## 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.

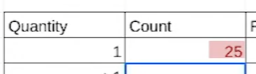
We'll also add borders to make the summary table nice and clear.

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.

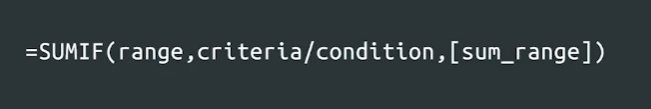


Getting this information helps us compare the data about quantity.

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.

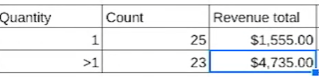
The condition we want the data to meet is for the values in the column to be equal to one. So we'll type a comma and then inside quotes an equal sign and the number one.

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.

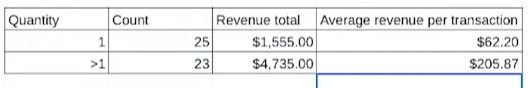
To complete our objective, we'll do two more quick calculations.

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**

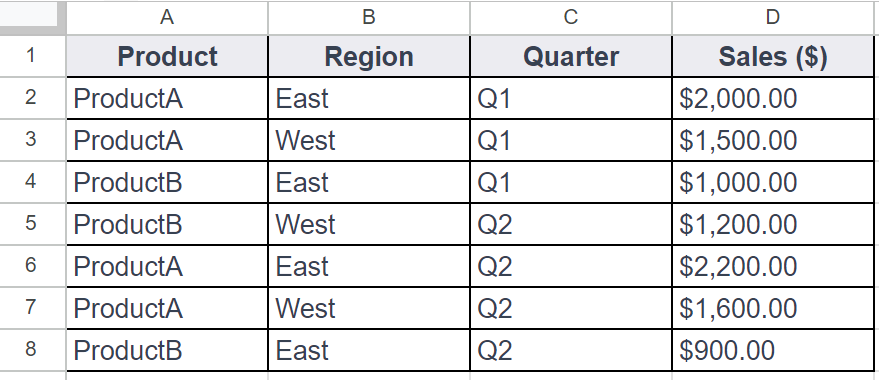
**Previously, you learned that the SUMIF function adds values in a particular range based on a single condition.**

**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 definition for **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 an 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)

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

### [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)

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

### [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)

## 