## MODULE 4 - PERFORM DATA CALCULATIONS

Calculations are a common task for data analysts. In this part of the course, you’ll explore formulas, functions, and pivot tables in spreadsheets and queries in SQL, all of which will help with your calculations. You’ll also learn about the benefits of using SQL to manage temporary tables.

### **Learning Objectives**

* Describe the use of functions to conduct basic calculations on data in spreadsheets
* Discuss the use of pivot tables to conduct calculations on data in spreadsheets
* Demonstrate how to use SQL queries to complete calculations in SQL
* Explain the importance of the data-validation process for ensuring accuracy and consistency in your analysis
* Discuss the use of SQL queries to manage temporary tables
* Reflect on how conditional statements can be used to create complex queries and functions
* Generate multiple points of summary based on a wide variety of conditions using COUNTIF, SUMIF, MAXIF, and AVERAGEIF

## GET STARTED WITH DATA CALCULATIONS

### [DATA CALCULATIONS](https://www.coursera.org/learn/analyze-data/lecture/HZjQ8/data-calculations)

As a data analyst, you'll use key tools and processes over and over, but you'll also learn new things as you grow in your job. It could be anything from building a new kind of analysis to a time-saving shortcut.

When I first got to Google, I relied on just a couple of programs and tools to access data and do my analysis. But I soon realized that I wasn't working as efficiently as I wanted to. Once I got comfortable pulling data and analyzing it using SQL, it allowed me to be a lot more efficient than before. And the better I got at SQL and pulling the data from data tables, the faster I completed my analysis. I was hooked.

Over the next few videos, **I will show you some ways to be as efficient as possible while completing calculations during your analysis**.

We'll start by revisiting spreadsheets where we'll look at **formulas for basic calculations**.

Then we'll move into **conditional formulas that use the IF function** to check whether a condition is met through a calculation.

After that, we'll explore the multifunctional **SUMPRODUCT function**. Try saying that five times quickly! SUMPRODUCT adds and multiplies all in one step, so it's very useful.

Next we'll take another look at **pivot tables**. If you've skipped around, and it's your first time learning about them, you'll get to know all about them. Pivot tables have tons of uses, including organizing your calculations.

We'll then pivot to SQL, pun intended. We'll show how **queries and calculations go hand in hand in SQL.**

We'll also look at **temporary tables in SQL**, which are helpful for temporarily storing your data during analysis.

We'll be covering lots of new concepts in these videos, so feel free to hit the pause button at any time to think through the problem or steps to try it on your own. And you can always review the videos as much as you need to.

So to recap, we'll have a little bit of a review, and then cover some all new concepts, all about calculations. Are you ready? Good. Me, too.

### [COMMON CALCULATION FORMULAS](https://www.coursera.org/learn/analyze-data/lecture/33kSX/common-calculation-formulas)

You probably do a lot of calculations in your daily life. Maybe it's figuring out how much to tip someone or balancing your budget. You might do some of these calculations in your head or with paper and pencil or the calculator on your phone. You might even have shortcuts to use to make the calculations easier. You'll perform a lot of calculations as a data analyst too. But they'll involve more numbers in a wider range of calculations. That's where you'll put your data analyst tools to work. We'll show you how you could use formulas in a spreadsheet to complete some of the most basic calculations. Formulas are one of the many shortcuts that data analysts use. But rest assured, even though they're shortcuts, they'll still calculate with complete accuracy. We've covered a lot of these calculations earlier in the program. But if you skip that part and want a refresher, we'll review them here. These calculations will also be more advanced than the ones we've covered so far. But they'll also be closer to what you might use on the job. We'll be using Google Sheets in this video, but you can also use Excel. The steps might look a little different in Excel, but the outcomes will be the same.

Let's try out some calculations with sales data from a discount store chain. We'll look at data for one of the stores in the chain. Our objective: use the existing sales data to find any trends. This is a great way to see a lot of the ways formulas can be useful in your analysis.

We'll start by finding annual sales over the years 2011-2020. The data is already organized in columns by month and in rows by year. **But we don't have the total sales for each year yet.** We can use a sum function to help us figure that out. We'll add the sales from 2011 first. We'll add a heading for the annual sales column, then we can type our sum function and a formula. All formulas begin with an equal sign. We'll type that first, followed by sum and then an open parenthesis. After the open parenthesis, we need to tell the formula which cells are being added. In this case, we need data from the whole row which begins in cell B2. B2 is a cell reference we'll use. Instead of typing each cell one by one, we can put them in the formula quickly by selecting cell B2 and dragging the fill handle across the row to the last cell with sales data, M2.Now we'll complete the formula by closing the parentheses and pressing Enter.

Just like that, we've calculated the total sales for 2011. Here's another shortcut we worked with in an earlier video. The fill handle is the tiny box in the corner of each sale. You can use it for lots of things like selecting multiple cells for a formula or continuing a pattern across several cells, the fill handle definitely qualifies as a shortcut. We can use the formula we created to calculate the total sales for the other years in the dataset. All we have to do is drag the fill handle down the other cells in the annual sales column and we'll have total sales data for the rest of the years in the dataset.

**Let's say, we also need to find the growth in annual sales from year to year.**

**This would be a good time to think through the problem before we try to solve it**. Do we have the data we need to solve this? Not yet. Thinking backwards like this helps us plan out the steps to move forward.

The first step we'll need to do is calculate the total sales per year. Then we'll measure the rate of change between years. We'll start by labeling a new column.

In this case, we won't need to use a function or parentheses, since we're only using data from two cells. We can just use the name of those cells, we'll type an equals sign and then click in "Cell N3", which automatically populates that sale in the formula. Next, we'll add a minus sign to the formula because we're subtracting to find the difference between two consecutive years. Clicking in "Cell N2" gives us the total from 2011, which we can then subtract from the total from 2012. Then we hit Enter and get our sales growth from 2011-2012. We're definitely getting some useful data here. Let's keep going. We can also use our sales growth to find the growth rate between the two years. We'll show this as a percentage. We'll head our column with the percent sign and growth. To do this, we'll divide the total in cell O3 by the annual sales from 2011 in cell N2. A slash is a symbol that a formula recognizes as division, so we'll place that between the two cell references and presto, there's the growth rate. Growth rates are usually shown as percentages, which can be easier than a decimal to read and understand. Let's change this number to a percentage. Time for another shortcut. All we have to do is click the percent style button and our growth rate will become a percentage. We can select the cells for both the total growth and the growth rate to populate the rest of the two columns.

We have some negative numbers, but that just means that there was negative growth from one year to the next. We've got just a few more things to calculate for our stakeholders. Next step is finding the average sales. We want to compare sales between months to learn if there's a trend. We'll add this in a row instead of a column. This will line up our averages under each month.

To find our averages, we’ll calculate the total and then divide that total by the number of values added to get it. We can do this by using the average function.

Between our parentheses will select the cells that contain the sales data for January, B2 through B11.

We'll duplicate that formula across the row through December to look for trends.

Right away, we know that summer months and December have the highest average sales.

Since our stakeholders will want to understand our findings quickly and easily, we'll add a little visualization to the data with conditional formatting. You'll learn more about data visualizations like conditional formatting soon. But here's a sneak peek. Conditional Formatting is a spreadsheet tool that changes how cells appear when values meet specific conditions. Let's apply conditional formatting to the cells with the average sales by month.

We'll use a color scale to show the range of averages. Well, the lowest monthly average remaining as white and we'll apply shades of green to the rest of the values.

The brighter the green, the higher the average. Now, when we share our analysis with our stakeholders, they will be able to tell right away which months have the highest average sales. Just a couple more steps to complete our analysis. Now we need to find the minimum and maximum for average monthly sales. With the dataset this small, it might be easy to find the minimum and maximum values without a formula, but it's still good practice to use one. Not to mention, using a formula helps prevent human error, will again rely on formulas with Functions to do these calculations, we'll start with the lowest monthly average.

Our function here is MIN, followed by the cells with the average month B12 through M12.

After we press Enter, the lowest monthly average is calculated. We can repeat the same steps to find the highest monthly average,

in this formula will use the same data, but we'll replace MIN with MAX for maximum.

For this store location, sales are strongest in December and weakest in January. We could share these findings with stakeholders if they've met our objectives. If they haven't, we might need to continue with our analysis. Either way, I hope you've learned how spreadsheet formulas can be valuable tools when doing calculations. Coming up, we'll check out more formulas. See you soon.

### [FUNCTIONS AND CONDITIONS](https://www.coursera.org/learn/analyze-data/lecture/3OFTV/functions-and-conditions)

One of the first calculations most kids learn how to do is counting. Soon after, they learn addition, and that doesn't go away. No matter what age we are, we're always counting or adding something, whether it's change at the grocery store or measurements in a recipe. Data analysts do a lot of counting and adding too. And with the amount of data you'll come across as a data analyst, you'll be grateful to have functions that can do the counting and adding for you.

So let's learn how these functions COUNTIF and SUMIF can help you do calculations for your analysis more easily and accurately. We'll start with the COUNTIF function.

You might remember COUNTIF from some of the earlier videos about data cleaning. **COUNTIF returns the number of cells that match a specified value. Earlier, we showed how COUNTIF can be used to find and count errors in a data set.**

Here we'll only be counting. Just a reminder though, while we won't be actively searching for errors in this video, you'll still want to watch out for any data that doesn't look right when doing your own analysis.

**As a data analyst, you'll look for and fix errors every step of the way.**

For this example, we'll look at a sample of data from an online kitchen supplies retailer.

Our stakeholders have asked us to answer a few questions about the data to understand more about customer transactions, including the revenue they're bringing in. We've added the questions we need to answer to the spreadsheet.

We'll set up a simple summary table, which is a table used to summarize statistical information about data. We'll use the questions to create the attributes for our table columns: count, revenue total, and average revenue per transaction.

Each of our questions ask about transactions with one item or transactions with more than one item, so those will be the observations for our rows.

We'll make Quantity the heading for our observations.

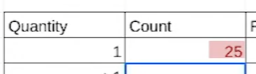
We'll also add borders to make the summary table nice and clear.

The first question asks, How many transactions include exactly one item? To answer this, we'll add a formula using the COUNTIF function in cell G11.

We'll begin with an equal sign, COUNTIF, and an open parenthesis.

Column B has data about quantity. So we'll select cells B3 through B50, followed by a comma.

Next, we need to tell the formula the value that we're looking for in the cells we've selected. We want to tell the data to count the number of transactions if they equal 1. In this case, between quotation marks, we'll type an equal sign and the number 1 because that's the exact value we need to count. When we add a closed parenthesis and press enter, we get the total count for transactions with only one item, which is 25.



We can follow the same steps to count values greater than one.

But this time, because we only want values greater than 1, we'll type a greater than sign in our formula inside of an equals sign.

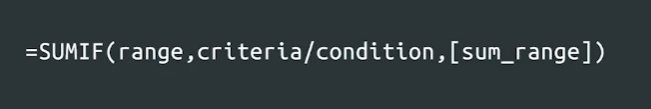


Getting this information helps us compare the data about quantity.

Okay, now we need to find out how much total revenue each transaction type brought in. Since the data isn't organized by quantity, we'll use the SUMIF function to help us add the revenue for transactions with one item and with one more item separately.

**SUMIF is a function that adds numeric data based on one condition.**

Building a formula with SUMIF is a bit different than one with COUNTIF. They both start the same way with an equal sign and the function, but a SUMIF formula contains the range of cells to be evaluated by your criteria, and the criteria. In other words, SUMIF has a list of cells to check based on the criteria you set in the formula. Then the range where we want to add the numbers is placed in the formula if that range is different from the range being evaluated.





