## FORMAT AND ADJUST DATA

As you move closer to analyzing your data, you’ll want to have it formatted and ready to go. In this part of the course, you’ll learn all about converting and formatting data, including how SQL queries can help you combine data. You’ll also find out the value of feedback and support from your colleagues and how it can lead to learnings that you can apply to your work.

**Learning Objectives**

* Explain what is involved in the conversion and formatting of data
* Demonstrate how to use SQL spreadsheets and SQL queries to combine multiple pieces of data
* Discuss the importance of seeking feedback and support from others

## FORMATTING FOR BETTER ANALYSIS

### [GET STARTED WITH DATA FORMATTING](https://www.coursera.org/learn/analyze-data/lecture/u1pom/get-started-with-data-formatting)

Hey, it's great to have you back! You've learned so much already, and now you're ready to start analyzing data. Coming up, we'll cover some final things you'll need to do for your analysis to make sure your data is formatted and adjusted correctly.

We'll start converting and formatting your data; using data validation in spreadsheets; and conditional formatting. You'll also learn how to combine multiple pieces of data. And finally, we'll talk about how to get support during your analysis and find resources whenever you're stuck.

These skills will help make sure that your data analysis process is as smooth as possible. And even when it isn't, you'll know how to tackle any problems that might come up.

A big piece of being an analyst is troubleshooting and problem-solving. You're as good of an analyst as your ability to ask the right questions, which is why we'll spend some time learning about problem- solving strategies you can use during analysis. So whenever you're ready to start learning about data formatting and solving problems, head to the next video, and we'll get started.

### 

### [STEP-BY-STEP: FROM ONE TYPE TO ANOTHER](https://www.coursera.org/learn/analyze-data/supplement/hxmAr/step-by-step-from-one-type-to-another)

This reading provides you with the steps the instructor performs in the following video, [From one type to another](https://www.coursera.org/learn/analyze-data/lecture/FOAwr/from-one-type-to-another). Watch as the instructor demonstrates how to format numbers and convert units of measurement in your spreadsheets.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

If you’d like to access the spreadsheets the instructor uses in this video, click the links to the dataset to create a copy. If you don’t have a Google account, you may download the data directly from the attachments below.

Link to movie data starter project: [Movie data starter project](https://docs.google.com/spreadsheets/d/1FLaUmMn62YlHYihV6pK1DJqWcFYCnuoqoxFWmm_o5b0/template/preview).

Link to weather table - data for convert: [Weather Table - Data for CONVERT](https://docs.google.com/spreadsheets/d/15VeWQLQ5lUKvywYJL-0cGqmehvE8OH8W9cOlJ2P0J_I/template/preview).

OR

[Movie Data Starter Project](https://d3c33hcgiwev3.cloudfront.net/7rkSg1b6TCeiM37bJ_iPQA_cc9c5dd03dbe4e09bc83a3c244e4d6e1_Movie-Data-Starter-Project.xlsx?Expires=1712793600&Signature=V38hV~TUx6loM-AHbVADkKsWAoKDDdyAzuZFzJHrFGt4oZzBHoejm2y6Srx-82nlgXffh~D09xoYdTu5YraNcgwgW2UeUO2sVUTrrzm-UJ2m0C9Y3WIUKu1OtxdFS116e52sHmtpMNTLpv4MRGphLWZOr8YlQxqtqSZkO6CtWw0_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)[Weather Table - Data for CONVERT](https://d3c33hcgiwev3.cloudfront.net/dtqtPkiTSV27NcIQALH5gw_e797db6bf7374b11b5e41841e34d71e1_Weather-Table---Data-for-CONVERT.xlsx?Expires=1712793600&Signature=khJsIYOxMAylEYExgMABIDC1r~Gmku2ZjxJ-hpRehvvqrapYBJL3NLUaCotu9YgUaHY1gmQeTO1zp7pTjOHIvKr8czUH535hW3dNDeXQBw19dHq6xqnyebnZdhONB1S4cQhkTXFZkRAuX57aq3iR~nCEbIJ9YH8QLGsnhilFfJ0_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[empty alt text](https://d3c33hcgiwev3.cloudfront.net/dtqtPkiTSV27NcIQALH5gw_e797db6bf7374b11b5e41841e34d71e1_Weather-Table---Data-for-CONVERT.xlsx?Expires=1712793600&Signature=khJsIYOxMAylEYExgMABIDC1r~Gmku2ZjxJ-hpRehvvqrapYBJL3NLUaCotu9YgUaHY1gmQeTO1zp7pTjOHIvKr8czUH535hW3dNDeXQBw19dHq6xqnyebnZdhONB1S4cQhkTXFZkRAuX57aq3iR~nCEbIJ9YH8QLGsnhilFfJ0_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

## **Example 1: Check and change data type**

Check your data for inconsistent units of measurement to prevent problems during data analysis.

1. Open the [**Movie Data Starter Project**](https://docs.google.com/spreadsheets/d/1FLaUmMn62YlHYihV6pK1DJqWcFYCnuoqoxFWmm_o5b0/template/preview) spreadsheet using the link in the video.
2. Select **Column M** [Budget ($)] and **Column N** [Box Office Revenue ($)].
3. From the menu, select **$**, the currency shortcut key.
4. Notice that the currency in **Columns M** and **N** are now formatted correctly.

## **Example 2: Convert temperatures from Fahrenheit to Celsius**

Use the **CONVERT** function to change units of measurement.

1. Open the [**Weather Table - Data for CONVERT**](https://docs.google.com/spreadsheets/d/15VeWQLQ5lUKvywYJL-0cGqmehvE8OH8W9cOlJ2P0J_I/template/preview) spreadsheet using the link in the video.
2. Select cell **F2** and begin typing the Convert function formula as **=CONVERT**.
3. Indicate the cell you want to convert. After **=CONVERT**, enter **(B2,**.
4. Indicate the conversion you’d like to make: from Fahrenheit to Celsius. Enter **"F", "C")**.
5. The formula in its entirety should look like this: **=CONVERT (B2, "F", “C”)**.
6. Cell **F2** now contains the temperature from cell **B2** in Celsius.
7. Calculate temperature in Celsius for the rest of the column. Hover over cell **F2** and select the fill handle, a small circle on a corner of the cell. Drag the fill handle to cell **F193** to convert the other cells in the column to Celsius.

**Note**: Would you like more practice? Try converting the wind speed in Column **D** from miles per hour (mph) to meters per second (m/s) using **CONVERT**. In cell **H2**, enter: **=CONVERT(D2, "mph", "m/s")**.

You can check if your conversion is correct by entering 8.5248 in a metric conversion tool, [metric-conversions.org/speed/miles-per-hour-to-meters-per-second.htm](https://www.metric-conversions.org/speed/miles-per-hour-to-meters-per-second.htm).

## **Example 3: Lock data in a table**

Using functions to convert data can lead to problems, which data professionals must be prepared to fix. For example, if a reference value changes, the calculated value also changes. Locking data in a table by changing it from a function to a value ensures a cell stays consistent even if the data around it changes.

