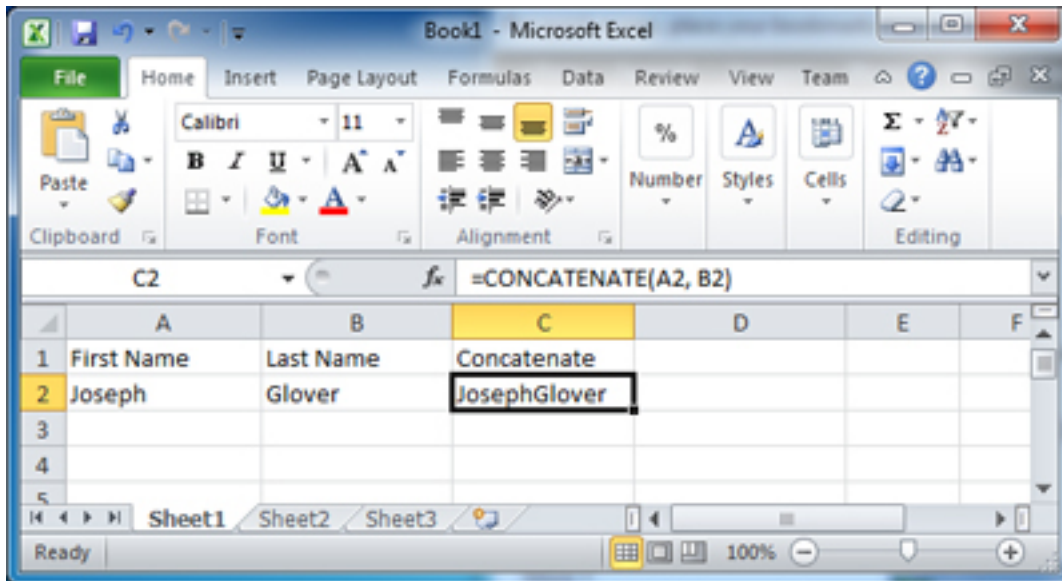

Of course, there's a bit of a problem: no space between first and last name. To handle that we need to modify the CONCATENATE function call to include another string, a blank space, between the first and last names. The revised function in cell C2 should now read =CONCATENATE(A2, " ", B2). This example is helpful because it demonstrates that the inputs to CONCATENATE don't just have to be other cells, you can put literal strings in there as well.

Try creating a new concatenation that creates a string with the form "Lastname, Firstname".

The HYPERLINK Function

You may not have occasion to use the HYPERLINK function everyday but when you do, it's a real gem. Its essential functionality is that it converts the function input into a link while assuming that the text you passed in is a legitimate URL. I use it primarily for making quick links out to Google maps as in the screen shot below.
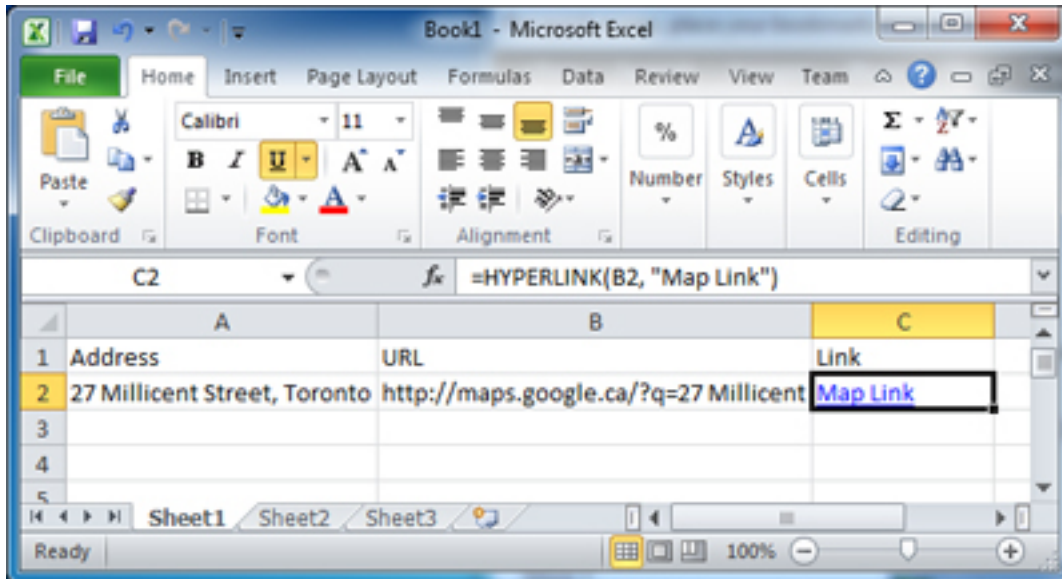

For the above example I used an address and, using CONCATEANTE, created the value in B2 that defines a URL to Google maps with the address as the q or query value. Finally, in C2 I used the HYPERLINK function to create a link by typing =HYPERLINK(B2, "Map Link"). The second value, Map Link, is optional but it offers a chance to replace the full linked URL with some more succinct link text. If you click the link your browser will should open up and take you to the address in Google Maps

In cell A3 put your address and in cell B3 use CONCATENATE to creare the URL to Microsoft Bing Maps instead of Google Maps (the Bing URL prefix is http://maps.bing.com/?q=). In C3 use HYPERLINK to create a clickable link to Bing.

The VALUE Function

Have you ever had a column of numbers that Excel thinks is text? And have you copied and pasted trying to get Excel to recognize that the values as numbers so that you can sort, sum, format, etc. as you would with a regular number? Well, VALUE is your function.