There's commas between each of these parts. Adding a space after each comma is optional. So let's try this. In cell H11, we'll type our formula. The range to be evaluated is in column B, so we'll select those cells.

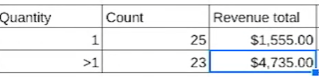
The condition we want the data to meet is for the values in the column to be equal to one. So we'll type a comma and then inside quotes an equal sign and the number one.

Then we'll select the range to be added based on whether the data from our first range is equal to one. This range is in column C, which lists the revenue for each transaction.

So every amount of revenue earned from a transaction with only one item will be added together. And there's our total. Since this is revenue, we'll change the format of the number to currency, so it shows up as dollars and cents.

So the transactions with exactly one item earned $1,555.00 in revenue. Let's see how much the transactions with more than one item earned.

Okay, let's check out the results. Just like with our COUNTIF examples, the second SUMIF formula will be the same as the first, except for the condition, which will make it greater than one.



When we run the formula, we discover that the revenue total is much higher, $4,735.00. This makes sense, since the revenue is coming from transactions with more than one item. Good news.

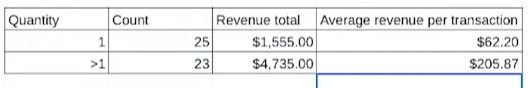
To complete our objective, we'll do two more quick calculations.

First, we'll find the average revenue per transaction by dividing each total by its count. This will show our stakeholders how much of a difference there is in revenue per transaction between one item and multiple item transactions. This information could be useful for lots of reasons. For example, figuring out whether to add a discount on purchases with more than one item to encourage customers to buy more. We'll put these calculations in the last column of our summary table. You might remember that we use a slash in a formula as the operator for division calculations.



The average revenue for transactions with one item is $62.20.

And the average revenue for transactions with more than one item is $205.87.



And that's it for our analysis. Our summary table now gives the stakeholders and team members a snapshot of the analysis that's easy to understand. Our COUNTIF and SUMIF functions played a big role here. Using these functions to complete calculations, especially in large datasets, can help speed up your analysis. They can also make counting and adding a little more interesting. Nothing wrong with that. And coming up, we'll explore more functions to make your calculations run smoothly. Bye for now.

### [FUNCTIONS WITH MULTIPLE CONDITIONS](https://www.coursera.org/learn/analyze-data/supplement/s9khi/functions-with-multiple-conditions)

As you’ve been learning, conditional functions and formulas perform calculations according to specific conditions. In addition, functions including **SUMIF** and **COUNTIF** only work in cases where there is one condition.

However, if you have more than one condition, you would need to use the **SUMIFS** or the **COUNTIFS** function instead. These functions enable you to perform calculations if you have two or more conditions. In this reading, you will learn more about conditional functions and how to construct functions with multiple conditions by exploring their basic syntax and checking out an example.

You will also be able to access resources for similar functions in Excel.

## **SUMIF to SUMIFS**

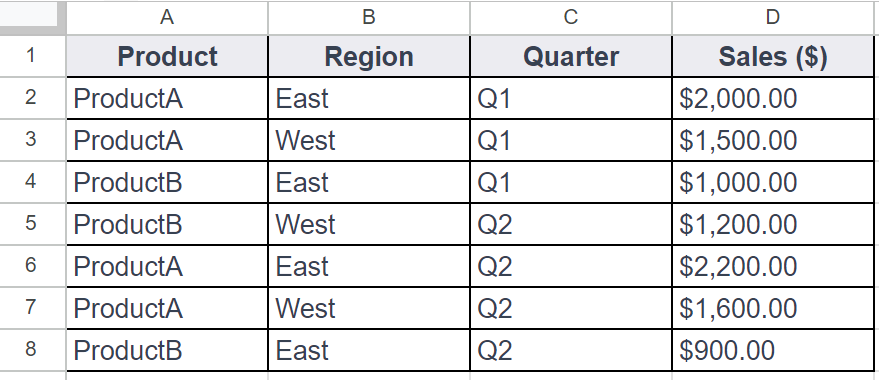
Previously, you learned that the **SUMIF function adds values in a particular range based on a single condition.**

**The basic syntax is =SUMIF(range, criterion, sum\_range).**

The first range is where the function will search for the condition that you have set. The criterion is the condition you are applying and the sum\_range is the range of cells that will be included in the calculation.

For example, in an accounting spreadsheet, you could use **SUMIF** to calculate the total expenses for a specific category, like Travel expenses, within a given month.

Or, you could find the total sales for automotive fuel treatment products– in this table, the ProductA is high octane fuel and ProductB is standard octane. Table 1 includes columns for Product, Region, Quarter, and Sales.



You could use **SUMIF** to calculate the total sales for Product A using a formula like this:

**=SUMIF(A2:A8, "ProductA", D2:D8)**

But, you could also build in multiple conditions by using the **SUMIFS** function. **SUMIF** and **SUMIFS** are very similar: They add up values in a range.

**But SUMIFS can include multiple condition**s. This gives you more control over your summing criteria, which, in turn, allows you to perform more complex data analysis easily.

**The basic syntax is: =SUMIFS(sum\_range, criteria\_range1, criterion1, [criteria\_range2, criterion2, ...])**

The square brackets let you know that this is optional.

The ellipsis at the end of the statement enables as many repetitions of these parameters as needed.

For example, if you wanted to calculate the sum of sales for ProductA in the East district in the first quarter, you could create a **SUMIFS** statement with multiple conditions, like this:

**=SUMIFS(D2:D8, A2:A8, "ProductA", B2:B8, "East", C2:C8, "Q1")**

In this example, B2:B8 is the second criterion\_range and East is the second condition. The third criterion\_range is C2:C8 and the third condition is Q1.

**As long as you follow the basic syntax, you can add up to 127 conditions to a SUMIFS statement!**

## **COUNTIF to COUNTIFS**

Just like the **SUMIFS** function, **COUNTIFS** allows you to create a **COUNTIF** function with multiple conditions.

The definitionfor **COUNTIF is a function that counts the number of cells in a range that meet a single condition**.

For example, using **COUNTIF** to track the number of days a temporary employee was absent in an attendance record.

**The basic syntax is: =COUNTIF(range, criterion)**

Just like **SUMIF**, you set the range and then the condition that needs to be met. For example, in Table 1, if you wanted to count the number of transactions for ProductA, you could use a **COUNTIF** function like this:

**=COUNTIF(A2:A8, "ProductA")**

**COUNTIFS has the same basic syntax as SUMIFS: =COUNTIFS(criteria\_range1, criterion1, [criteria\_range2, criterion2, ...])**

The criteria\_range and criterion are in the same order, and you can add more conditions to the end of the function. So, if you wanted to find the number of sales transactions for ProductA in the East region in the first quarter, you could use **COUNTIFS** to apply those conditions, like this:

**=COUNTIFS(A2:A8, "ProductA", B2:B8, "East", C2:C8, "Q2")**

This enables you to find every instance where both of conditions (East and Q1) are true.

## **For more information**

**SUMIFS** and **COUNTIFS** are just two examples of functions with multiple conditions. They help demonstrate how multiple conditions can be built into the basic syntax of a function. There are other functions with multiple conditions that you can use in your data analysis and many resources available online to help you get started:

* [**How to use the Excel IFS function**](https://exceljet.net/excel-functions/excel-ifs-function): This includes an explanation and example of the **IFS** function in Excel. It’s a great reference if you’re interested in learning more about **IFS**. The example is a useful way to understand this function and how it can be used.
* [**VLOOKUP in Excel with multiple criteria**](https://exceljet.net/formula/vlookup-with-multiple-criteria): Similar to the previous resource, this resource goes into more detail about how to use **VLOOKUP** with multiple criteria. Being able to apply **VLOOKUP** with multiple criteria will be a useful skill, so check out this resource for more guidance on how you can start using it on your own spreadsheet data.
* [**INDEX and MATCH in Excel with multiple criteria**](https://www.coursera.org/learn/analyze-data/supplement/s9khi/functions-with-multiple-conditions): This resource explains how to use the **INDEX** and **MATCH** functions with multiple criteria. It also includes an example, which demonstrates how these functions work with multiple criteria and actual data.
* [**Using IF with AND, OR, and NOT functions in Excel**](https://support.microsoft.com/en-us/office/using-if-with-and-or-and-not-functions-d895f58c-b36c-419e-b1f2-5c193a236d97): This resource combines IF with AND, OR, and NOT functions to create more complex functions. By combining these functions, you can perform your tasks more efficiently and cover more criteria at once.

### [HANDS-ON ACTIVITY: WORK WITH CONDITIONS](https://www.coursera.org/learn/analyze-data/quiz/abUf7/hands-on-activity-work-with-conditions)



## **Activity Overview**

**empty alt text**

In previous activities, you used basic spreadsheet functions such as **COUNT**, **SUM**, **AVERAGE**, and **MAX**. In this activity, you will work with the conditional versions of these functions: **COUNTIF**, **SUMIF**, **AVERAGEIF**, and **MAXIFS**.

Conditional functions are functions that perform a specific task, but only on cells that satisfy some defined criteria. They are usually identified with an **IF** suffix adjoined to the desired operation. They are frequently used when constructing more complex queries that cannot be accomplished using more basic functions.

By the time you complete this activity, you will be able to use conditional functions and understand when and why they are appropriate. This will enable you to do more complex analysis with spreadsheets as you continue to develop your data analyst skill set.

### 

### **Step-By-Step Instructions**

### **Step 1: Access the template**

To get started, you will need to access the Working with Conditions spreadsheet.

Click the link to the spreadsheet to create a copy. If you don’t have a Google account, you may download the spreadsheet directly from the attachments below. Make sure to select “Use Template” on the downloadable item.