1. Select cell **F2**. In the formula bar, notice that the contents of this cell are the function you entered in the previous example.
2. Right-click cell **F** and select **Copy** from the drop-down menu.
3. Right-click cell **G** and select **Paste special** from the drop-down menu. Then, select **Paste values only**. This option pastes only the values from the original selection, removing any formatting, functions, or other information.
4. Select cell **G2**.
5. In the formula bar, notice that the contents of this cell is a value. This means that the value won’t change when other cells change.

### 

### [FROM ONE TYPE TO ANOTHER](https://www.coursera.org/learn/analyze-data/lecture/FOAwr/from-one-type-to-another)

Hey there! So far, we've learned about **typecasting data with SQL** as a way of converting data from one type to another in databases. Now I want to check out another way to format data types within spreadsheets. In this video, we'll talk more about why making sure your data is formatted properly is so important and how to format numbers and convert units of measurement in your spreadsheets. Let's get started.

Sometimes, you need to convert data when you're working with spreadsheets. That might mean changing numbers into dates, strings, percentages, or even currency.

It's important to double check that all of your data is in the right format for your analysis. Sometimes even after cleaning and processing data, it still might not be in the right format you need.

Let's think back to the table with movie data from before. There were a lot of different data types that included numbers, such as dates, budgets, and text strings, like actors' names.

These are distinct values, but the spreadsheet doesn't always automatically know that. Here's an example. Let's say you wanted to sort the movies in this spreadsheet by most recent. If the spreadsheet cast them as strings instead of dates, it might sort them alphabetically. Until you change the data type, you won't be able to sort them the way you want. It's also possible that your datasets contain inconsistent units of measurement that you'll need to convert. Like say, a table that includes both US dollars and English pounds. That's why it's important to check those data types again, so you don't run into any problems during the actual analysis. Think about the incorrectly cast dates in our movie table. If your boss needed a list of the 20 most recent movies, but your spreadsheet was organized alphabetically instead of the most recent, you wouldn't be giving her the list of movies she needed. Incorrectly formatted data can lead to time-consuming mistakes in your analysis, and might end up affecting your stakeholders' decision-making.

But taking the time early on to convert and format your data can help you avoid that.

And now that you know why you'll need to convert data types while working in spreadsheets, let's find out how.

First, let me show you a really useful menu for specifying data types in spreadsheets. Here's the movie data table we used before, but now the money columns aren't typed as currency. On the toolbar at the top of the sheet, you'll find a menu that can help you convert these numbers into specific data types. It gives you a lot of choices just from the drop-down menu, such as number, currency, date, percentage.... And if you click to open the full menu, there's even more options, including one for a custom number format. We know that we want these columns to be in currency format, so let's do that. All I have to do is select this column and then hit the currency shortcut.

And now it's all typed correctly. But it doesn't stop there. You can go even further and convert the unit of measurement you're using. For this example, let's check out a different table. Imagine that you're working with a weather channel to gather data about daily temperatures. You have a table with some data about daily observations on the temperature, wind speed, and precipitation in this area. Right now, the temperatures are in Fahrenheit, but for your analysis you need them to be in Celsius. No problem. All you need to do is use the CONVERT function to change the unit of measurement. We'll use this empty column here. Here's the first temperature in the table. We'll input the CONVERT function in our new column to change it to Celsius. Then we need to put what cell we want converted. And finally, we're going to convert.

And presto! Now this cell has the right unit of measurement for your analysis. You can simply apply it to the rest of this column. Now this temperature data is all in Celsius, and your unit of measurement is consistent across the table. And here's another tip. When adding data to tables using a formula, go back and paste the data in as values afterwards. That way they're locked in. Otherwise the cell stays as a formula and could get confusing when you start working with the data. So let's do that now. We'll copy the values and then right click in a new column. There's an option for "Paste special." And there's an option to "Paste values only." And now we have the static values in this column. Making sure your data is in the right format before you start analysis is so important. Do this, and your analysis will return the kinds of answers you're really searching for. And now you know some ways to typecast numbers and convert units of measurement in spreadsheets. You can feel confident your data is formatted the right way.

### 

### [CONVERT DATA IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/supplement/H7GTe/convert-data-in-spreadsheets)

In this reading, you will learn about converting data from one format to another. One of the ways to help ensure that you have an accurate analysis of your data is by putting all of it in the correct format. This is true even if you have already cleaned and processed your data. As a part of getting your data ready for analysis, you will need to convert and format your data early on in the process.



As a data analyst, there are lots of scenarios when you might need to convert data in a spreadsheet:

### 

### **String to date**

* [**How to convert text to date in Excel**](https://www.ablebits.com/office-addins-blog/2015/03/26/excel-convert-text-date/#:~:text=Excel%20DATEVALUE%20function%20%2D%20change%20text,Excel%20recognizes%20as%20a%20date.&text=So%2C%20the%20formula%20to%20convert,stored%20as%20a%20text%20string.): Transforming a series of numbers into dates is a common scenario you will encounter. This resource will help you learn how to use Excel functions to convert text and numbers to dates, and how to turn text strings into dates without a formula.
* [**Google Sheets: Change date format:**](https://www.ablebits.com/office-addins-blog/2019/08/13/google-sheets-change-date-format/) If you are working with Google Sheets, this resource will demonstrate how to convert your text strings to dates and how to apply the different date formats available in Google Sheets.

### **String to numbers**

* [**How to convert text to number in Excel:**](https://www.ablebits.com/office-addins-blog/2018/07/18/excel-convert-text-to-number/)Even though you will have values in your spreadsheet that resemble numbers, they may not actually be numbers. This conversion is important because it will allow your numbers to add up and be used in formulas without errors in Excel.
* [**How to convert text to numbers in Google Sheets:**](https://productivityspot.com/convert-text-to-numbers-google-sheets/)This resource is useful if you are working in Google Sheets; it will demonstrate how to convert text strings to numbers in Google Sheets. It also includes multiple formulas you can apply to your own sheets, so you can find the method that works best for you.

### **Combining columns**

* [**Convert text from two or more cells:**](https://support.microsoft.com/en-us/office/combine-text-from-two-or-more-cells-into-one-cell-81ba0946-ce78-42ed-b3c3-21340eb164a6)Sometimes you may need to merge text from two or more cells. This Microsoft Support page guides you through two distinct ways you can accomplish this task without losing or altering your data. It also includes a step-by-step video tutorial to help guide you through the process.
* [**How to split or combine cells in Google Sheets:**](https://www.techrepublic.com/article/how-to-split-or-combine-text-cells-with-google-sheets/)This guide will demonstrate how to to split or combine cells using Google Sheets specifically. If you are using Google Sheets, this is a useful resource to reference if you need to combine cells. It includes an example using real data.

