

Data Cheat Sheet: Vlookup and If/Then

Disclaimer: Both of these are difficult to explain in text. If you need this, I recommend that you look up “vlookup excel” or “if statement excel” on YouTube. There are many helpful YouTube videos that go into formulas and seeing the screen as someone goes through it may help illuminate the concept.

VLOOKUP: Formula that allows you to join two datasets together. For example, in spreadsheet 1, you have a group of businesses with the state they are located in. In spreadsheet 2, you have a list of all 50 states and the percentage of their citizens that have advanced degrees. You want to combine these so that in each row for your business, you have the percentage of citizens that have advanced degrees in their home state.

You'll need:

- Two different sheets. It helps if you move the data into one “workbook,” ie there are tabs at the bottom of your spreadsheet where you can toggle between the data. They are not in different browser tabs.
- A key: This is the value that is the same in both spreadsheets. It must be identical.
 - In spreadsheet 2, move the key to the first column. The key must be in the first column of spreadsheet 2. It doesn't matter which column it is in spreadsheet 1.
- Create a blank column in spreadsheet 1. This is where your new data will go. In the example above, this is where the percentage of citizens with advanced degrees for your homestate would go.

The formula is `=VLOOKUP(A2,'Email data'!$1:$993,2,FALSE)`. The formula should go in the first blank cell of the blank column you just created. So, it should be in row 2 (under you title, which you obviously gave it because everything is better with a title. Right?)

In the formula above...

- A2 is the cell in spreadsheet 1 that is your key. So, in the example above, it would be the state.

- The bit of craziness after the comma is the range of values you want to “look” for the data in, ie sheet 2. Do not type this in yourself. You should click over to sheet 2 with the formula dialog box up, then highlight your spreadsheet. It will automatically fill in your range. Then, go back to sheet 1 and type comma.
- 2 is the column number of the data point you want to input into the first spreadsheet. So, our percentage is currently column B in spreadsheet 2. Remember that column A in spreadsheet 2 is our key.
- FALSE. Always false. It’s not really important to know why. It has to do with the efficiency of the search.

Important background knowledge: It may help you remember this if you think of the alligator always “eating” the bigger value.

>	Greater than
<	Less than

IF: Allows you to further categorize rows based on a value. For example, categorizing states into regions, dates into seasons or businesses into industries. This is often used to categorize numeric values. For example, categorizing a grade out of 100 into A-F, categorizing a value into low, medium or high or categorizing numbers above a certain threshold as too high or above average, etc.

The most simple formula would be: =IF(condition-to-test, what-to-do-if-true, what-to-do-if-false)

Seeing this in a real example. Let’s pretend you put the following formula in cell B2: =IF(A2>0, “positive”, “negative”). This example would look at cell A2. If A2 is any number greater than zero, B2 will display as “positive.” If A2 is not greater than zero, B2 will display “negative.”

You can create several categories with one if statement by stacking them. Where our previous example had “negative,” would become another IF statement. Like so: =IF(A2>0, “positive”, IF(A2=0, “zero”, “negative”). So, positive would work as above. If positive wasn’t triggered though, the computer would move to the next if statement (A2=0). If that one wasn’t true, meaning both statements weren’t true, “negative” would display in B2.

A word on unexpected issues: When you use an IF formula like above, the last value (“negative” in both cases) isn’t actually tested. In other words, there is no IF statement that says if B2<0. This can lead to issues because you may not be thinking about all possible values when you are writing your IF statement. To get around this, it’s good practice to write an IF statement for each of your “tests” and to make the last thing after the comma “ERROR.” This will input “ERROR” if something unexpected happens and will save you a lot of heartache.

Example from above: =IF(A2>0, “positive”, IF(A2=0, “zero”, IF(A2<0, “negative”, “ERROR”)))