Link to spreadsheet: [Working with Conditions](https://docs.google.com/spreadsheets/d/1sKNqBzV63Na76jjoKLukjTtDyDMwVW0Vjz4mM-sJDEM/template/preview#gid=895883562)OR[Working with Conditions](https://d3c33hcgiwev3.cloudfront.net/qLRKJmPJTR6yRiMQRNIcfg_918c1ebc887045bba3c61fcdfdc080e1_Working-with-Conditions.xlsx?Expires=1713052800&Signature=g0KaoAE7k-fyyHkwXdcUI5ayTMQNNJWtFixdK24eCTk-62-q~KZr8Znl6U3mN-CEs0OyWXxG1aHuwH8qGPNidkrbMNQjg2WxrMAJITbPOp6F62qwhUikmipeyqh3sgq-qs6ljfHsgCVK6Mu8BAwPRUVlkvpi3NoMsXud05EFY2Y_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

### 

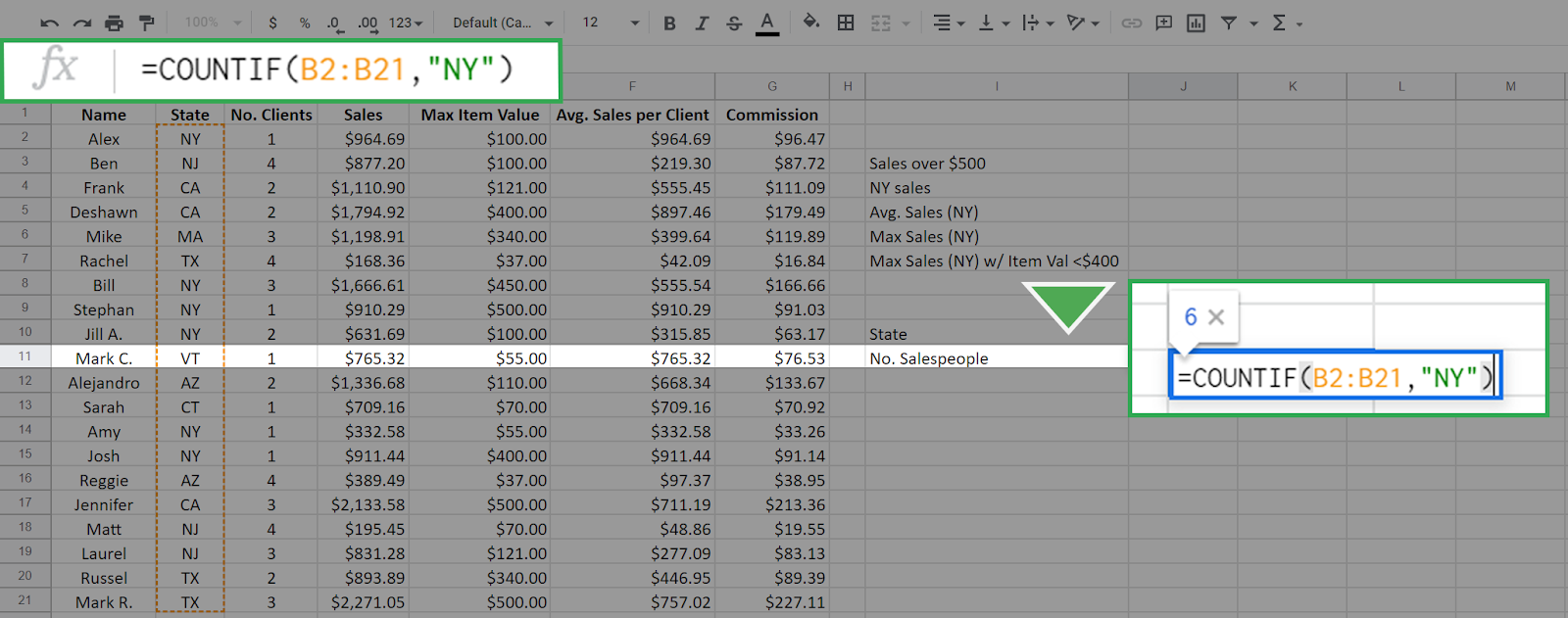
### **Step 2: Use the COUNTIF function**

First, open the Working with Conditions spreadsheet.

In this scenario, your stakeholders have asked you to calculate the number of salespeople that the company has in New York state. The **COUNTIF** function allows you to do this easily. **COUNTIF** counts and returns the number of records in a table that meet a certain criteria—in this case, the number of salespeople a company has in the state of New York. The syntax is **=COUNTIF(**range, criteria).

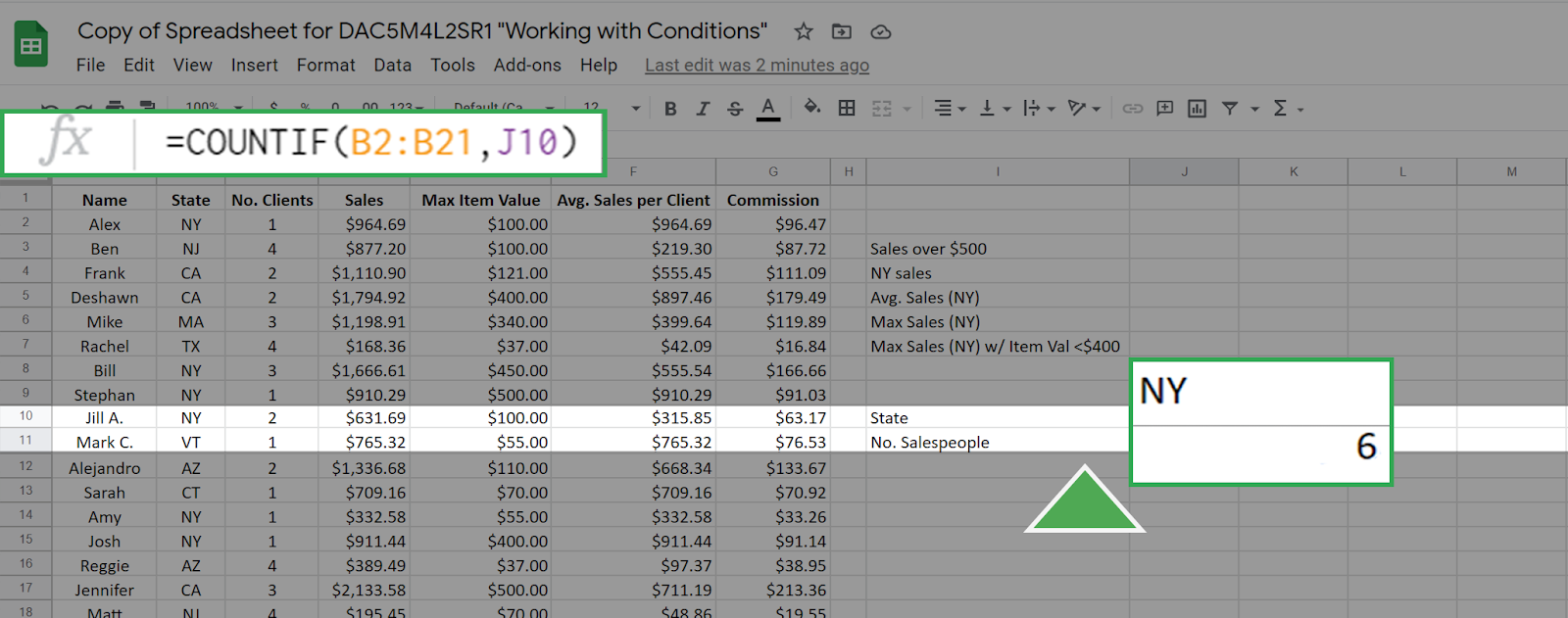
The range is the array (or collection) of cells that you are checking and the criteria is what you are checking for. All cells in the array that match the provided criteria will be counted and this number returned as the value of the function.

To use this function to count the number of salespeople working from “NY,” click on an open cell. In the function bar, enter **=COUNTIF(B2:B21, "NY")**.

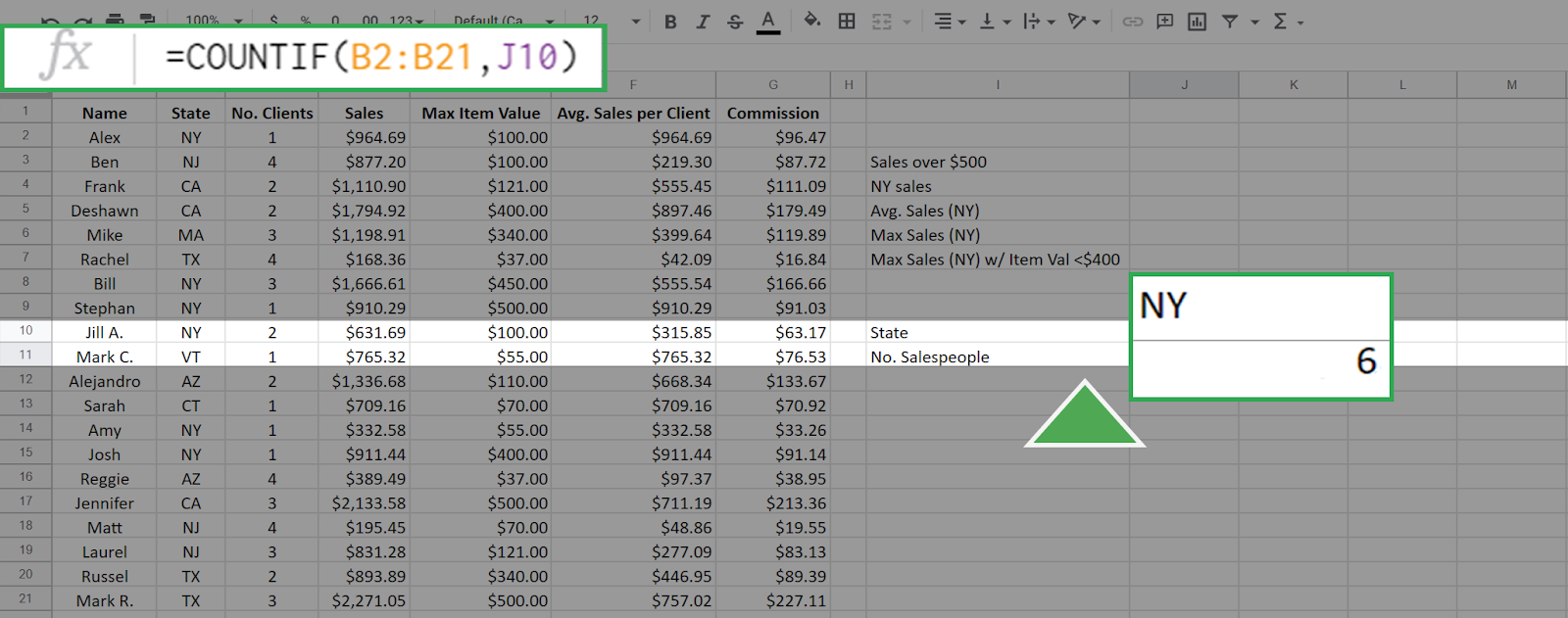


The range is the array of cells from B2 to B21. This is all from column B with the exception of the column header. The function checks all cells in this array against the value “NY” (entered in quotes) which is the criteria. Every cell in this array with a value of “NY” will be counted, and the result is returned in the cell. It is 6 in this case.

Press Enter/Return. The result should display like this:



**You can achieve the same result by entering a cell address as the criteria**. For example, the cell J10 has the value "NY." If you enter this in the function bar the **COUNTIF** function will seek out the value in cell J10 and use it as the criteria. This gives the same result as before:



Other examples of using **COUNTIF** include a hotel manager checking on how many rooms are currently occupied for staffing and inventory, or it could be used in a factory to count the number of defective items in a batch.