What VALUE does is take a string representation of a number and convert it, in the mind of Excel, into a number. It's very straightforward to use, just type =VALUE(A2) to convert a string number in A2 into a number number in B2. Hint: you can tell that it worked because Excel left justifies text by default and right justifies numbers by default.
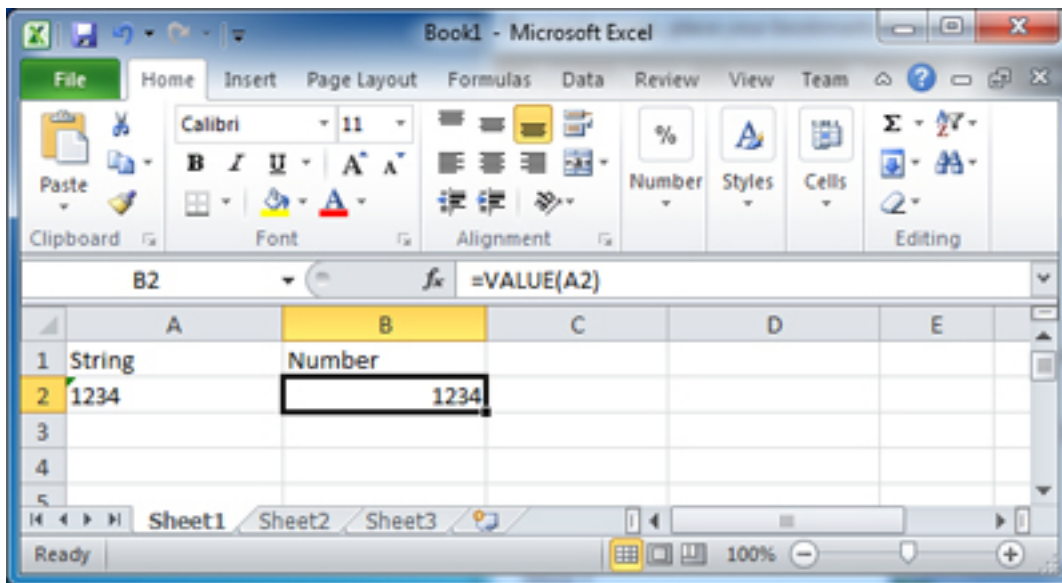
Exercise #1

Ok, phew, that was a lot of functions. But now it's your turn. In the Excel workbook named 30_Functions_Exercises.xlsx are two sheets of interest: the first has a bunch of messy data representing outstanding warrants and the second, named "Exercise 1" has a lovely, albeit limited, summation of the data. It's your job to use the formulas discussed above to transform the raw data into the polished form. You'll see that there are some other sheets in the workbook, those are for later exercises

Hints:

- Feel free to create extra worksheets for scratch work
- You shouldn't need to edit any text by hand. Everything can be done with a function.
- The address out of the database has a weird spacing issue around commas. You should probably clean that up when working with the address.

Date Functions

After text the next most common type of data analysts work with is dates. Dates in Excel can be frustrating if you don't know how they work under the hood. The following functions and exercise should help you get up to speed.

The DATE Function

When a date arrives in Excel as a date there's not usually a problem. The problems arise when cells are filled with values that we know are dates but Excel thinks are strings. In these situations we need to leverage our understanding of Excel text functions to transform date strings into date dates. In the following screen shot there's a date in an annoying string format.
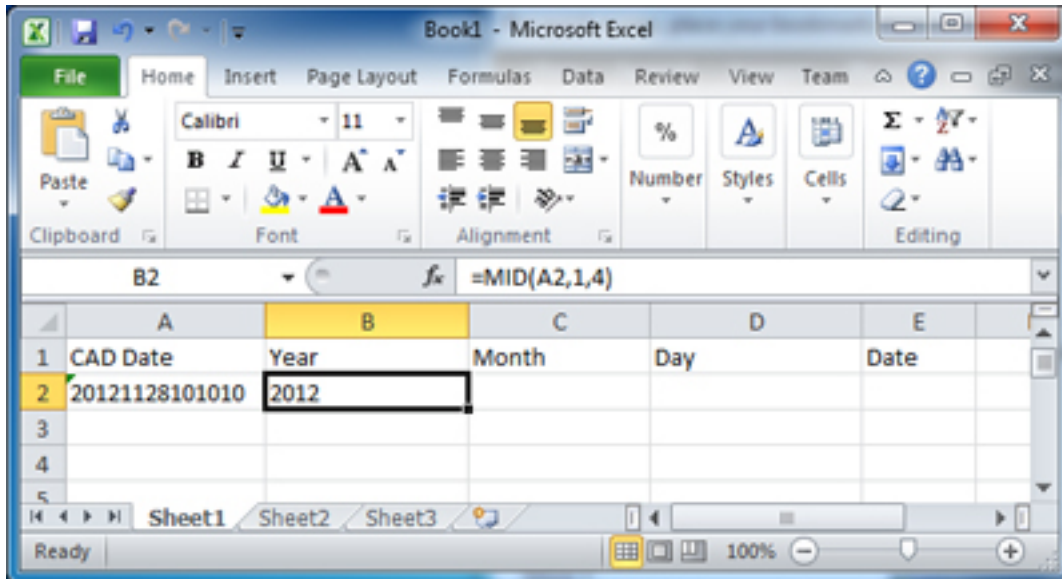

This format represents the date and time as a string in the order of year, month, day, hour, minute and second. Unfortunately, the format is mostly useless to us for manipulating dates so we're going to break out MID to help us. In the above screen shot we're using MID to extract the year into cell B2 (start at position 1, extract 4 characters).

Now try using MID to extract the month in cell C2 and the day into cell D2.

Now that we have the year, month and day broken out we need to combine them into an Excel date and this is where the DATE function comes into play. In the following screen shot I've used DATE in cell E2 and passed in the year, month and day, in that order, as the inputs.
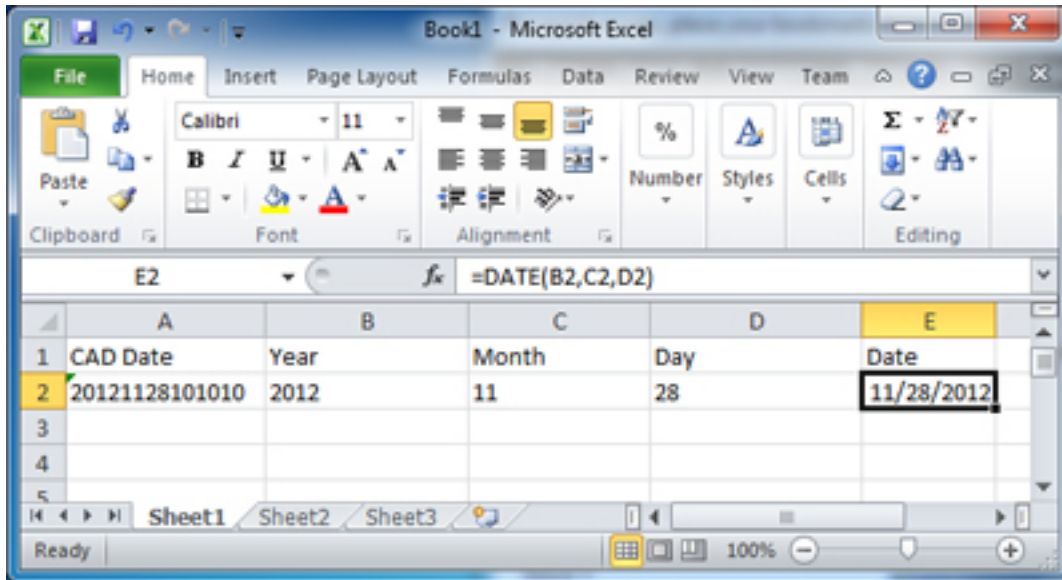

You can verify that the value in cell E2 is actually a date by changing its number format. Click cell E2 and then activate the Home ribbon and select from the Number drop down the value 'Long Date'.