### 

### **Number to percentage**

* [**Format numbers as percentages:**](https://support.microsoft.com/en-us/office/format-numbers-as-percentages-de49167b-d603-4450-bcaa-31fba6c7b6b4) Formatting numbers as percentages is a useful skill to have on any project. This Microsoft Support page will provide several techniques and tips for how to display your numbers as percentages.
* [**TO\_PERCENT:**](https://support.google.com/docs/answer/3094284?hl=en) This Google Sheets support page demonstrates how to use the **TO\_PERCENT** formula to convert numbers to percentages. It also includes links to other formulas that can help you convert strings.

**Pro tip:** Keep in mind that you may have lots of columns of data that require different formats. Consistency is key, and best practice is to make sure an entire column has the same format.

## **Additional resources**

If you find yourself needing to convert other types of data, you can find resources on [**Microsoft Support**](https://support.microsoft.com/)for Excel or [**Google Docs Editor Help**](https://support.google.com/docs/?hl=en#topic=1382883) for Google Sheets.

Converting data is quick and easy, and the same functions can be used again and again. You can also keep these links bookmarked for future use, so you will always have them ready in case any of these issues arise. Now that you know how to convert data, you are on your way to becoming a successful data analyst.

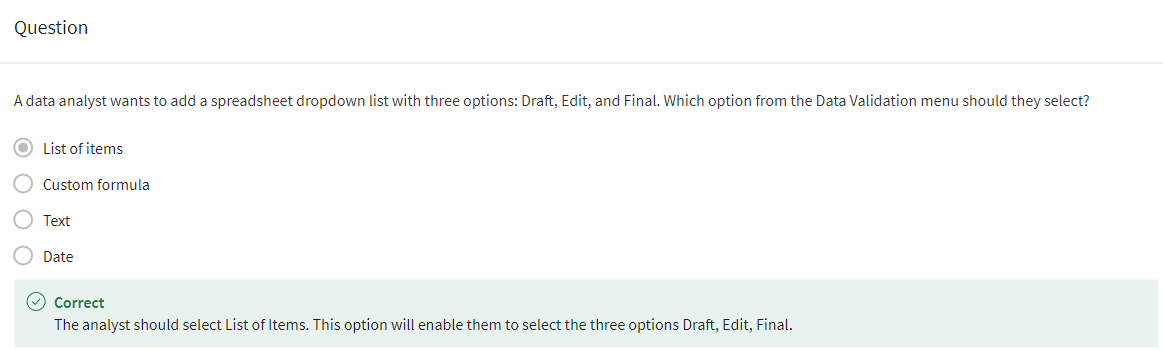
### [DATA VALIDATION](https://www.coursera.org/learn/analyze-data/lecture/glrSU/data-validation)

Welcome back! While we're learning about formatting data, I want to talk to you about another spreadsheet feature: **data validation**.

I'll teach you a little bit about data validation and show you how to use it. **For now, when I say data validation, I'm talking about the function, which is different from the data validation process**. First, let's talk about **what data validation does in spreadsheets**. Basically, it allows you to control what can and can't be entered in your worksheet.

Usually, data validation is used to add drop-down lists to cells with predetermined options for users to choose from. If you have a spreadsheet with a lot of collaborators, this can make it easier for them to interact with your table. You can think of it like a multiple choice question on a quiz. Since you control what's being entered into the worksheet, it cuts down on how much data cleaning you have to do later on.

For this example, we'll work on a project with a lot of milestones and deadlines to keep track of. Let's say our team has a spreadsheet that tracks everyone's progress. But instead of making everyone write in where they are in their task individually, we can provide a drop-down menu with multiple options, like "Not Yet Started," "In Progress," and "Ready." So we'll select the column that we want to add the drop-down menus to, in this case, the "Status" column. Then we'll go to the Data pull-down menu here at the top and click "Data validation." This brings up a pop-up menu with options for data validation. In this case, we know that we want to add a list of items for other users to choose from. So we'll select the "list of items" option from the possible criteria and type in the selections we want to create. Then hit Save, and now all of those cells have drop-down menus that we can use to easily mark progress for each task. But there's other things that you can do with data validation and spreadsheets, too, like creating custom checkboxes. To do this, let's select the cells under the "Review" column to make a checkbox that will let us know if tasks have been approved or not. We'll go back to the data validation menu. But instead of choosing "List from a range," we'll choose "Checkbox." There's an option to use custom cell values. Let's choose that and put in "Approved" and "Not approved." Now these tasks can be checked off by whoever's reviewing them, like a project manager, for example. Another way we can use data validation is to protect structured data and formulas. The more people that are working together in a spreadsheet, the more likely someone can accidentally break a formula.



But good news: the data validation menu has an option to reject invalid inputs, which helps make sure our custom tools will continue to run correctly, even if someone puts the wrong data in by mistake. All right, now you know three uses for data validation in your spreadsheets: adding drop-down lists, creating custom checkboxes, and protecting structured data and formulas. Data validation can help your team track progress, protect your tables from breaking when working in big teams, and help you customize tables to your needs. Coming up, we'll learn more about conditional formatting and some ways you can use conditional formatting and data validation together.

### [CONDITIONAL FORMATTING](https://www.coursera.org/learn/analyze-data/lecture/0WVmi/conditional-formatting)

Earlier we talked about **conditional formatting** as a spreadsheet tool that changes how cells appear when values meet specific conditions. This lets you add visual cues to your spreadsheets that make it easier to understand your table at a glance, and it makes the information in the worksheet clearer to your stakeholders.

We'll take that even further by combining conditional formatting and data validation to create custom tools for our spreadsheets. So far, we've used conditional formatting to highlight empty cells that still needed data so that we could quickly pinpoint what information our table was missing and add it in. Now, let's build on that by using it to make our scheduling table easier to read at a glance.

Here's a table we worked with when we covered data validation. It's tracking the status of different tasks on our project for our team to check on. But now there's even more tasks than the last time we looked at it. This table has useful information, but it takes a second to understand. Right now we don't have a visual on how many tasks are in progress or how many upcoming deadlines there are. But if we color-coded those elements of the table, we could quickly see key pieces of data really easily. Let's start with the Status column, column C. In the last example, we created these drop- down menus with the data validation tool. Now we can use conditional formatting to add some color. Let's go to the conditional formatting option under the Format menu.

This brings up a sidebar where we can select our range rule in formatting style. We need to decide which rows to apply our formatting to when the condition we set is met. We can click this button in the range options to select all of the rows we're applying the formatting to instead of typing it in. Now that we have those cells selected, we can choose the rule that we want to apply to these cells. We already have drop-down menus with specific text. So we can choose "Format Cells if... Text is exactly" from the rules. For our first rule, let's write "Not Yet Started" as the text condition. Then we'll choose a color to apply to those cells that have "Not Yet Started" in them. Let's use red. Now all cells that have "Not Yet Started" selected from the drop- down menu will be red.