### 

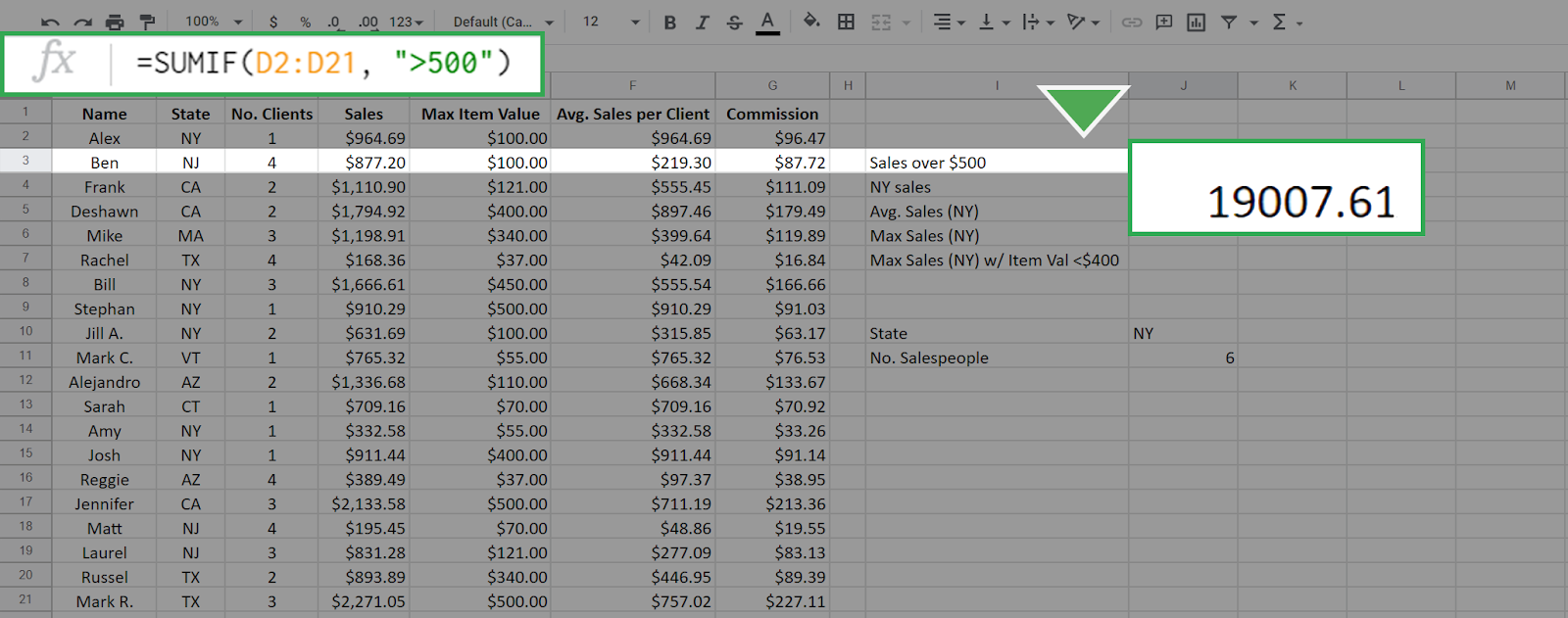
### **Step 3: Use the SUMIF function**

The **SUMIF** function is used to create a sum of the values of cells that meet a specific criteria. **SUMIF** supports the **logical operators** (>, <, <>, =). The syntax for this function is **=SUMIF(range, criteria, [sum\_range] )**.

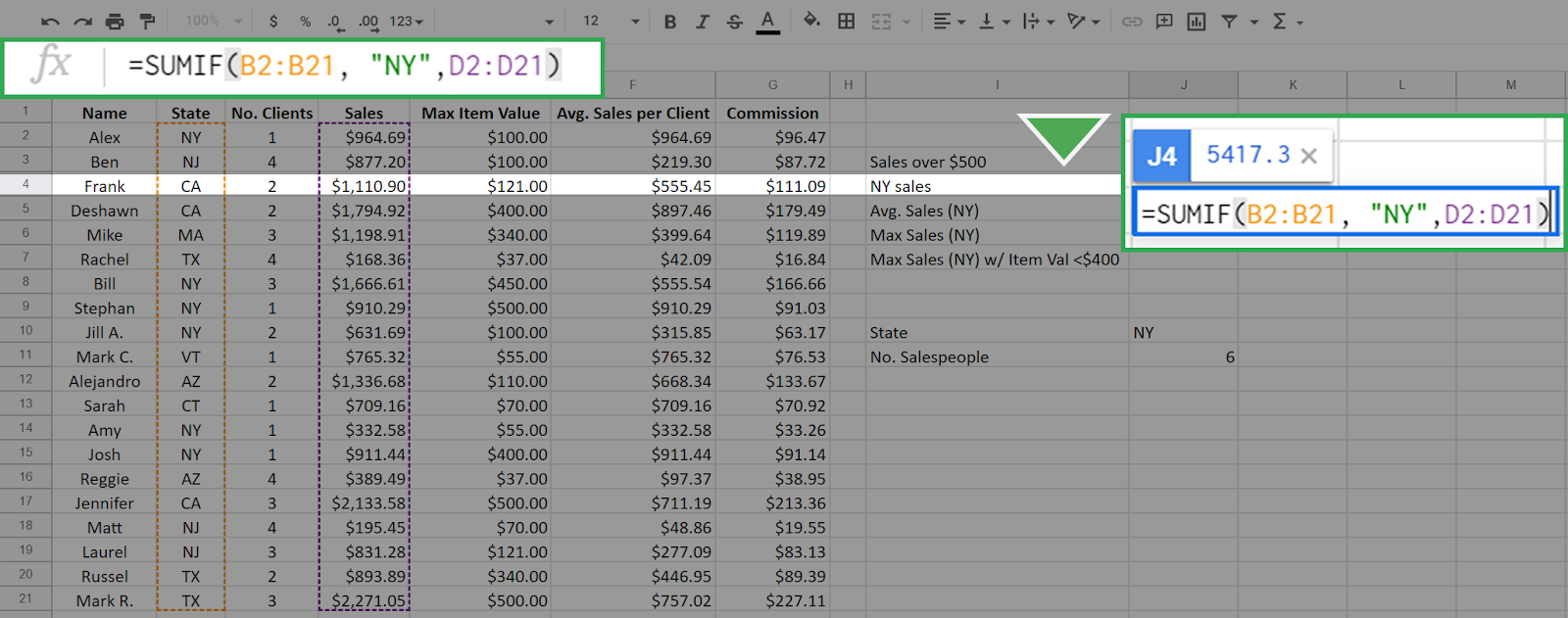
The range, also called the input range, is the array of cells that you check against the value of criteria. The **sum\_range** is the array of values that you will sum up if the criteria is met. **In this syntax above, the square brackets around sum\_range indicate that this input is optional. However, you do not add square brackets when writing the function**. If the argument **sum\_range** is absent, then the **SUMIF** will sum the values in range by default.

As an example of this function, suppose that you want to create a sum of all sales more than $500.00. This can be executed as **=SUMIF(D2:D21, ">500")**.

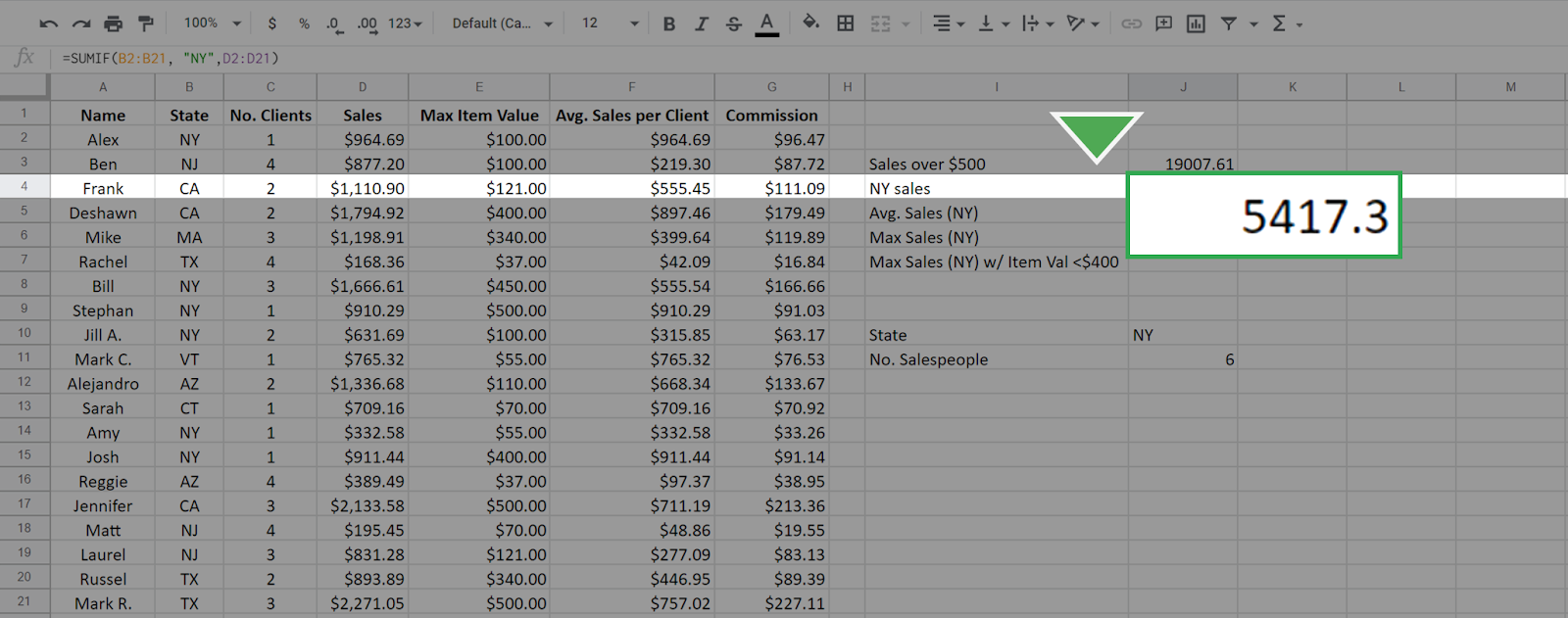
The result is 19007.61:



Because you didn't include the **sum\_range** input, all the values in the cells D2 to D21 that match the criteria were summed by default. To sum only the sales from New York, but not restrict to those greater than $500, enter the following function: **=SUMIF(B2:B21, "NY", D2:D21)**.



This results in 5417.3:



The first input, B2:B21, is the range of cells that are checked for the criteria "NY" and the summing is done across the **sum\_range** of cells D2:D21 that have the state meeting the criteria "NY." This is different than in the first case. In that case, the array that you check is the same array that you sum across.

Other examples of using **SUMIF** include tracking the total value of a specific product to better manage inventory. A medical accountant might use **SUMIF** to total the expenses related to a particular department or piece of equipment, such as an MRI machine.