You should also change the number format to 'General' and note that the value in cell E2 is now 41241. This is because Excel stores all dates as numbers behind the scenes and the number represents the number of days since January 1st, 1900. It is possible to have fractions of day (which correspond to time) and we'll cover that in a bit.

The TIME Function

The TIME function is the companion to DATE. Just as DATE takes a year, month and day and returns a proper Excel date, TIME takes an hour, minute and second and returns a proper EXCEL time. Revisting example above it is starightfoward to use MID to extract the hours, minutes and seconds. In the screen shot below is the function call for hours, try and do minutes and seconds yourself.
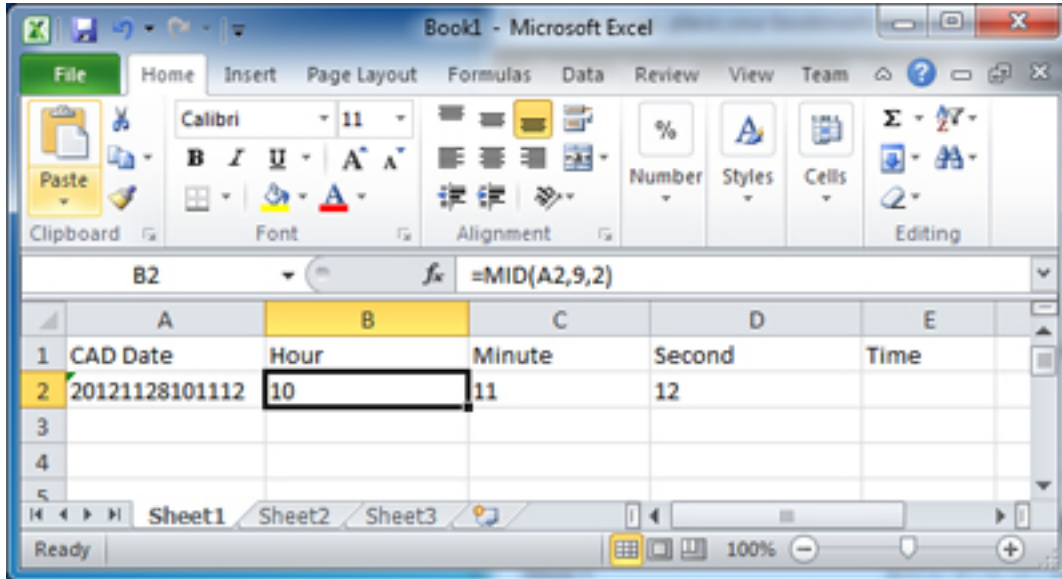
Just like with DATE you can combine them together using TIME as in the screen shot below.
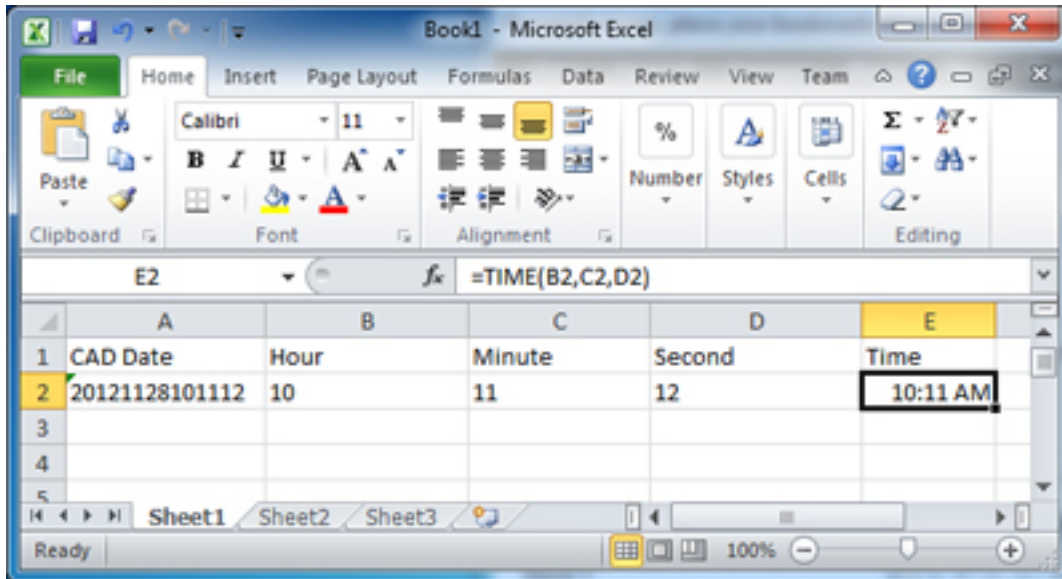

Note that, just like a date, Excel stores times as a number behind the scenes. In the case of hours and minutes though it stores a fraction. Since Excel says a day is equal to 1 than half a day, or 12 hours, is equal to 0.5. The beauty of this system is that if you want to determine the full datetime for something all you need to is add together the date and time cells like this =E2 + I2.

Add together the date from above with the time to convince yourself this works.