Let's hit the "Add another rule" button to add conditional formatting to other status options. Let's add the condition "In Progress" next. We can make that one yellow. And then we'll add one last rule for "Ready." Let's choose green. And there. Now we have an easy-to- understand visual cue that tells us how many tasks are in progress, and how many are completed. We can also combine data validation and conditional formatting to track upcoming deadlines. We have a column of dates called "Review By This Date." First, let's use the data validation functionality to make sure users only enter valid dates. We'll go back to the Data dropdown at the top, pull up Data validation, and select Date as our criteria.

Then we can go to the Format menu at the top. Go down to conditional formatting and open the sidebar again. We'll click the "Select range" icon and select the "Review By This Date" column. Now under Format rules, we can select "Date is after," which will give us another option. Let's choose "today."

And finally, let's choose the color for these cells. So if the date listed in these rows is after today, it'll be filled in orange. You can also choose a specific locked date if needed. But for now, let's go with today. Now all of the upcoming review dates have an easy-to-see color code, so anyone using this table can quickly reference these deadlines. You'll find that some spreadsheet programs, like Excel, have built-in color codes that you can use, too. And there you go. Now you know how to use data validation and conditional formatting to create custom tools and visual cues that make your information easy to understand. There's a lot of different ways to use these tools, so feel free to experiment with them in your own spreadsheets. Coming up, we'll keep learning about new tools for spreadsheets and SQL. Bye for now.

### 

### 

### [TRANSFORM DATA WITH SQL](https://www.coursera.org/learn/analyze-data/supplement/HqeNj/transform-data-with-sql)

Data analysts usually need to convert data from one format to another to complete an analysis. But what if you are using SQL rather than a spreadsheet? Just like spreadsheets, SQL uses standard rules to convert one type of data to another. If you are wondering why data transformation is an important skill to have as a data analyst, think of it like being a driver who is able to change a flat tire. Being able to convert data to the right format speeds you along in your analysis. You don’t have to wait for someone else to convert the data for you.



In this reading, you will go over the conversions that can be done using the **CAST** function. There are also more specialized functions like **COERCION** to work with big numbers, and **UNIX\_DATE** to work with dates. **UNIX\_DATE** returns the number of days that have passed since January 1, 1970 and is used to compare and work with dates across multiple time zones. You will likely use **CAST** most often.

## 

## **Common conversions**

The following table summarizes some of the more common conversions made with the **CAST** function. Refer to [Conversion Rules in Standard SQL](https://cloud.google.com/bigquery/docs/reference/standard-sql/conversion_rules) for a full list of functions and associated rules.

| **Starting with** | **CAST function can convert to:** |
| --- | --- |
| Numeric (number) | - Integer  - Numeric (number)  - Big number  - Floating integer  - String |
| String | - Boolean  - Integer  - Numeric (number)  - Big number  - Floating integer  - String  - Bytes  - Date  - Date time  - Time  - Timestamp |
| Date | - String  - Date  - Date time  - Timestamp |

## **The CAST function (syntax and examples)**

**CAST** is an American National Standards Institute (ANSI) function used in lots of programming languages, including BigQuery. This section provides the BigQuery syntax and examples of converting the data types in the first column of the previous table. The syntax for the **CAST** function is as follows:

CAST(expression AS typename)

Where **expression** is the data to be converted and **typename** is the data type to be returned.

### **Converting a number to a string**

The following **CAST** statement returns a string from a numeric identified by the variable **MyCount** in the table called **MyTable**.

SELECT CAST(MyCount AS STRING) FROM MyTable

In the above SQL statement, the following occurs:

* **SELECT** indicates that you will be selecting data from a table
* **CAST** indicates that you will be converting the data you select to a different data type
* **AS** comes before and identifies the data type which you are casting to
* **STRING** indicates that you are converting the data to a string
* **FROM** indicates which table you are selecting the data from

### **Converting a string to a number**

The following **CAST** statement returns an integer from a string identified by the variable **MyVarcharCol** in the table called **MyTable**. (An integer is any whole number.)

SELECT CAST(MyVarcharCol AS INT) FROM MyTable

In the above SQL statement, the following occurs:

* **SELECT** indicates that you will be selecting data from a table
* **CAST** indicates that you will be converting the data you select to a different data type
* **AS** comes before and identifies the data type which you are casting to
* **INT** indicates that you are converting the data to an integer
* **FROM** indicates which table you are selecting the data from

### **Converting a date to a string**

The following **CAST** statement returns a string from a date identified by the variable **MyDate** in the table called **MyTable**.

SELECT CAST(MyDate AS STRING) FROM MyTable

In the above SQL statement, the following occurs:

* **SELECT** indicates that you will be selecting data from a table
* **CAST** indicates that you will be converting the data you select to a different data type
* **AS** comes before and identifies the data type which you are casting to
* **STRING** indicates that you are converting the data to a string
* **FROM** indicates which table you are selecting the data from

### **Converting a date to a datetime**

Datetime values have the format of YYYY-MM-DD hh: mm: ss format, so date and time are retained together. The following **CAST** statement returns a datetime value from a date.

SELECT CAST (MyDate AS DATETIME) FROM MyTable

In the above SQL statement, the following occurs:

* **SELECT** indicates that you will be selecting data from a table
* **CAST** indicates that you will be converting the data you select to a different data type
* **AS** comes before and identifies the data type which you are casting to
* **DATETIME** indicates that you are converting the data to a datetime value
* **FROM** indicates which table you are selecting the data from

## **The SAFE\_CAST function**

Using the **CAST** function in a query that fails returns an error in BigQuery. To avoid errors in the event of a failed query, use the **SAFE\_CAST** function instead. The **SAFE\_CAST** function returns a value of Null instead of an error when a query fails.

The syntax for **SAFE\_CAST** is the same as for **CAST**. Simply substitute the function directly in your queries. The following **SAFE\_CAST** statement returns a string from a date.

