Using IF and VLOOKUP functions in Google Sheets
MaryJo Webster
maryjo.webster@startribune.com

Video tutorials are available: tinyurl.com/mj-data-academy

Data to go with this class: tinyurl.com/nicar25-mj-sheets

# IF statements

IF() is one of several logical functions in spreadsheet programs. It works the same in Sheets and Microsoft Excel. The basic concept is that you can use it to populate a new column, and have that be different depending on whether it meets the criteria you've defined. I find it useful for some basic data cleaning or for categorizing my data.

**Sheet: "BasicIF"**
This is salary data for the St. Paul Police Department. The city has announced that everyone is going to get a 1% raise, with a minimum of $350. That means that if 1% of a person's pay is less than $350, they will get the $350. Everyone else will get 1% of their pay. Goal here is to calculate what those raises will be for each person, so we can ultimately add up the dollars.

Here's how an IF statement is set up:
=IF(logical_expression, value_if_true, value_if_false)

The logical_expression is where you tell it how to figure out which value to ultimately put in the cell. The value_if_true is what to put in the new column if the logical_expression is true and then the value_if_false is what to put if it's false.

This example is a simple one with only two options. In the next exercise you'll see that it's possible to have 3 or more things to evaluate (those are known as nested IF statements)

Here's the formula we'll use. The blue part is the logical_expression, the red is the value if true and the black is value if false.

=if(F6*.01<=350, 350, F6*.01)

**Sheet: "Nested IF"**

This is data on the results of NFL games from the 2018 season. The data has the scores of each game, but doesn't have fields identifying which team won the game so we can tally up the totals wins and losses for each team. I also want to add a column that indicates if the home team won or if the visit team won. We also need to deal with tie games because there were several during this season. This is an example of categorizing rows of data – a common thing that is needed for analysis.

We'll start by populating the "HomeVisit" column with either "home", "visit" or "tie" depending on who won the game (or if there's a tie)

We need to use 2 IF statements. The first one will evaluate whether it was a tie game (we're doing this first because it's a simple either/or answer.) The second one determines if the "home" team won or the "visiting" team won.

The blue part is the first IF statement…. Notice there really isn't a value_if_false for that first one. It's technically the whole second IF statement.

=if(f2=g2, "tie", if(f2>g2, "home", "visit"))

A common thing that trips you up with nested IF statements is not getting enough parentheses in there. Always count how many opening ones and make sure you have the same number of closing ones. If you get an error message, first thing to check is number of parentheses and whether they are all in the correct place.

Next, let's create a column that has the name of the winning team. Notice the names are stored in columns D (home team) and E (visiting team).

=if(f2=g2, "tie", if(f2>g2, d2, e2))

Next, in the column called "Vikings", let's just create a "yes" or "no" depending if the row is for the Minnesota Vikings. However we need to use a Nested IF statement because on some rows the Vikings are the home team and others they are the visiting team. So we need to look in both the D column and the E column.

Notice we start looking in the D column and if it find "Minnesota Vikings", then it drops a "yes" in the new column. If not, it moves on to the second IF statement and looks in column E. Note: this relies on the values being consistent and the name of the team appearing exactly like that in both columns. If it's not consistent, you should either clean it first or there are ways to put a wildcard in this search.

=if(d2="Minnesota Vikings", "yes", if(e2="Minnesota Vikings", "yes", "no"))

# VLOOKUP

VLOOKUP is a function that allows you to "join" data from one sheet to another. You can transfer one column of information over to another sheet, matching based on common values. Both sets of data – ideally stored in separate sheets within the same workbook – need to have one column that is the same.

Sheets: "census_biz" and "county_lookup"

In the “census_biz” sheet we have partial data from a Census dataset that tallies up the number of workers and businesses in each county, plus total payroll. (this one is kind of old, btw)  Notice that we don’t have the names of the state or the counties. FIPS code 27 is Minnesota, but I don’t know the FIPS codes of the counties (column B).

In “county_lookup” we have a code book that translates those county FIPS into names. Our goal is to transfer that name over to the “census_biz” sheet.

First, let me explain how VLOOKUP works.

```
=VLOOKUP(search_key, range, index, [is_sorted])
```

Search_key is the cell in your data that you want to find a match for. In “census_biz” that’s going to be B2 – the first value in the county FIPS column.