The TEXT Function

It might seem strange to list a function named TEXT in the section covering dates but TEXT is your best friend when it comes to formatting a date for a particular purpose. Have a look at the screen shot below that shows six different examples of using TEXT with different formats.
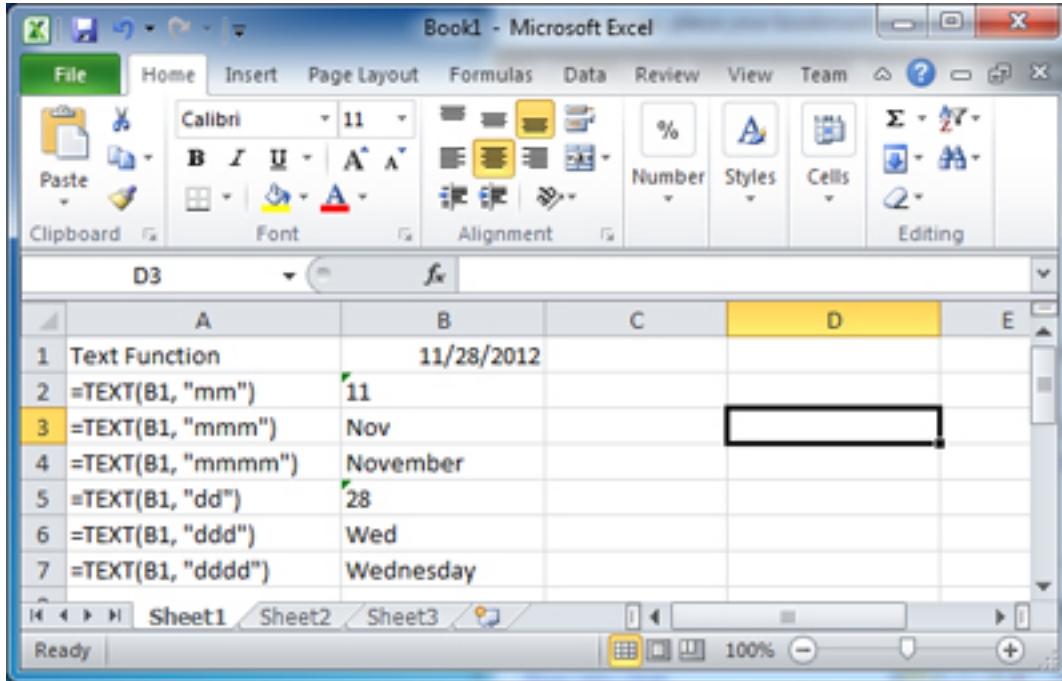
The key difference between each of the TEXT calls is the format that is passed as the second input. Depending on what you pass in you can get out different representations of the month or day (including day of week, good for temporal analysis). TEXT has a lot of other formats as well but I've just listed the ones I use most day to day. Chances are that if you're looking for a way to represent a date then Excel has a format you can use.

The TODAY and NOW Functions

These two functions couldn't be simpler as neither takes an input. The NOW function, as you might surmise, returns the current date and time while the TODAY function returns today's date with no time attached.
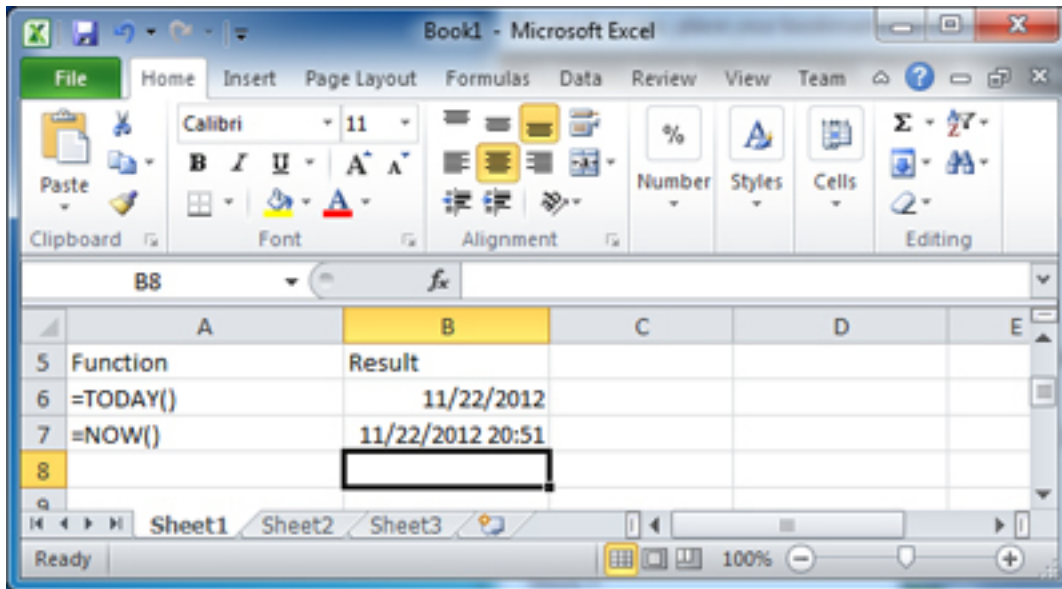

You'll notice that the values in the screen shot are neither today nor now. Try the functions on your own to see the true result.

There is one hitch with both of these functions, when you call them you need to include the empty parenthesis at the end, like this =NOW(). If you just call the function like this =TODAY you'll get an error.

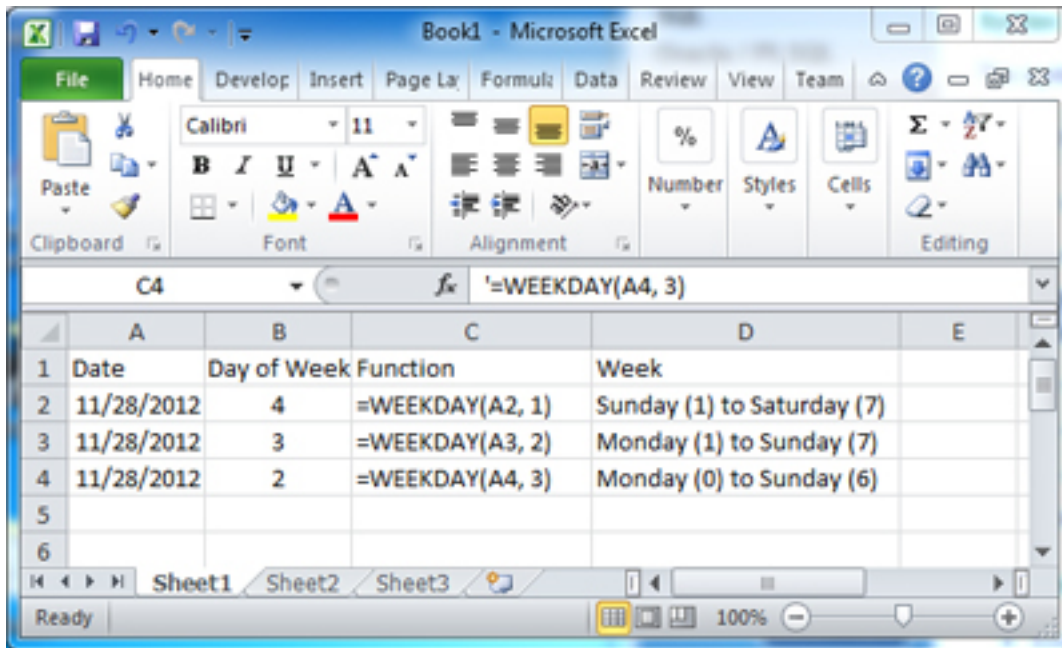## The YEAR, MONTH, DAY, HOUR, MINUTE and SECOND Functions