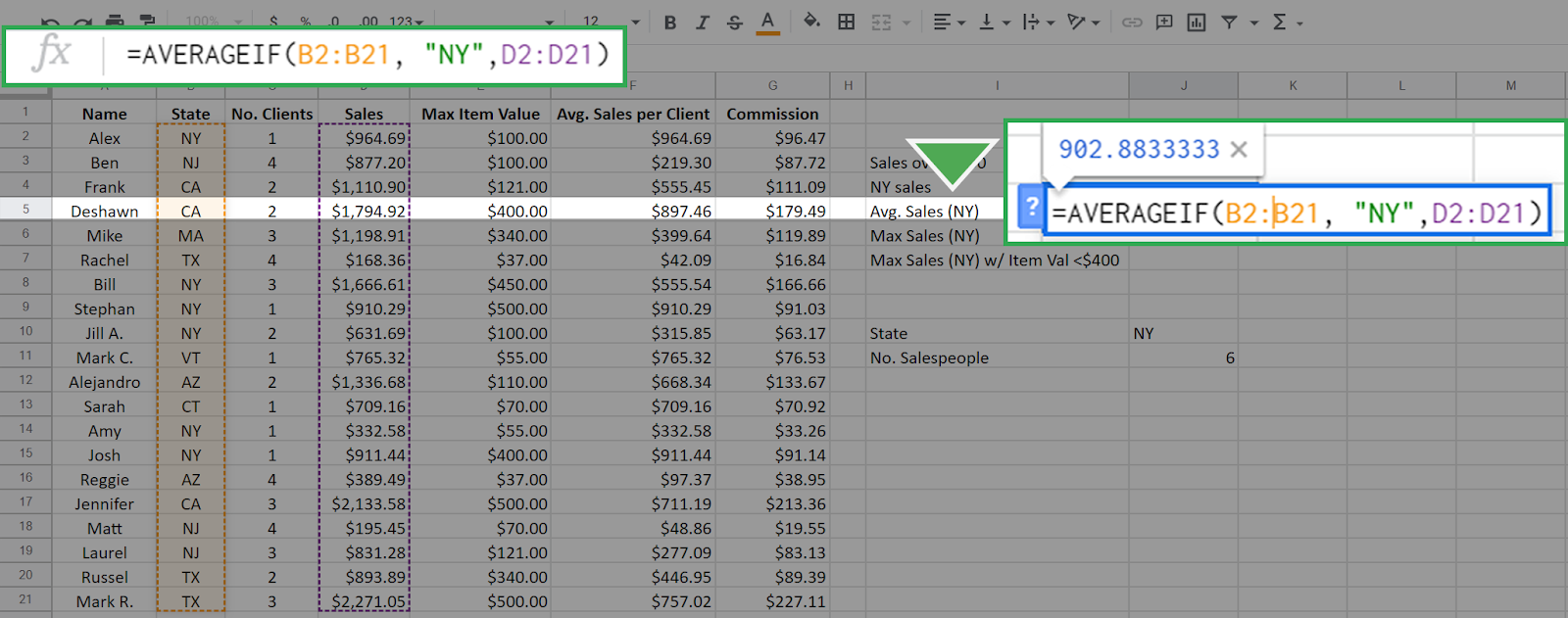
### 

### **Step 4: Use the AVERAGEIF function**

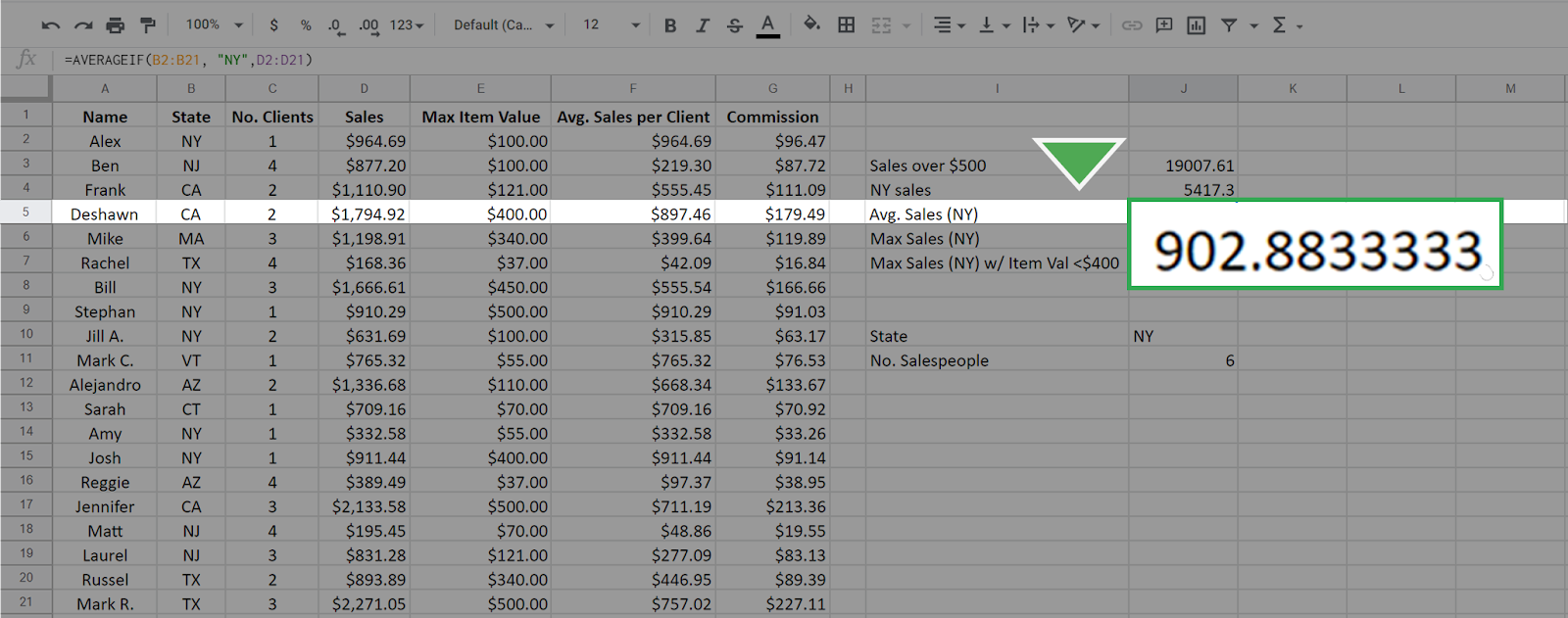
Just like the previous two functions, the **AVERAGEIF** function will average the values in an array based on a specific criteria. The syntax is **=AVERAGEIF(range, criteria, [sum\_range])**.

The inputs to this function—**range**, **criteria**, and **sum\_range**—work in exactly the same manner as in the **SUMIF** function. Again, the **sum\_range** is optional.

You know from your previous queries that there are six salespeople in New York. To find the average sales per for these salespeople, enter the following function: **=AVERAGEIF(B2:B21, "NY", D2:D21)**.



The result is 902.83333, which means that the average salesperson for this company in New York state sells $902.83 worth of merchandise.



Averaging data conditionally allows you to get more specific insights into your dataset. For example, you might want to know the average sales that are made on weekends to plan promotional events more effectively.

### 

### 

### **Step 5: Use the MAXIFS function**

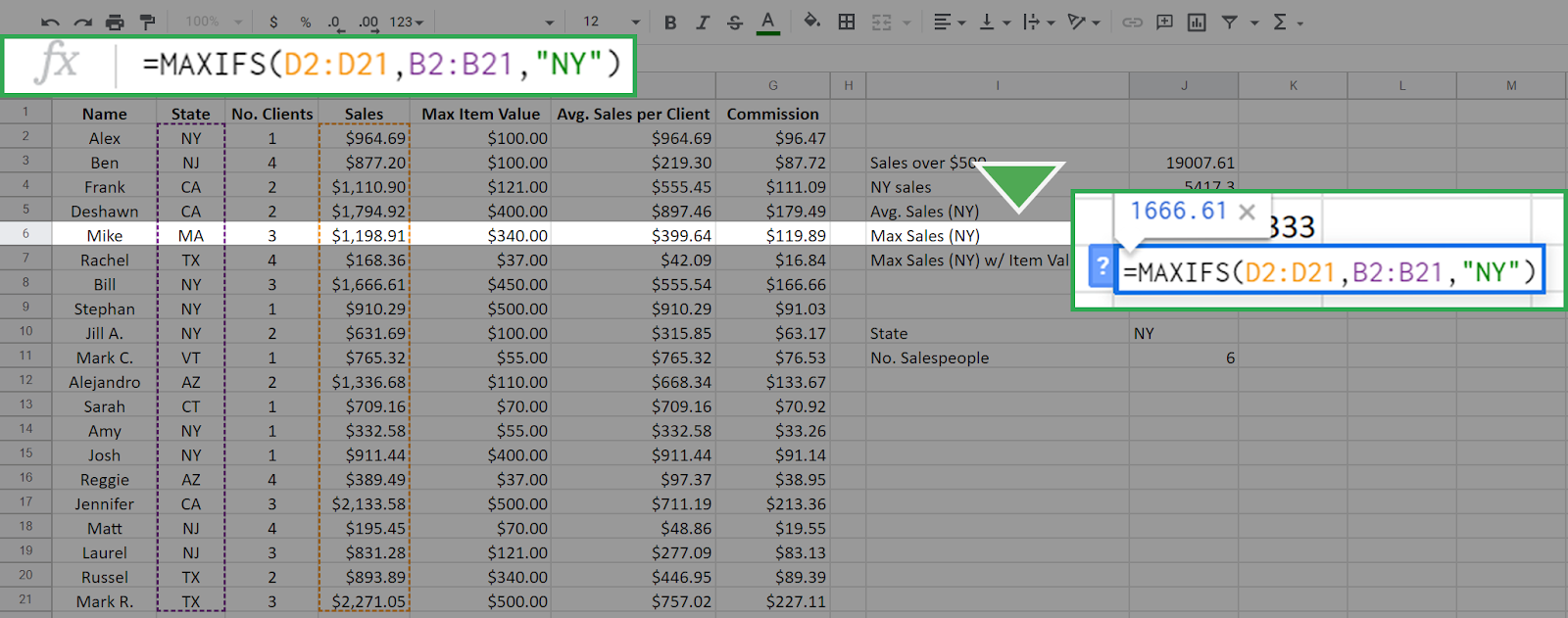
The **MAXIFS** function is slightly different from the other three functions—**it returns the maximum value in a range based on one or more conditions.**

The easiest way to observe the difference is to examine the **syntax: =MAXIFS(max\_range, range1, criteria1, [range2], [criteria2], ...).**

Note for Microsoft Excel: **MAXIFS** can only be used with an Office 365 subscription on Excel 2016 or newer. If you cannot use a version of Excel that allows the function **MAXIFS**, use Google Sheets for this part of the activity.

1. The first argument, **max\_range**, is the array for which you want to find the maximum value.
2. The second argument **(range1)** is the array you are checking.
3. The third argument **(criteria1)** is the value that you are checking for.
4. The inputs in the square brackets are for optional additional constraints.

To use **MAXIFS** to find the maximum sales from any salesperson in New York, enter the following: **=MAXIFS(D2:D21, B2:B21, "NY")**.



The resulting calculation is 1666.61. It’s important to keep in mind that the order in which you enter the inputs matters. If you reverse the position of the arrays and enter **=MAXIFS(B2:B21, D2:D21, "NY")** the result is 0.

This is because you are now asking the function to find the maximum of the array B2:B21 where the sales equal "NY". This is impossible because the values in the array D2:D21 (the sales array) are numerical—none of them equals "NY." The function returns 0 when nothing in the range meets the criteria.

The **MAXIFS** **function can input more than one constraint.**

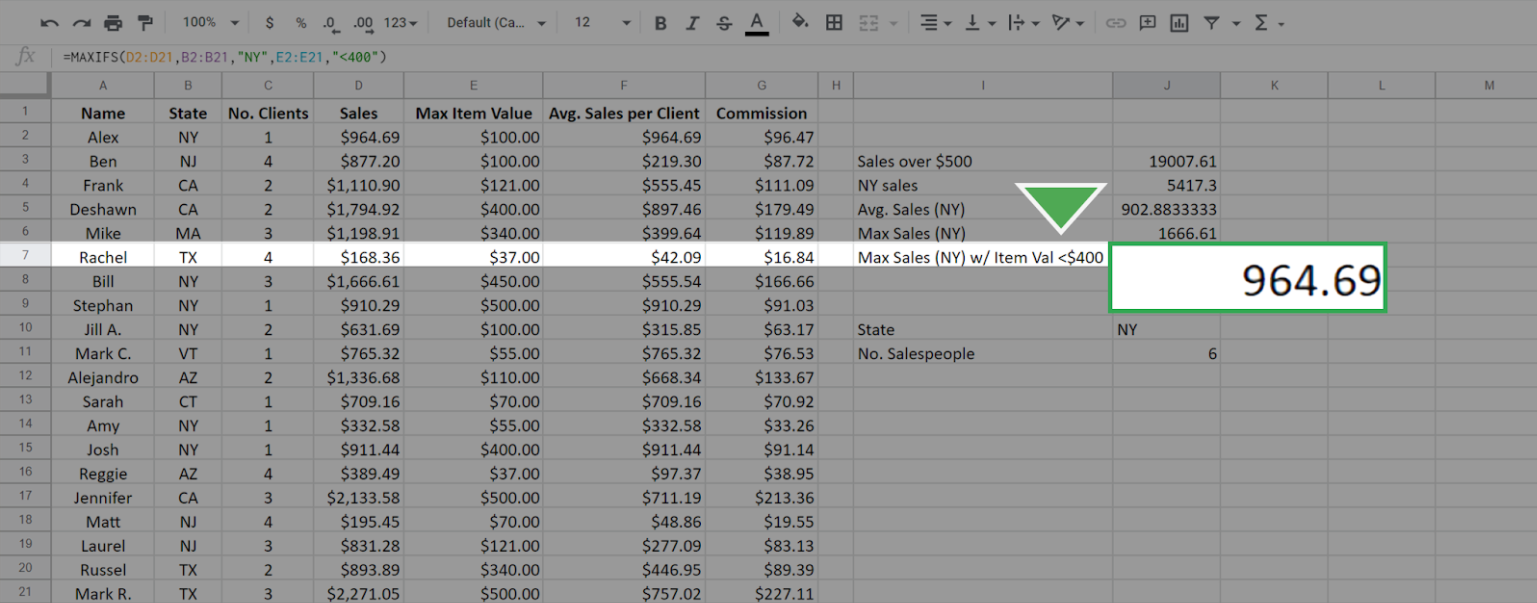
This is where the optional **range2** and **constraint2** come into play.