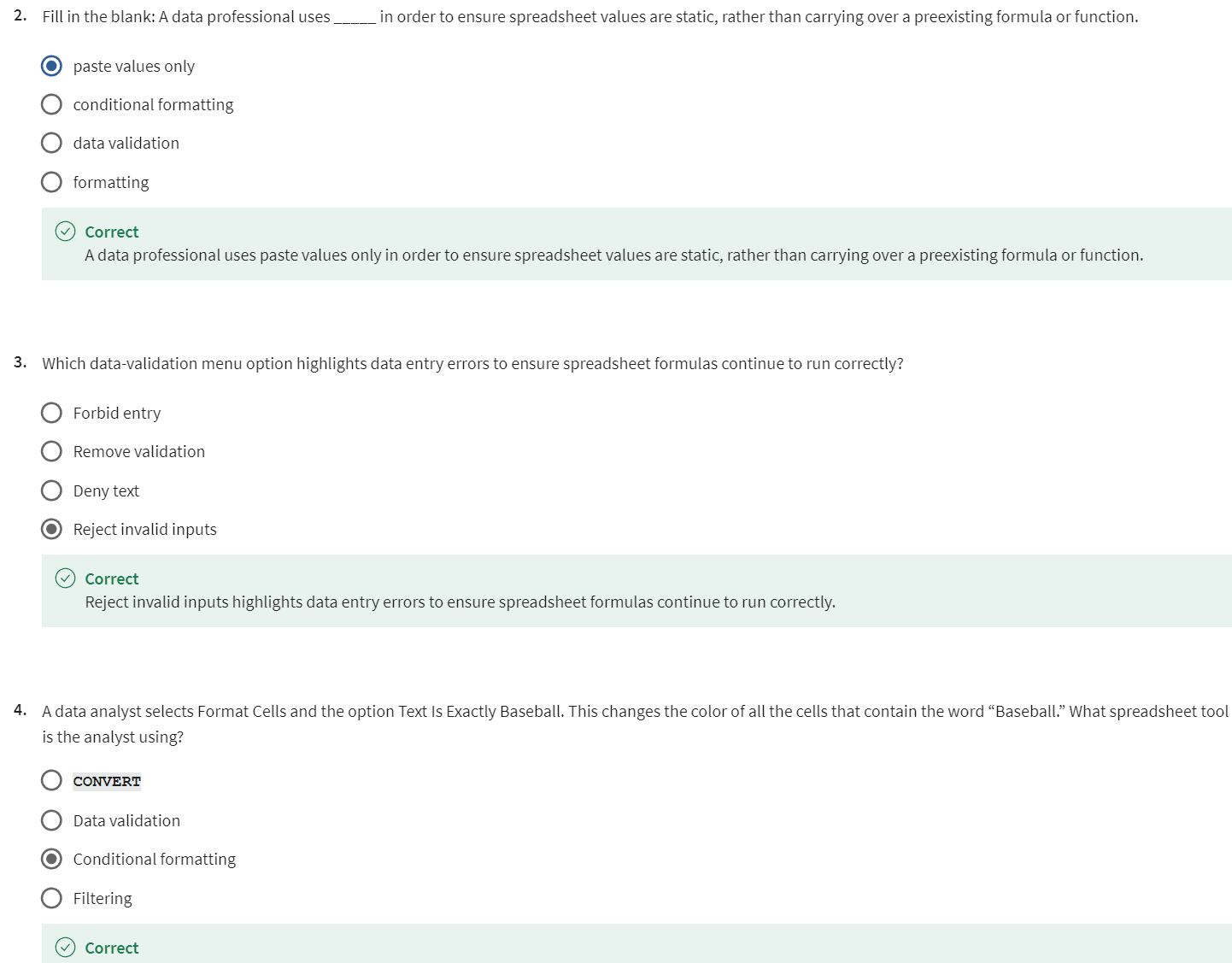
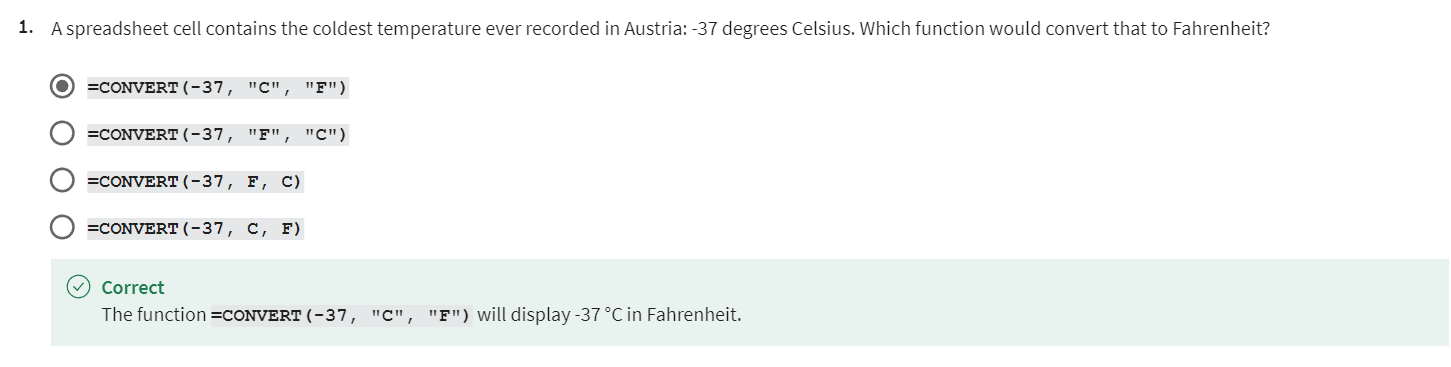
SELECT SAFE\_CAST(MyDate AS STRING) FROM MyTable

**More information**

Browse these resources for more information about data conversion using other SQL dialects (instead of BigQuery):

* [CAST and CONVERT](https://docs.microsoft.com/en-us/sql/t-sql/functions/cast-and-convert-transact-sql?view=sql-server-ver15): SQL Server reference documentation
* [MySQL CAST Functions and Operators](https://dev.mysql.com/doc/refman/8.0/en/cast-functions.html): MySQL reference documentation
* [How to: SQL Type Casting](https://www.blendo.co/blog/how-to-sql-type-casting/): Blog about type casting that has links to other SQL short guides

### [EXERCISE ON CONVERTING & FORMATTING DATA](https://www.coursera.org/learn/analyze-data/quiz/4twnh/test-your-knowledge-on-converting-and-formatting-data)



## 

## COMBINE FOR MULTIPLE DATASETS [IMPORT AND COMBINE DATA IN SPREADSHEETS AND DATABASES](https://www.coursera.org/learn/analyze-data/supplement/QFoFj/import-and-combine-data-in-spreadsheets-and-databases)

In earlier lessons, you discovered how to use the **IMPORTRANGE** and **CONCATENATE** functions in spreadsheets. In this reading, you will have the opportunity to extend your knowledge about these concepts to SQL queries.

## **Import data**

As a data analyst, there are many occasions where you will need to import data from one file or location to another. Both spreadsheets and SQL include functionality that enables you to import data.

### **Import data in spreadsheets**

As you learned earlier, in spreadsheets you use the **IMPORTRANGE** function to import a range of cells from another spreadsheet into your current spreadsheet. The syntax is: **=IMPORTRANGE(spreadsheet\_url, range\_string)**.

In this formula, **spreadsheet\_url** is the URL of the spreadsheet from which you want to import data. The specific cells you want to import, such as A2:B6, are specified by **range\_string**. If the spreadsheet has multiple tabs, you also need to specify the name of the tab as part of the range.

An example of this is a company that needs to track who made retirement contributions so that it can make sure the company match is correctly distributed. The analysts would use **IMPORTRANGE** to pull all retirement contribution information into a spreadsheet that contains all of the employees year-end salaries and bonuses. This enables them to determine which employees made contributions and are eligible for matching funds.