All of these functions take a date/time as input and spit out the particular piece that you're asking for. They differ from TEXT above in that they always return integer numbers and are primarily used doing further processing rather than presentation.
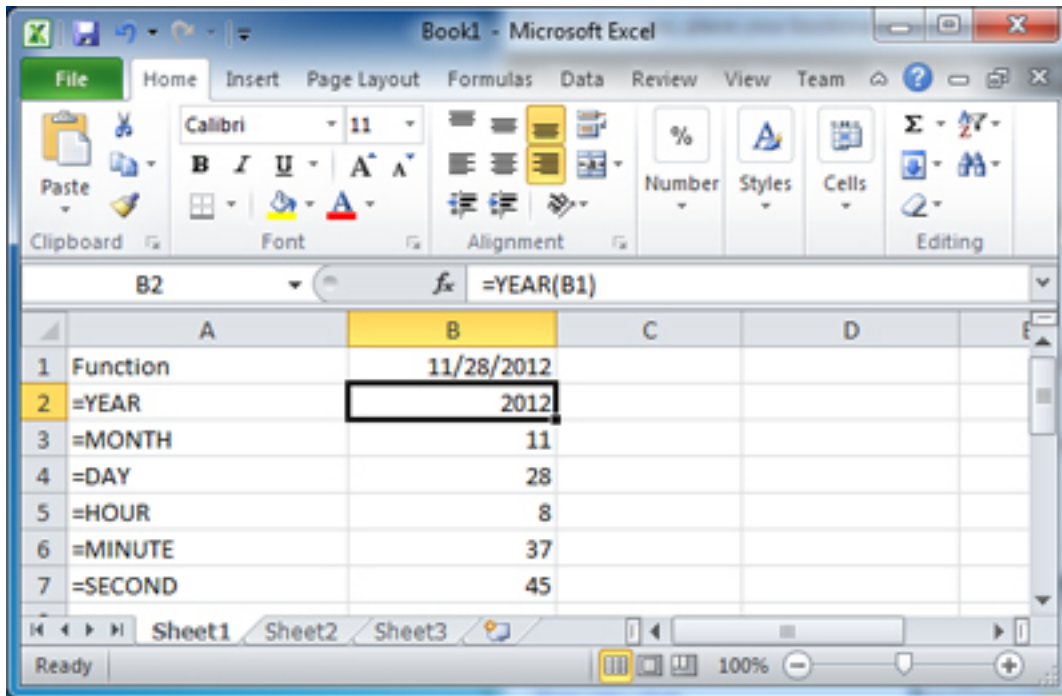
Try using NOW to get the current date and time in a cell and then use YEAR, MONTH, DAY, HOUR, MINUTE and SECOND to pull out the individual pieces.

The WEEKDAY Function

This WEEKDAY function will return the day of the week for a given date. Since different organizations align weeks differently there is also a second optional input for defining when the week starts. In the screen shot below all three examples are demonstrated.

	A	B	C	D	E
1	Date	Day of Week	Function	Week	
2	11/28/2012	4	=WEEKDAY(A2, 1)	Sunday (1) to Saturday (7)	
3	11/28/2012	3	=WEEKDAY(A3, 2)	Monday (1) to Sunday (7)	
4	11/28/2012	2	=WEEKDAY(A4, 3)	Monday (0) to Sunday (6)	
5					
6					

C4 = '=WEEKDAY(A4, 3)

This function is great if you're doing a day of week analysis. Try it with your brithday to figure out what day of the week you were born on.

The DATEDIF Function

The DATEDIF function allows you to find the difference between two dates in terms of year, months or days. It's useful for calculating, in conjunction with the TODAY function, the age of a person.

In the screen shot below the DATEDIF function is used to calculate the current age of a person from their DOB.
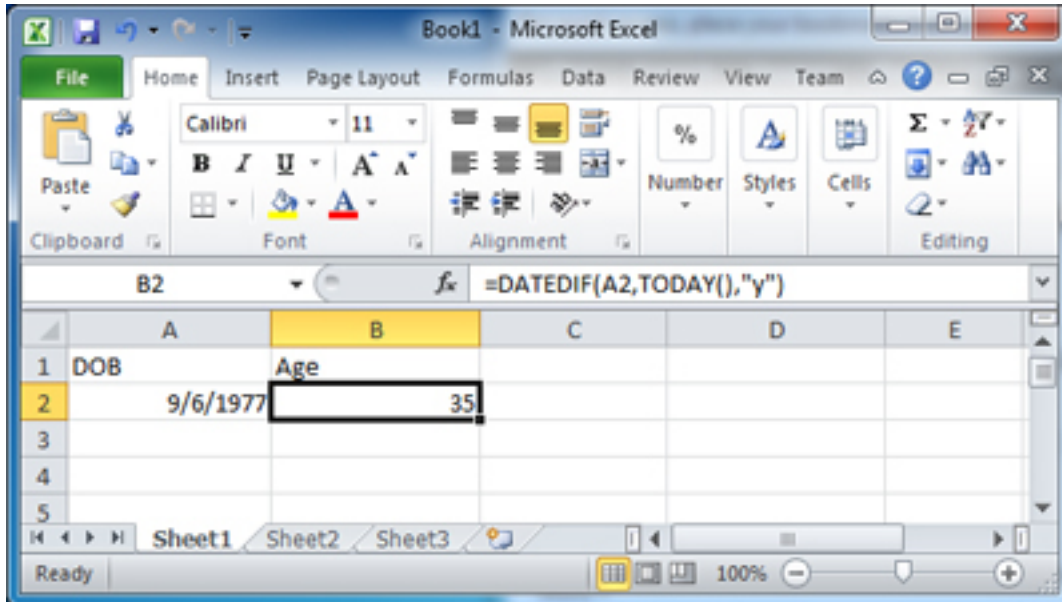
As you can see, in cell B2 the function is =DATEDIF(A2, TODAY(), "y"). The function can also be used to determine someone's age in month or days by swapping the out the "y" for a "m" or a "d". You can, of course, use any other date you'd like in place of the TODAY function call if you are attempting to determine the interval between two dates

Try putting your own date of birth into the cell A2 and calculating how old you are in years, months and days.