For example, to find the maximum sales in New York where the Max Item Cost is below $400, enter the following into the function bar: **=MAXIFS(D2:D21, B2:B21, "NY", E2:E21, "<400")**.



Additional constraints follow the logic that every constraint must be satisfied for a cell in the **max\_range** to be considered.

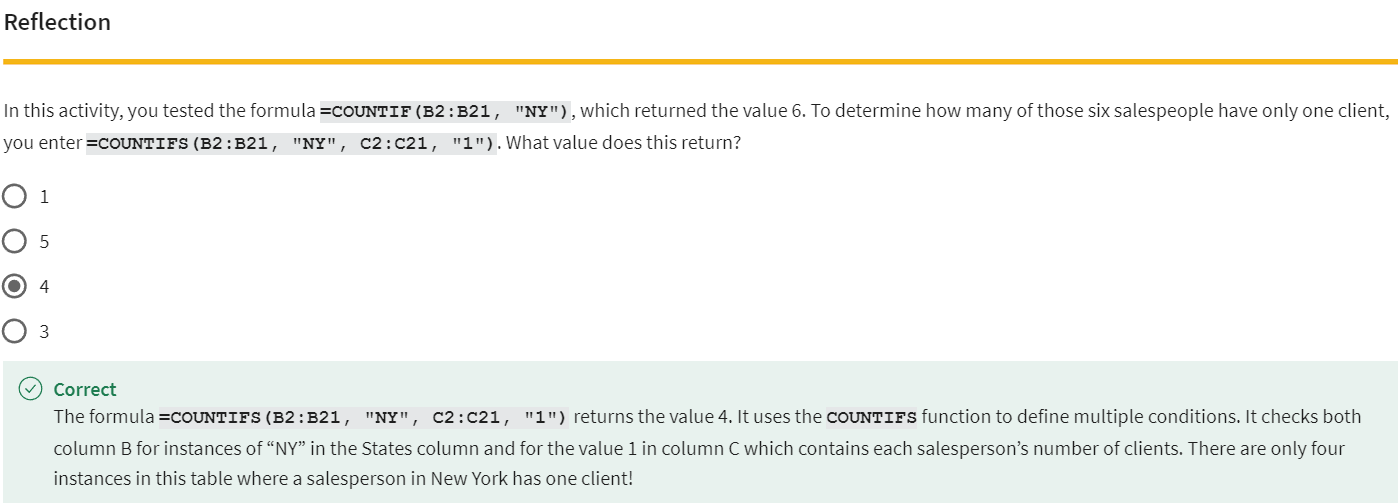
The first three inputs are the same as above, but now you've added the additional constraint that Max Item Value must be less than $400. The array E2:E21 is the Max Item array and its cells are checked against the criteria <400. The function returns the following, which is the maximum sales of any New York salesperson who did not sell any single item over (or equal to) $400.



### 

### **Step 6: Compare the conditional functions**

Each of the previous functions—**COUNTIF**, **SUMIF**, and **AVERAGEIF**—have equivalents that work similarly to **MAXIFS**. These include **COUNTIFS**, **SUMIFS**, and **AVERAGEIFS**. The syntax and functionality of these functions, apart from the specific calculation, are identical to **MAXIFS**. For example, the **SUMIFS** function will give the sum for single and multiple constraints just like **MAXIFS** function does for the maximum. It also has the same syntax as **MAXIFS**



.

### [COMPOSITE FUNCTIONS](https://www.coursera.org/learn/analyze-data/lecture/VNgdu/composite-functions)

Data analysts love discovering new ways to work on their analysis, especially when those new ways simplify their work. I know I'm a big fan of learning new tricks to complete tricky tasks.

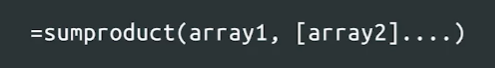
Instead of trying to find a new way to do something every time I do an analysis, I try to learn from other people by asking questions and getting help when I need it. **The people I work with like to use the phrase, stealing with pride**.

All this means is that you should feel no shame for using a process in your analysis that you learned from someone or somewhere else. Fellow team members, message board posts, online searches, I've used all of these resources for ideas. With pride! Of course, I always cite my sources when I do. That's a super important step to remember.

The SUMPRODUCT function is also one of those tricks that analysts come across either on their own or from another source. **You can also think of it as a shortcut for doing more complex calculations**. We'll show you how SUMPRODUCT works and when you might use it to make your work life simpler.

**SUMPRODUCT is a function that multiplies arrays and returns the sum of those products.**

Here's what the SUMPRODUCT formula looks like;

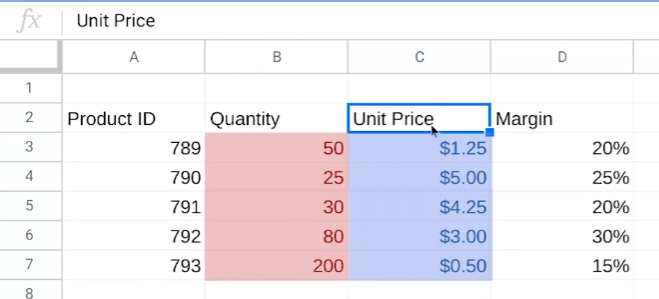


An array is kind of like a range in a spreadsheet. But keep in mind, **an array is a collection of values in cells, not the cells themselves.**

**Use this table to follow the example below:**[KP Kitchen Supplies, Profit](https://docs.google.com/spreadsheets/d/1tew5mVmK-wkAxJVzS2nZul6IJvK4tdbDfm3IjJfVWsE/template/preview)

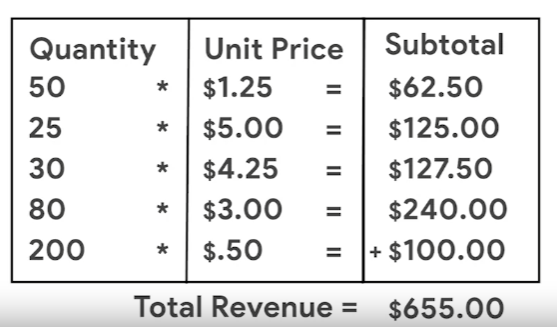
**When added to a formula, the SUMPRODUCT function multiplies each of the values in two or more arrays together**.

For example, each value in the array of cells B3 through B7 can be multiplied by its corresponding value in the array of cells C3 through C7. B3 times C3, B4 times C4, and so on.

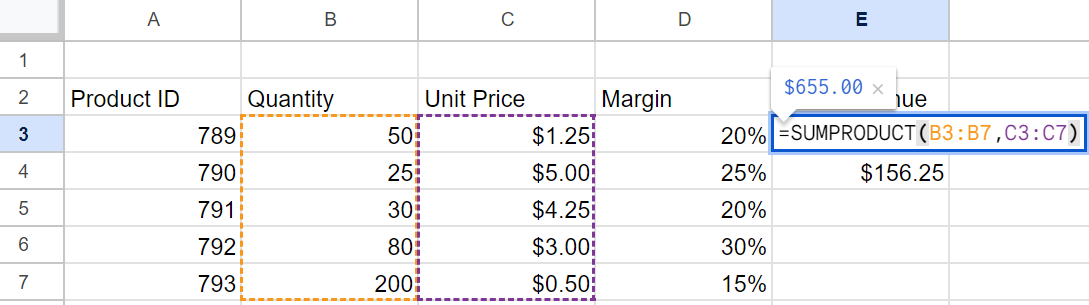


It will then return the sum of all of those multiplications.  
  
You might remember this example from our COUNTIF and SUMIF video.   
We've been given some data about a product order, including the quantity of each product that was sold in the order and the unit price, which tells how much one of each product cost. Our job is to use the data in these two columns to find out the total revenue for this order. That's where SUMPRODUCT comes in.

To find the total revenue, we need to do both addition and multiplication calculations. First off, we need to find the revenue that each item brought in separately. If we did this without SUMPRODUCT, we'd have to multiply each quantity by its unit price: 50 times $1.25, 25 times $5, and so on. Then we'd have to add all of those revenue amounts together to get the total revenue.



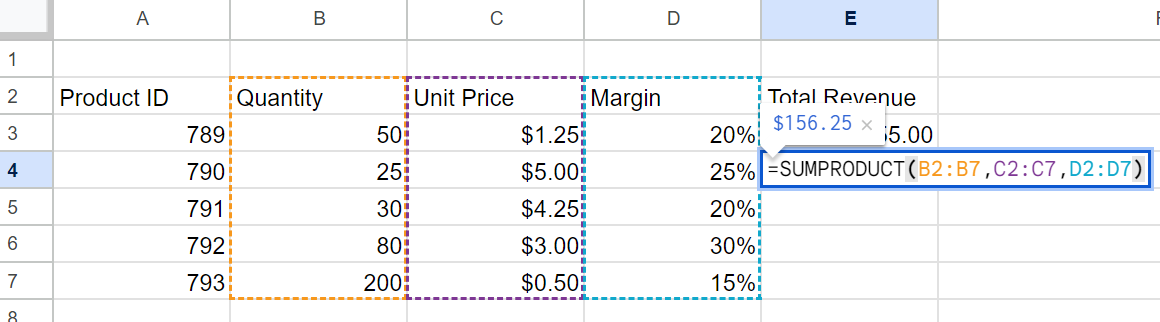
Fortunately, the SUMPRODUCT function does all of that for us. Let's add the label Total Revenue in cell G5 and then click G6 to input our formula. We'll then start our formula with an equal sign and the function followed by an open parenthesis. It's good to remind ourselves that the arrays we add to our formula should always be inside the parentheses. Next, we'll select cells B3 through B7 for the first array followed by a comma. The comma acts as a separator between the two arrays and the formula. Then, we'll select cell C3 through C7 for the second array, followed by a closed parenthesis to complete our formula. We don't need to include the brackets in our actual formula. We included them in the syntax example to clearly define each array for you. Then we press Enter to get our total revenue.