Range – this is the range of cells that make up your lookup table. This will be everything from A1:B88 in the “county_lookup” table. IMPORTANT: the left-most column in this range needs to be the one you are matching/joining on.

Index – this is the number of the column in the “range” that you want to move back to the dataset and drop it in the column where you are typing this formula. Notice I said “number of the column” – I didn’t say the letter. Our range – the data in “county_lookup”-- consists of 2 columns. So we will tell it we want column 2.

[is_sorted] is an optional argument that won’t be needed.

If you Google how to do a VLOOKUP, it won’t include this first step I’m going to show you. But I can guarantee you, my approach saves a lot of headache and simplifies the ultimate formula.

What we’re going to do is assign a name to the “range.” If we don’t assign a name, we would have to write a complicated jumble to ensure Sheets understands where that A1:B88 is located. IMPORTANT: the left-most column of your range needs to be the one you are joining on. Notice that the county FIPS code is the left-most column here.

Go to the “county_lookup” sheet and highlight the range of cells we want (everything from A1:B88).  Then right-mouse click and choose “View more cell actions” (all the way at the bottom) and then choose “Define Named Range.” Over on the right it will give you a change to name this area – let’s call it “countynames” .  Click Done.

Go back to “census_biz” and in the first blank column, we’ll start building our VLOOKUP.

=VLOOKUP(B2, countynames, 2)

Just for giggles, here’s what that formula would have to look like if you DIDN’T use the named range:
=vlookup(B2, county_lookup!$A$1:$B$88, 2)

Let's do another example. This is one I use for some simple data cleanup.

**Sheets: "death_data" and "race_lookup"**

The data in "death_data" sheet is a cut from Minnesota death certificates of people where opioids were a factor in their deaths. I included just a few of the fields.

Do a quick pivot table on the "race" column and notice that we've got quite a few inconsistencies. When I was prepping this data, I took this column out of the pivot table and pasted it into the sheet that is called "race_lookup." Notice that it's the column called "original." And then I added a new column where I typed the value I wanted to replace each one with.

We'll use a VLOOKUP to match those original values back in the dataset and create a new column where we'll drop the new values. This is a very easy data cleanup option when you have inconsistencies but it isn't hundreds of them.

Like we did in the last example, first let's do the "Define Named Range" step. Highlight the table in the "race_lookup' sheet; right-mouse click, go to More cell actions and then "Define Named Range." let's name this "clean_race"

Then go back to the "death_data" worksheet and in the first blank column, let's drop our VLOOKUP.
Our matching value is the original race column in H2.
=VLOOKUP(h2, clean_race, 2)

Optional: You can repeat that last step for the Hispanic ethnicity column. You would do a Pivot Table and copy the column out to a new sheet and make your own lookup table. Then all the other steps would be the same.

**Sheets: "crime_data" and "shift_lookup"**
This is partial crime incident data in the city of Minneapolis. It has the date and TIME that the call came in, but I want to categorize each row by the shift for the police officers (morning, afternoon or night). This is known as an "inexact" matching situation. The ones we did previously we all exact matches.

Notice the "shift_lookup" table has the start time of each shift and then a column with the name of the shift that we want to transfer over to the "crime_data" sheet.

Setting up a lookup table for an inexact match is a little bit harder than an exact match. Also it has limited applications – mainly dealing with anything numeric: numbers, dates and times.

When dealing with numbers, the first row of the lookup table needs to have "0" as the first start value.

In this case, because we're dealing with time, the first value needs to be midnight. Notice that we have two values for the "night" shift. I've purposely made the first value in all capital letters ("NIGHT") so that you can how the formula is working when it's applied to the data.

The first row with "NIGHT" covers midnight to 5:59 a.m.  The second "night" value at the end of the table covers the first half of the night shift from 10 pm to 11:59 p.m.

First, let's go to the "shift_lookup" sheet and define the range. Highlight the little table and right-mouse click to get to "Define Named Range" and let's call this "shifts"

Next, go back to "crime_data" and in the first blank column let's put our VLOOKUP
=VLOOKUP(I2, shifts, 2)

Notice that anything that falls between midnight and 5:59 a.m. has "NIGHT" (in capital letters) and the ones between 10 pm and 11:59 p.m. have "Night" in proper case. If we had created our lookup table starting at 6 am, those midnight to 5:59 records wouldn't have found a "match"

If you want your data to be standardized, you can go into the lookup table and change NIGHT to proper case to match the other one and your formula will automatically adjust.