Figuring Out the Day of Year

Considering that Excel has so many date functions you'd think that they'd have one for a day of year, but they don't. You can create your own though using what we've already learned in this workshop.

First, let's review what we know:

- We know that Excel stores dates as numbers where a day is equal to 1
- We know that we can add (and subtract) dates as if they were regular numbers
- We know that we can create a date using the DATE function by passing in a year, month, and day

Check out the screen shot below to see how to combine these things we know into a day of year.
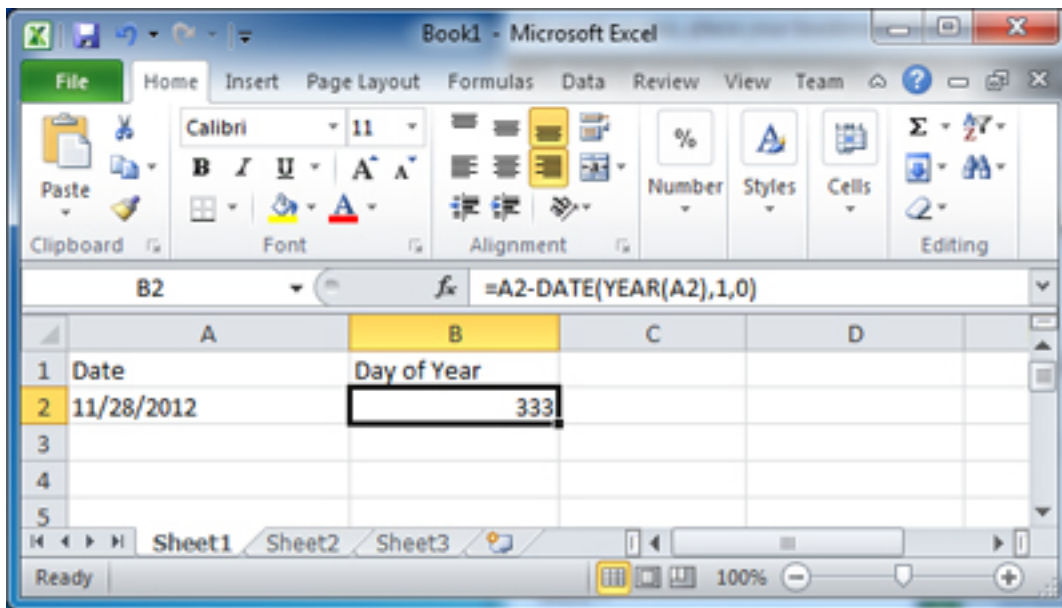
How does it work?

First, we use the YEAR function to determine the year of the date in call A2.

Second, we use the year as input to the DATE function along with month 1 (January) and day 0 (which actually makes it December 31).

Third, we subtract the output of the DATE function from the date in cell A2.

In essence we asking Excel how many days have elapsed since the start of the year, which is the same thing as asking what day of the year it happens to be.

I think we can agree that this is pretty complicated for determining such a simple thing and in the VBA workshop later I will demonstrate how to turn this into a user defined function that hides all the work.

Exercise #2

For exercise #2 we're going to continue working on the spreadsheet we started in Exercise #1. Please have a look at the "Exercise 2" sheet and add the additional columns found there to your own. As you can probably guess, the new columns have to do with calculating dates and ages.

Hints:

- While the DOB looks like a proper Excel date, it's not. You'll need to turn into a proper date using MID.
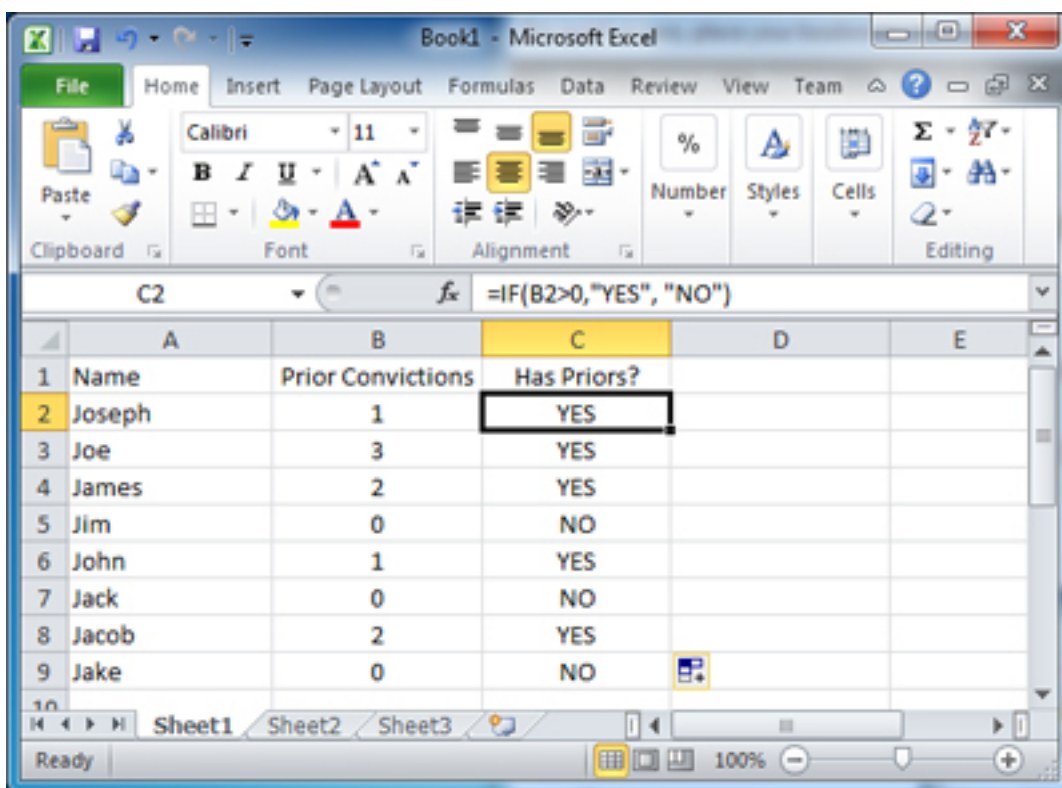
Utility Functions

The final group of functions we're going to look at fall under a general utility category. These functions are used to achieve certain actions in Excel and can be used with a variety of data types. They are a bit more complicated then the ones we've looked at so far but in exchange for complexity you get power.