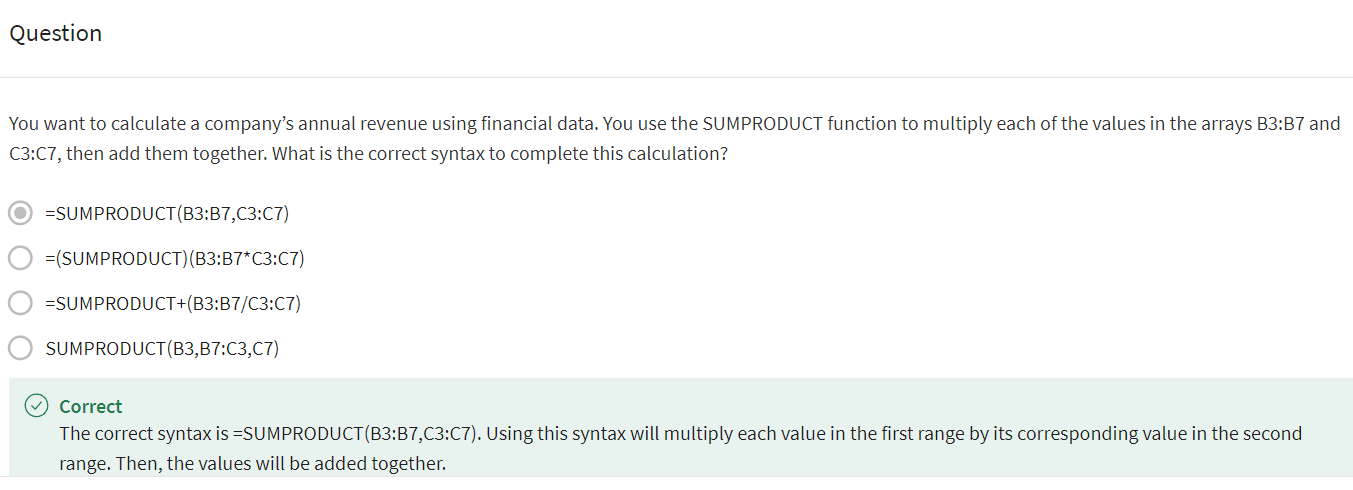
Since we're dealing with revenue, we'll format the number as currency. We've learned the total revenue is $655.

But that's not the actual profit from the sales of these kitchen supplies because we haven't included the profit margin in our calculations. The profit margin is a percentage that indicates how many cents of profit have been generated for each dollar of sale. In our dataset, product # 789 has a profit margin of 20 percent, meaning each product sold earns a total profit of $0.20 for every dollar. And just like the calculation for revenue, we can save time finding profit margin by using the SUMPRODUCT function. **There's only one difference between the formula for profit margin and revenue in this spreadsheet. But it's an important difference.** To start, in cell G7 we type the same first part of the formula. Then we include the two arrays in the same way as well.

But instead of ending our formula, we add another comma followed by another array. This time, we'll select the cells with a profit margin, D3 through D7. We'll finish our formula, and our calculation is complete.



The SUMPRODUCT function saved us from having to multiply each individual revenue amount by each profit margin percentage, then add each profit margin amount together. Using SUMPRODUCT for calculations is a time-saver and helps you avoid making mistakes. Definitely a trick worth remembering, and there's more worth remembering about calculations coming up next.



### [TEST YOUR KNOWLEDGE ON DATA CALCULATIONS](https://www.coursera.org/learn/analyze-data/quiz/HdCkh/test-your-knowledge-on-data-calculations)

### 

### 

## PIVOT…PIVOT…PIVOT…

### [START WORKING WITH PIVOT TABLES](https://www.coursera.org/learn/analyze-data/lecture/HCOme/start-working-with-pivot-tables)

By now, we've learned a lot about functions and formulas. **They are very helpful tools for your toolbox and great for finding shortcuts to complete calculations. But there's another tool out there that does some of the same things in a spreadsheet: the pivot table**.

As a quick reminder, pivot tables let you view data in multiple ways to find insights and trends. **We've talked before about how pivot tables help with cleaning and organizing your data including sorting and grouping data.**

**But pivot tables can also help with calculations. For example, they're great for quickly calculating sums and averages**.

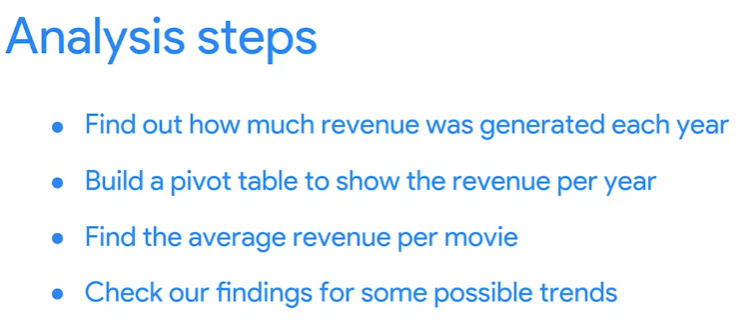
Let's revisit our movie data set to show you how pivot tables and calculations worked hand in hand.

[Movie Data Starter Project](https://docs.google.com/spreadsheets/d/1FLaUmMn62YlHYihV6pK1DJqWcFYCnuoqoxFWmm_o5b0/template/preview)

Earlier, we summarized and organized this data in pivot tables. We'll do that here too. But in this case the organization is a bonus to using pivot tables for calculations. You can do these steps in Excel as well, though some of the steps might look different. If you're using Exce.

In this example, your manager asks you to find some trends to help them think through new movie ideas using revenue calculations. This spreadsheet has data about movies from several years ago. So it probably wouldn't be as useful right now.Still, the steps we take to analyze the data absolutely apply then and now. So let's get into it.

First, we need to find out how much revenue was generated each year. A pivot table is a good way to organize this.So we'll build a pivot table to show this. In our pivot table, we can also find the average revenue per movie. We can then check our findings for some possible trends.



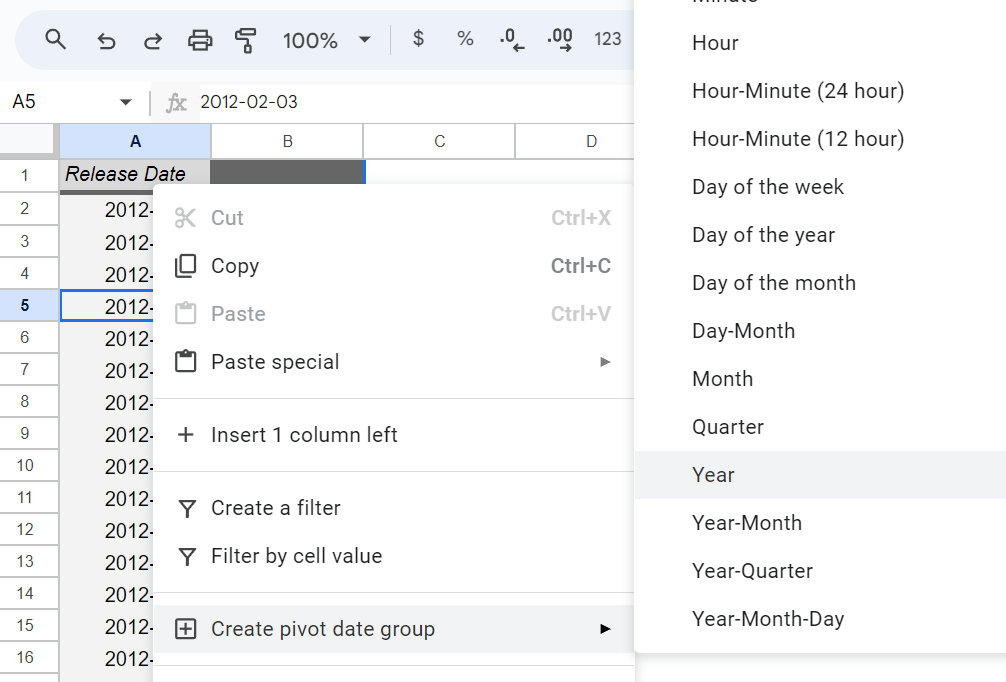
We'll start by finding the revenue generated each year. This gives us the release date for each movie in column B and the box office revenue in column N. Instead of reorganizing the table by year and building a formula to calculate the revenue per year we'll create a pivot table. Well, at the pivot table in a new sheet keeping the data range of cell A1 to cell N509.

Adding a new sheet is especially helpful when working in a large dataset like this. It helps keep our calculations together in one place and separate from the rest of the data.

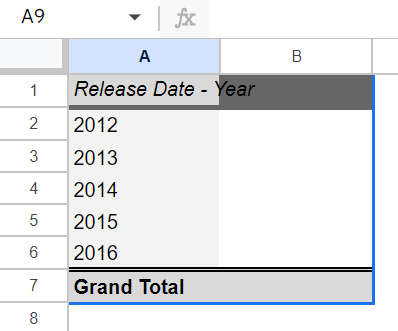
We'll rename this sheet revenue to call out where our calculations are, both for ourselves and for anyone else on our team who might need our analysis. Now we can build our pivot table, starting with the rows.

We'll sort the rows by release date to find out the revenue for each year. You might notice this creates a row for every date on which one or more movies in this dataset was released.

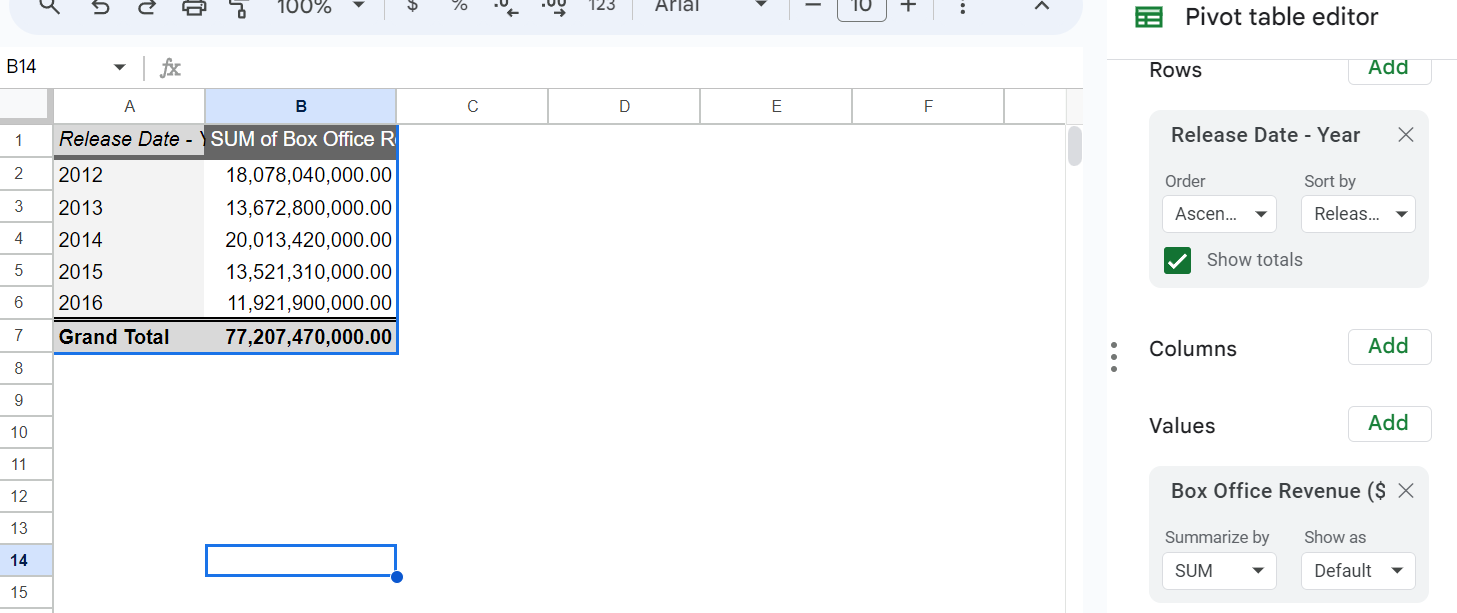
Since we only need the years, we'll right click in one of the cells in the release date column to create a pivot date group and group by year.