### **Import data in SQL**

**In contrast to spreadsheets, SQL does not include a function for importing data**. Instead, a method you can use to import data from one table to another is to use the **INSERT INTO** command together with a **SELECT** statement. The syntax is:

INSERT INTO [destination\_table\_name]

SELECT [column names, separated by commas, or \* for all columns]

FROM [source\_table\_name]

WHERE [condition]

In this syntax, the SQL query inserts rows from a source table into a destination table based on the **WHERE** clause.

For example, imagine you work for a retail company that stores its sales and customer information in a SQL database. The marketing director asks you to provide them with a table containing the names and addresses of customers who have not made a purchase this year and who live in specific postal codes. One way you could gather this information is to use the **INSERT INTO** along with the **SELECT** and **WHERE** commands, as follows:

INSERT INTO customer\_promotion

SELECT \*

FROM customers

WHERE total\_sales = 0 AND postal\_code = '12345'

## **Combine data**

Another tool in your data analyst toolkit is your ability to join together two or more text strings that are stored in separate columns or fields. For example, you might want to combine a customer’s first and last name to create mailing labels for a marketing promotion. In both spreadsheets and SQL, joining together text strings is referred to as *concatenation*.

### **Combine data in spreadsheets**

In spreadsheets, you use the **CONCATENATE** function to join together two or more text strings, such as combining street addresses and primary contacts in a business’ vendor database.

The basic syntax is **=CONCATENATE(item 1, item 2)**. You can add multiple items by separating them with commas. Where appropriate, such as when you’re combining a customer’s first and last name, you should add a space between the items you’re combining by typing quotation marks space quotation marks [“ ”] between the items. Separate this information by a comma as well. This would change the formula to: **=CONCATENATE(item 1, " ", item 2)**.

### **Combine data in SQL**

In SQL, use the **CONCAT** function to join strings together to create new text strings. You might combine data simply to improve the readability of reports (such as combining a customer’s first and last name when generating a customer list). Or, you might combine data to generate a unique identifier for the rows in a table. Here is the basic syntax:

SELECT CONCAT(field1, " ", field2)

FROM [table\_name]

Notice that this syntax includes " " so that there is a space between the combined fields. With this syntax, SQL combines field1 and field2 with a space between them.

By default, SQL includes the field names as headers when you run a query. However, if you use the **CONCAT** function, SQL doesn’t know what to use as a header. For this reason, you should include an alias for the combined fields to help with readability. You give the combined fields an alias by using **AS**:

FROM [table\_name]

SELECT CONCAT(field1, " ", field2) AS alias

For example, if you plan to use **CONCAT** to combine the first and last names of your company’s customers into a single expression, you could use this query:

SELECT CONCAT(first\_name, " ", last\_name) AS Customer\_Name

FROM [table\_name]

## **Key takeaways**

**Data can be imported and combined in both spreadsheets and SQL databases.**

To import data into a spreadsheet, use the **IMPORTRANGE** function.

To import data into a SQL table, use the **INSERT INTO**, **SELECT**, and **WHERE** commands.

Use **CONCATENATE** to combine two or more data strings in spreadsheets.

In SQL, use the **CONCAT** function to combine fields.

### [STEP-BY-STEP: MERGE TEXT STRINGS TO GAIN INSIGHTS](https://www.coursera.org/learn/analyze-data/supplement/vYBY0/step-by-step-merge-text-strings-to-gain-insights)

This reading outlines the steps the instructor performs in the following video, [Merge text strings to gain insights](https://www.coursera.org/learn/analyze-data/lecture/9V6L5). In the video, the instructor uses SQL’s **CONCAT** function to combine strings from multiple columns to create a new column. Additionally, the instructor uses other SQL commands such as **AVG**, **GROUP BY**, and **ORDER BY** to gain insights about the new column.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

## 

## **What you’ll need**

If you would like to follow along with the instructor, you will need to log in to your BigQuery account to use the open (public) dataset called **new\_york\_citibike**. The table you will use is called **citibike\_trips**. You previously used a BigQuery public dataset in the activity, [Hands-On Activity: Analyze weather data in BigQuery](https://www.coursera.org/learn/analyze-data/quiz/yRIIz/hands-on-activity-analyze-weather-data-in-bigquery). Review that activity if you need a refresher on loading public datasets!

empty alt text

## **Example: Use CONCAT on the bike sharing dataset**

The **CONCAT** function can combine data from separate columns to provide new insights.

* In the BigQuery editor, enter **SELECT** and press Enter (Windows) or Return (Mac).
* Enter **usertype,** on line 2.
* On line 3, enter **CONCAT(start\_station\_name," to ", end\_station\_name)** to combine the names of the beginning and ending stations for each trip in a new column. This will create one column of routes.
* Enter **AS route,** at the end of line 3 to name the column route.
* On line 4, enter **COUNT (\*) as num\_trips,** to count the number of trips. The asterisk tells SQL to count the number of rows you’re selecting. Each row represents a trip, so you can count all of the rows you’ve selected to count the number of trips.

Next, calculate the average trip duration for each route. On line 5, enter:

* **ROUND(AVG(cast(tripduration as int64)/60),2) AS duration**

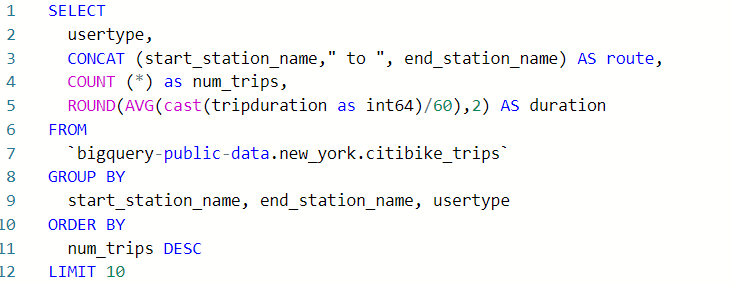
This line of code accomplishes several tasks:

* It uses the **CAST** function to cast **tripduration** as an integer and divides that number by 60 to convert the number from seconds to minutes.
* It uses the **AVG** function to find the average duration of each route.
* It uses the **ROUND** function to round the output to 2 decimal places.
* It uses the **AS** command to give this output the alias **duration**.

**Note 1:** BigQuery stores numbers in a 64-bit memory system, which is why there's a 64 after integer in this case.

**Note 2:** While explaining this code, the instructor says "divide by the number of rows." Instead, they meant "divide by 60."

* Enter **FROM** on line 6 and press return.
* Enter **`bigquery-public-data.new\_york.citibike\_trips`** on line 7 (enclosed in back-ticks).
* Enter **GROUP BY** on line 8.
* Enter **start\_station\_name, end\_station\_name, usertype** on line 9.
* Enter **ORDER BY** on line 10 to tell SQL how to organize this data.
* Enter **num\_trips DESC** on line 11 to sort it in descending order.
* Enter **LIMIT 10** on line 12.
* Your completed query should match the following code:



* Select **RUN** to view the results.