The IF Function

The IF function allows you to change your output based on the nature of the input. That sounds a little wishy-washy so I'm going to provide an example.

In the screen shot below is a list of individuals with the corresponding number of prior convictions. Let's say that we want to create a third column that identifies whether or not a person has any prior convictions. To do this I have typed in cell C2 the function =IF(B2>0, "YES", "NO").

The result is a YES in column C for everyone except Jim, Jack and Jake. What's going on is that Excel is first evaluating the expression B2>0 to see if it is true or false. If it happens to be true then it outputs the first value ("YES" in this case) if the is false then it outputs the second value ("NO" in this case).

This is how all IF functions work. Some criteria is tested to see if it is true or false and then the corresponding output is provided.

IF is powerful because you can use other functions inside the IF function to produce complex output. For example, below is a list of names and DOBs, except that now one of them is missing a name. In cell C2 the function is =IF(LEN(A2)>0, "", "MISSING NAME").
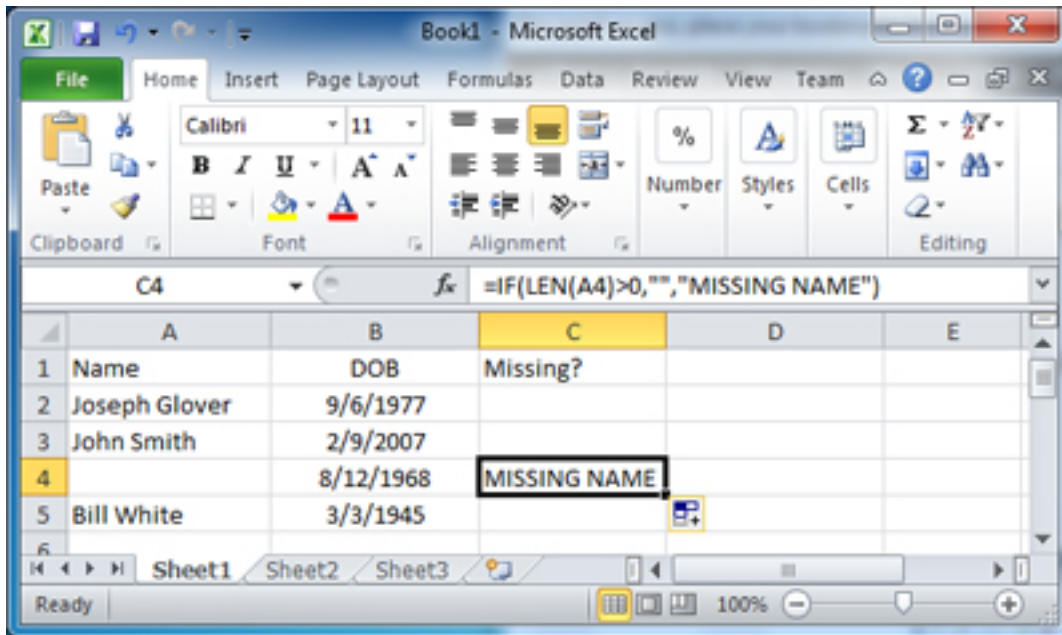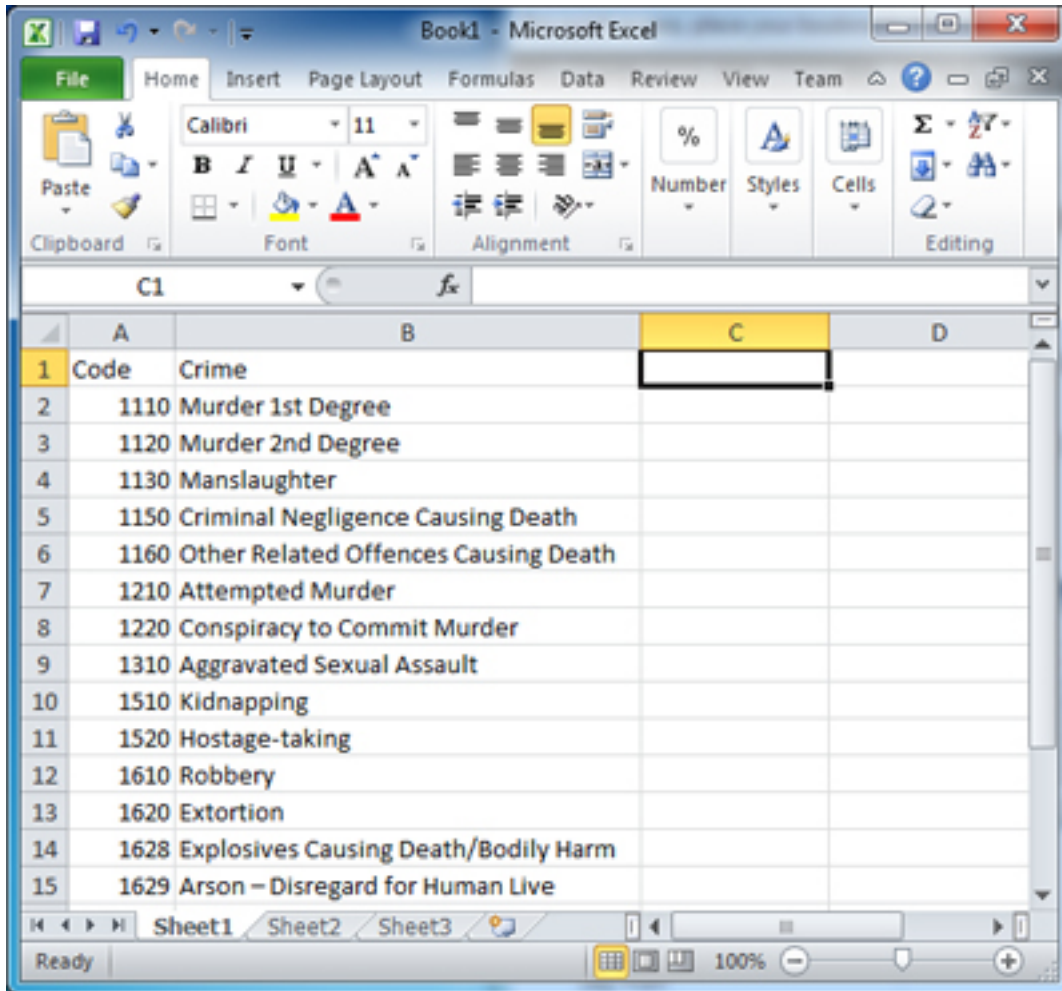

As you can see, filling that function down leaves all of the cells empty except one, the one with the missing name. The IF function is evaluating the length of each name to see if it is longer than 0 and if it is it returns an empty string ("") but if not it makes sure that we can see that the name is missing.