Now we have rows, one for each year in which these movies were released



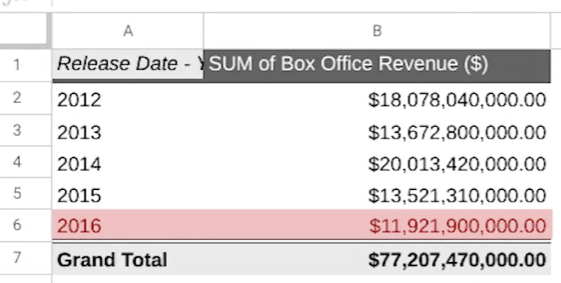
Next, we want to work with the values, add the box office revenue data on the Values field



This populates the columns next to the release dates with the total box office revenue and each year. These calculations are automatic because the pivot table is already set to summarize the data using the sum function. So no need to change this setting.

There are other functions and are summarized by menu though, such as min for the minimum amount of revenue each year and count for the number of movies that generated revenue in each year.

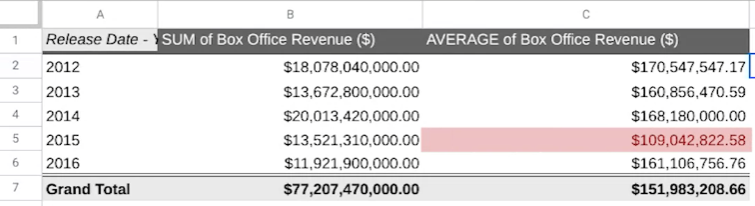
Okay, let's check out what we've got here.



This data shows that 2014 has the highest revenue, while 2016 has the lowest. This might be useful information, but finding the average revenue per movie would most likely be more useful since there was a different number of movies released each year.

So we'll add another column for the average revenue earned by each year's movie. We can do this in the same pivot table. We'll add another value and change the function that we use to summarize from sum to average.

The average function gives us the average revenue per year for the movies in the dataset. We can see that the average revenue in 2015 was much lower than the other years.



Since this data stands out so much, let's keep exploring to find out why.

Taking your analysis to another level like this is a sign of a great analyst. When you're in your job, you want to answer the questions that your manager and stakeholders ask. But you also want to answer the ones that come up while you're doing your analysis.

So let's try to figure this out. First, we'll know how many movies from each year were included in the dataset, we'll add a new value and use the count function this time.

This shows us that there are more movies in the data set from 2015 than from any other year. But 2015 still has the second lowest total box office revenue. This could mean a few things. It's likely that a lot of the movies from 2015 just didn't earn much revenue compared to the other years, which would bring down the overall average revenue. Even if the total revenue remained on par with the other years. We'll explore just this one possibility here. But you can always go further when you analyze data in your own job. It will depend on your objectives and the questions you need to answer. For now let's copy and paste our pivot table so we can test our hypothesis.We'll rename the columns and our copy table to differentiate them from our original table. We'll name them based on the data we'll be looking at, which I'll explain in the next video.

Now our copied pivot table is ready for us to test our hypothesis. Next, we're going to use filters to find out how many movies earned less than $10 million revenue in 2015. Then we'll also create a calculated field to determine what percentage of the total movies from that year they represent. I'll be here when you're ready to learn more about pivot tables.

### [PIVOT TABLES CONTINUED](https://www.coursera.org/learn/analyze-data/lecture/nXA6B/pivot-tables-continued)

In the last video, we created a pivot table of movie data and revenue calculations to help our manager think through new movie ideas. We used our pivot table to make some initial observations about annual revenue. We also discovered that the average revenue for 2015 was lower than other years even though more movies were released that year. We hypothesized that this was because more movies that earn less than $10 million in revenue were released in 2015. To test this theory, we created a copy of our original pivot table. Now we are going to apply filters in calculated fields to explore the data more. Let's get started. You all remember that the filter option lets us view only the values we need. We'll select a cell in our copied pivot table and add a filter to the box office revenue column. The filter will then be applied to the entire table. When we open the status menu, we can choose to filter the data to show specific values.

But in our case, we want to filter by condition so we can figure out how many movies in each year earn less than $10 million. The condition we'll use in our filter is less than and our value will be $10 million which is why we renamed these columns earlier. We'll type our number in a dollar and cents format so the condition matches the data in our pivot table. This might not be necessary, but it prevents potential errors from happening. Now we know that 20 movies released in 2015 made less than $10 million. This seems like a high number compared to the other years. But keep in mind, there were more movies from our data set released in 2015. Before we move on, let's use a calculated field to verify our average because it was copied from another pivot table before we filtered it. That way we can check that it's correct. We'll create a customized column called a calculated field using our values menu. A calculated field is a new field within a pivot table that carries out certain calculations based on the values of other fields. You can do this in Excel too using field settings and the create formula menu. For the formula in our calculated field, we'll use the sum function and divide the sum of the box office revenue data from our original table by the count of the same data. Because we applied our filter to this pivot table earlier, this formula will only return the average revenue of movies under $10 million. That worked. We were able to check the accuracy of some of our data before analyzing it. Always a good thing. But it's still difficult to tell how much of an impact these lower earning movies had on the average revenue. Let's run a quick formula to find the percentage of movies for each year that earned less than $10 million. This will make it easier to compare from year to year. Instead of a calculated field, we'll add this as a formula in a new column, that way we can pull data from both of our pivot tables. We'll put a header for our table in cell G10 and name it percent of total movies. Then we'll add our formula to the next cell in the column. Divide the number of movies in the copy table by the number of movies in the original table. Then we'll use the fill handle in the cell with a formula and drag it to apply the formula to the rest of the years. Finally, we'll format these numbers as percentages. Now our analysis shows that 16 percent of the movies released in 2015 earned less than $10 million of revenue. The other years are all close to 10 percent. This is one possible explanation for why the average revenue is comparatively low in 2015. In real life, we'd most likely need to take our analysis even further depending on our goals. But for now, we're all set. You've learned how you can use pivot tables to perform data calculations. It will take practice, but pivot tables are worth it because they do more than calculate. They organize and filter data too. Together we've covered functions, formulas, and pivot tables. All great tools to use in analysis. With practice and experience, it will feel like you've used them forever. Just take your time getting to know how they work. Keep exploring these videos and the readings. Great work.

### [ELEMENTS OF A PIVOT TABLE](https://www.coursera.org/learn/analyze-data/supplement/j6w9Z/elements-of-a-pivot-table)

### [USE PIVOT TABLES IN ANALYSIS](https://www.coursera.org/learn/analyze-data/supplement/qRo2l/use-pivot-tables-in-analysis)

### [HANDS-ON ACTIVITY: EXPLORE MOVIE DATA WITH PIVOT TABLES](https://www.coursera.org/learn/analyze-data/quiz/vuioG/hands-on-activity-explore-movie-data-with-pivot-tables)

### [TEST YOUR KNOWLEDGE ON PIVOT TABLES](https://www.coursera.org/learn/analyze-data/quiz/Qs5pX/test-your-knowledge-on-pivot-tables)

## LEARN MORE SQL CALCULATIONS

## [QUERIES AND CALCULATIONS](https://www.coursera.org/learn/analyze-data/lecture/mUbhl/queries-and-calculations)

## [UPLOAD THE AVOCADO DATASET TO BIGQUERY](https://www.coursera.org/learn/analyze-data/supplement/Y6c0d/upload-the-avocado-dataset-to-bigquery)

## [STEP-BY-STEP: EMBED SIMPLE CALCULATIONS WITH SQL](https://www.coursera.org/learn/analyze-data/supplement/NBGTM/step-by-step-embed-simple-calculations-with-sql)

## [EMBED SIMPLE CALCULATIONS WITH SQL](https://www.coursera.org/learn/analyze-data/lecture/RlnmJ/embed-simple-calculations-with-sql)

## [CALCULATIONS WITH OTHER STATEMENTS](https://www.coursera.org/learn/analyze-data/lecture/eiuvu/calculations-with-other-statements)

## [HANDS-ON ACTIVITY: CALCULATIONS WITH SQL](https://www.coursera.org/learn/analyze-data/quiz/IEXum/hands-on-activity-calculations-with-sql)

## [TEST YOUR KNOWLEDGE ON SQL CALCULATIONS](https://www.coursera.org/learn/analyze-data/quiz/CGpT6/test-your-knowledge-on-sql-calculations)

## THE DATA-VALIDATION PROCESS

## [CHECK AND RECHECK](https://www.coursera.org/learn/analyze-data/lecture/6tBpf/check-and-recheck)

## [TYPES OF DATA VALIDATION](https://www.coursera.org/learn/analyze-data/supplement/tQAED/types-of-data-validation)

## [HANDS-ON ACTIVITY: FROM SPREADSHEETS TO BIGQUERY](https://www.coursera.org/learn/analyze-data/quiz/tPlCp/hands-on-activity-from-spreadsheets-to-bigquery)

## [START VALIDATING](https://www.coursera.org/learn/analyze-data/ungradedWidget/VomOd/start-validating)

## SQL AND TEMPORARY TABLES

## [TEMPORARY TABLES](https://www.coursera.org/learn/analyze-data/lecture/tMnmz/temporary-tables)

## [HANDS-ON ACTIVITY: CREATE TEMPORARY TABLES](https://www.coursera.org/learn/analyze-data/quiz/Ac9jc/hands-on-activity-create-temporary-tables)

## [MULTIPLE TABLE VARIATIONS](https://www.coursera.org/learn/analyze-data/lecture/C1bQ7/multiple-table-variations)

## [WORK WITH TEMPORARY TABLES](https://www.coursera.org/learn/analyze-data/supplement/oGADZ/work-with-temporary-tables)

## [YOUR INTERMEDIATE GUIDE TO SQL](https://www.coursera.org/learn/analyze-data/supplement/W4DUE/your-intermediate-guide-to-sql)

## [USE CONNECTED SHEETS WITH BIGQUERY](https://www.coursera.org/learn/analyze-data/supplement/U5ojF/use-connected-sheets-with-bigquery)

## [TEST YOUR KNOWLEDGE ON USING SQL WITH TEMPORARY TABLES](https://www.coursera.org/learn/analyze-data/quiz/HXUhR/test-your-knowledge-on-using-sql-with-temporary-tables)

## MODULE 4 CHALLENGES

### [GLOSSARY TERMS FROM MODULE 4](https://www.coursera.org/learn/analyze-data/supplement/0raUI/glossary-terms-from-module-4)

### [MODULE 4 CHALLENGE](https://www.coursera.org/learn/analyze-data/exam/GtaS5/module-4-challenge)

## COURSE WRAP-UP

## [COURSE 5 GLOSSARY](https://www.coursera.org/learn/analyze-data/supplement/0p8b6/course-5-glossary)

## [CONGRATULATIONS! COURSE WRAP-UP](https://www.coursera.org/learn/analyze-data/lecture/4HYUY/congratulations-course-wrap-up)

## [COMING UP NEXT...](https://www.coursera.org/learn/analyze-data/supplement/0Jz6z/coming-up-next)

## 