Now you can easily read these route names and trace them back to real places. You can also explore the types of customers taking each route. This type of information can help decision-makers at the bike-sharing company understand their user base in different parts of the city and where to keep more bikes for people to rent.

[MERGE TEXT STRINGS TO GAIN INSIGHTS](https://www.coursera.org/learn/analyze-data/lecture/9V6L5/merge-text-strings-to-gain-insights)

### 

### [HANDS-ON ACTIVITY: COMBINE MULTIPLE PIECES OF DATA](https://www.coursera.org/learn/analyze-data/quiz/OWUBG/hands-on-activity-combine-multiple-pieces-of-data)



## **Activity Overview**

****

In previous activities, you gained experience using spreadsheet functions for manipulating and cleaning data. In this activity, you’ll use the **CONCAT** and **CONCATENATE** functions to help you quickly and efficiently combine multiple pieces of raw data into new data.

By the time you complete this activity, you will be able to use these functions to combine data. This will enable you to simplify and condense data, which is important for processing and cleaning data in your career as a data analyst.

### Step-By-Step Instructions

Follow the instructions to complete each step of the activity. Then answer the questions at the end of the activity before going to the next course item.

### Step 1: Access the template

To get started, you will need the **CONCAT** function exercise spreadsheet.

To use the template for the spreadsheet, click the link below and select “Use Template.”

Link to template: [CONCAT Function Exercise Spreadsheet](https://docs.google.com/spreadsheets/d/1Na9M4xb-3CA746BHECX75n3B9V04pxLo0xNs9i8jMNo/template/preview)

OR

If you don’t have a Google account, you can download the spreadsheet directly from the attachment below.

[Dataset for Project\_ CONCAT function](https://d3c33hcgiwev3.cloudfront.net/s7PLxn9URlSzy8Z_VDZUqQ_90e1fe6271aa400cb7a05c5ba5c40a56_Dataset-for-Project_-CONCAT-function.xlsx?Expires=1712793600&Signature=XiR2V9fckMLTArSumSX55d5OY5ua0~YAM7WXW9rSorR5bFithLFuJxr8Canszhd8sAmhVvA8hjkdEwLmLhqZVm8jgPiHGtmFKnXolxjUI2VKATAkstwATg~Gnpc13qQlMYdjJC33P2W3EARO42OYUzyUEdfumbNuLXRoc5LK~qk_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

### Step 2: Why Use the CONCAT and CONCATENATE functions

Occasionally, you will encounter a dataset with data values in separate cells that you want to combine as a single value in a single cell. This is common when dealing with names and dates. The dataset may have separate columns for first names and last names, but you may want a column with the full names.

City / state and month / year combinations are also often desirable to have together, as they are likely to be recorded together.

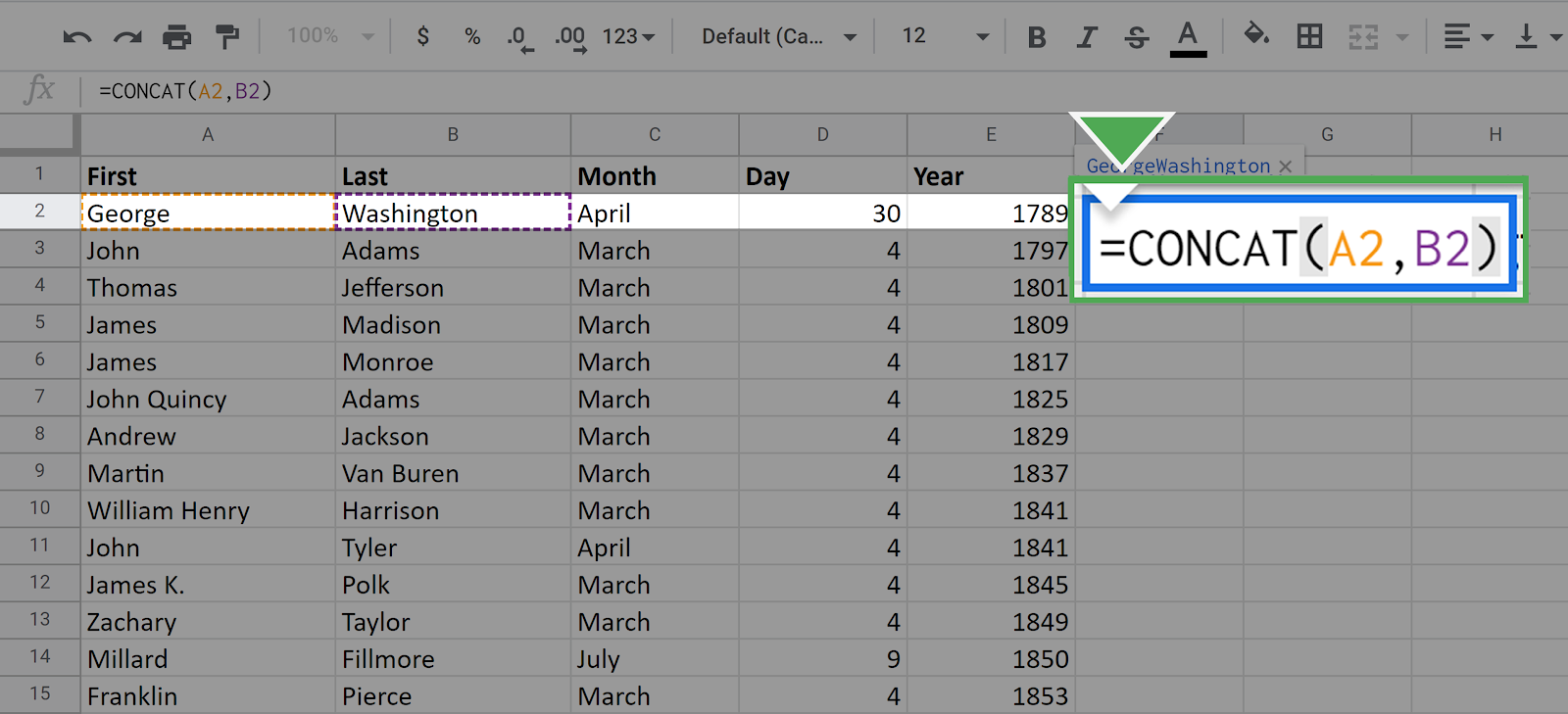
The **CONCAT** function in spreadsheets can combine these kinds of data.

### Step 3: Combine data from two cells

First, using the spreadsheet you downloaded, you’ll combine the two sets of names in columns First Name and Last Name in a new column called Full Name.

To do this, follow these steps:

1. Click on cell F2. This is where you start the data for the new column. After you click on the cell, enter **=CONCAT(A2,B2)** into the function bar and hit Enter (Windows) or Return (Mac).