Try changing the IF function to alert you if any of the people in the list is under the age of 17. This will involve using DATEDIF inside the IF function. It is tricky and you can create an age column if you want to do it in two steps.

Depending on how deep you go into Excel the IF statement may become a daily tool. I know I certainly use it every day for analyzing data.

The VLOOKUP Function

VLOOKUP is one of those functions that causes a lot of people grief. It is kind of complex to implement and most people don't use it enough to commit to memory exactly how it works. At its core VLOOKUP (the V stands for vertical by the way) is used for looking up a value based on a reference. As an example of a look up table, have a look at the screen shot below where a list of UCR codes (the look up value) is connected to a crime (the reference).


But why do we need look up tables? Because it's likely that in your day to day data management work you don't have "Murder 1st Degree" in all your spreadsheets but you may have the code 1110. It's shorter and easier to manage. But while the code is easier to work with there are times when you want to display that 1110 is "Murder 1st Degree" and you can use VLOOKUP to look it up ... vertically.

Continued on the next page ...

The key thing to remember about VLOOKUP is that you always have two different sets of data: the first is the data you are working on and the second is the lookup table. The screen shot below shows a sheet with both the data that is being worked on (on the right) and the lookup table on the left (in gray). Notice that the data being worked on has an occurrence number and a UCR code but no Crime.

In the screen shot below I have set up VLOOKUP to fill in the UCR Names for cell C2 from the lookup table by typing the function =VLOOKUP(B2, $F$1:$G$16,2,FALSE).
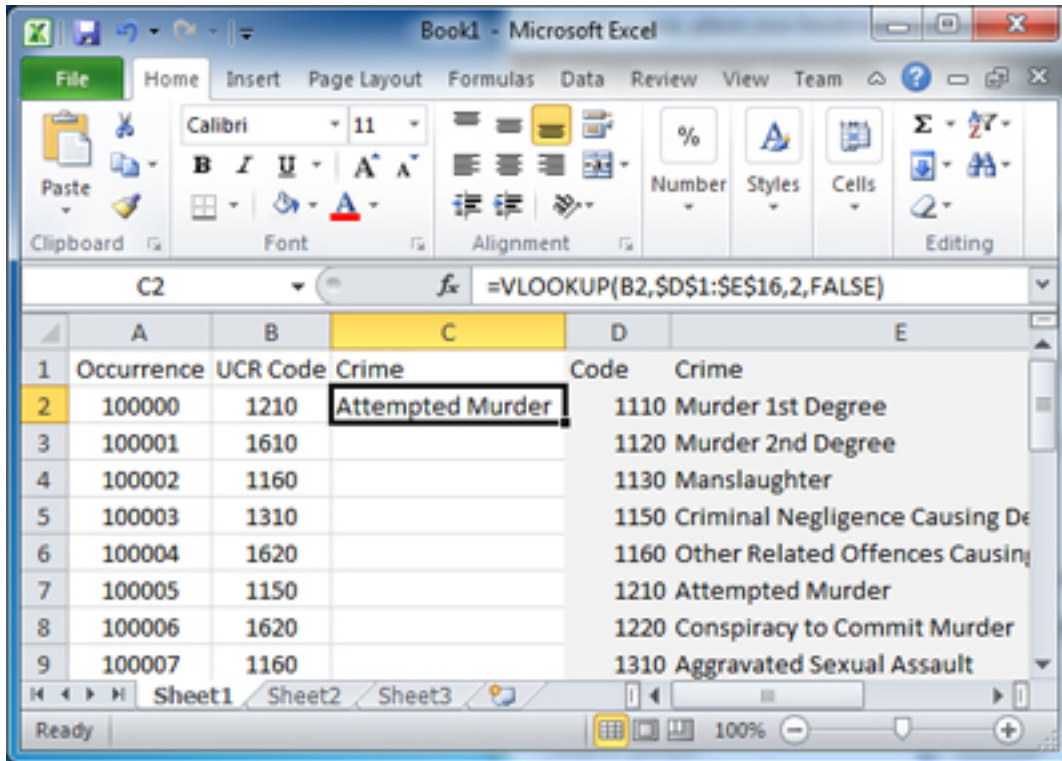
So what is the function doing? First, we're instructing VLOOKUP to use the value in B2 to find the value for C2 and we're telling it to look in the range $D$1:$E$16 which is the entire range occupied by the lookup table. You'll notice that the range values all have dollar signs in front of each letter and number. This is because when we fill this function down we don't Excel to act 'intelligently' and start shifting the references. Using the dollar sign locks the reference so that no matter how we fill it, it will always point to the same set of cells.

A second thing to note is the 2 that is serving as the third argument. Our reference table has two columns, one that is the code and one that is the crime. Since we already know the code what we want to output is the crime and since the crime is in the second column we specify that we want the result to come from column 2.

The final 'FALSE' is there because of the kind of lookup we are doing. There is almost no reason for this value to be TRUE so just always include it as FALSE (it can't be left blank or Excel assumes TRUE) and you'll be fine.

Exercise #3

This one is going to be tough. We're going to finish off the spreadsheet we've been working on for two exercises by incorporating some additional data using IF statements and a VLOOKUP. The final result can be found on the worksheet named "Exercise 3"

Hints:

- Use IF statements to build the information for the single cell that contains the name and cautions. Use LEN to determine what should and should not be concatenated.
- The VLOOKUP part of the exercise is to convert the UCR code into the crime name. You can find the VLOOKUP table in the sheet named "Lookup". Don't be alarmed but all of the people on this list are wanted for some serious stuff.

This concludes the workshop on 30 Excel Functions Every Analyst Should Know. Thank you.