Once you press enter, the following data should appear in the cell:

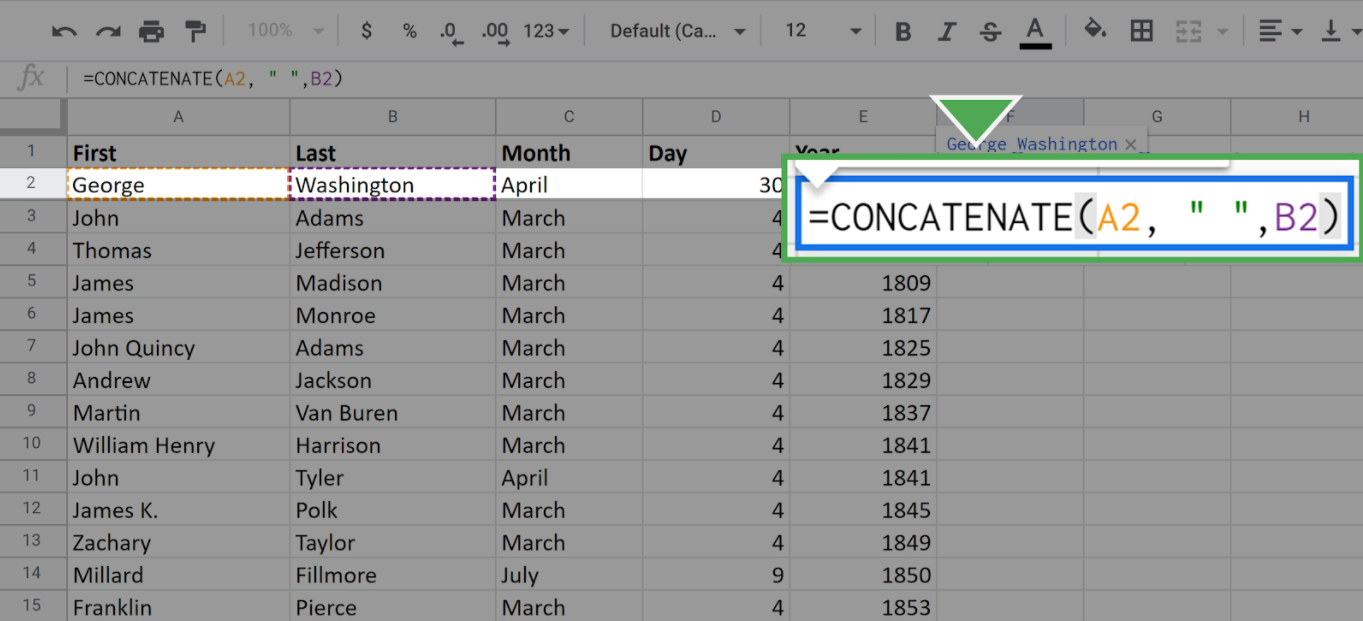


You have merged or, technically, concatenated the two data values from cells A2 and B2. Because you listed A2 first in the **CONCAT** function argument, it comes first in the final result.

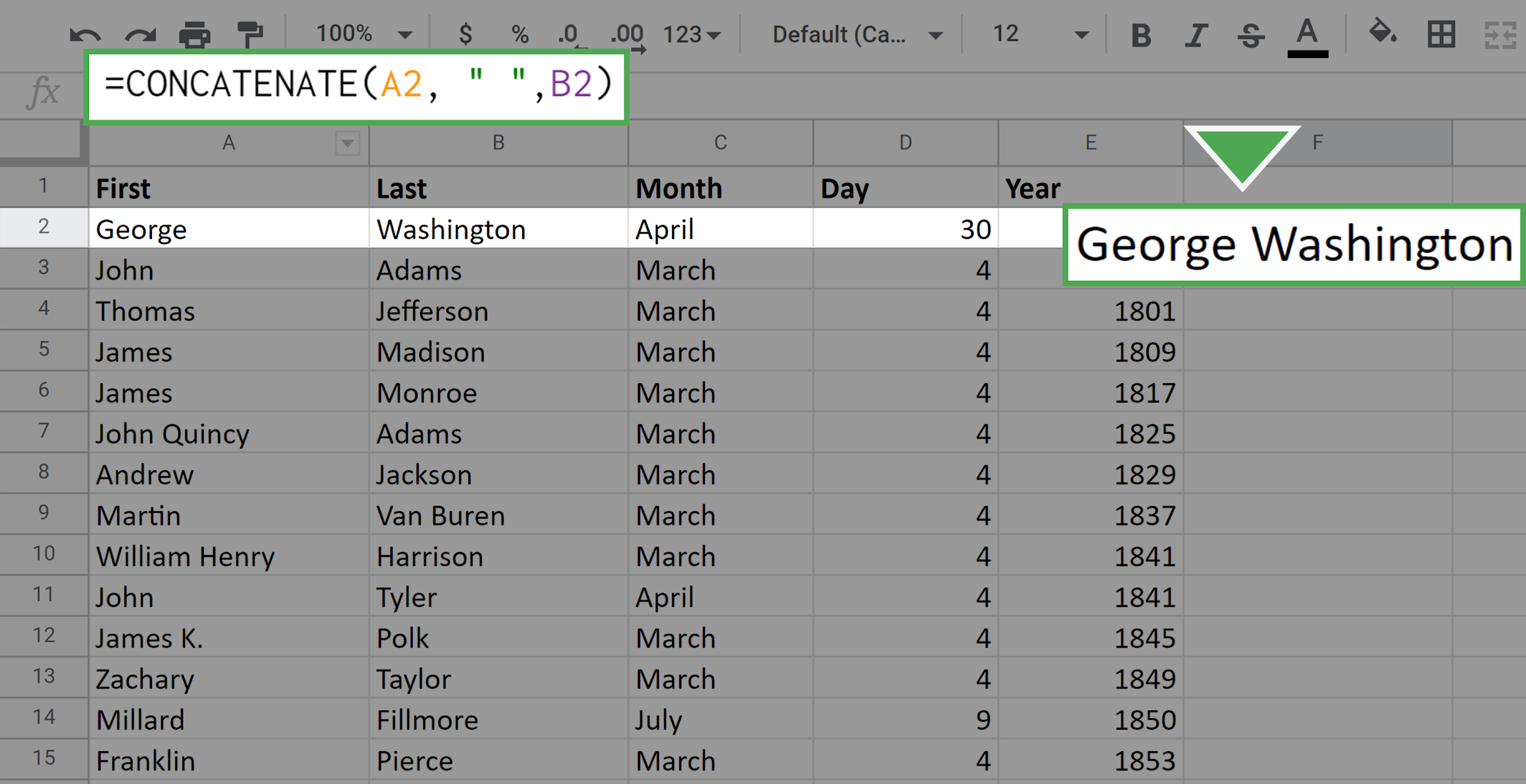
Notice that the two names were combined without a space between them.

If you want to put the space in between, you need to use the full **CONCATENATE** function, which allows you to combine multiple strings.

2. Click again on the cell F2. In the function call, place a space in quotes between A2 and B2 separated by commas.



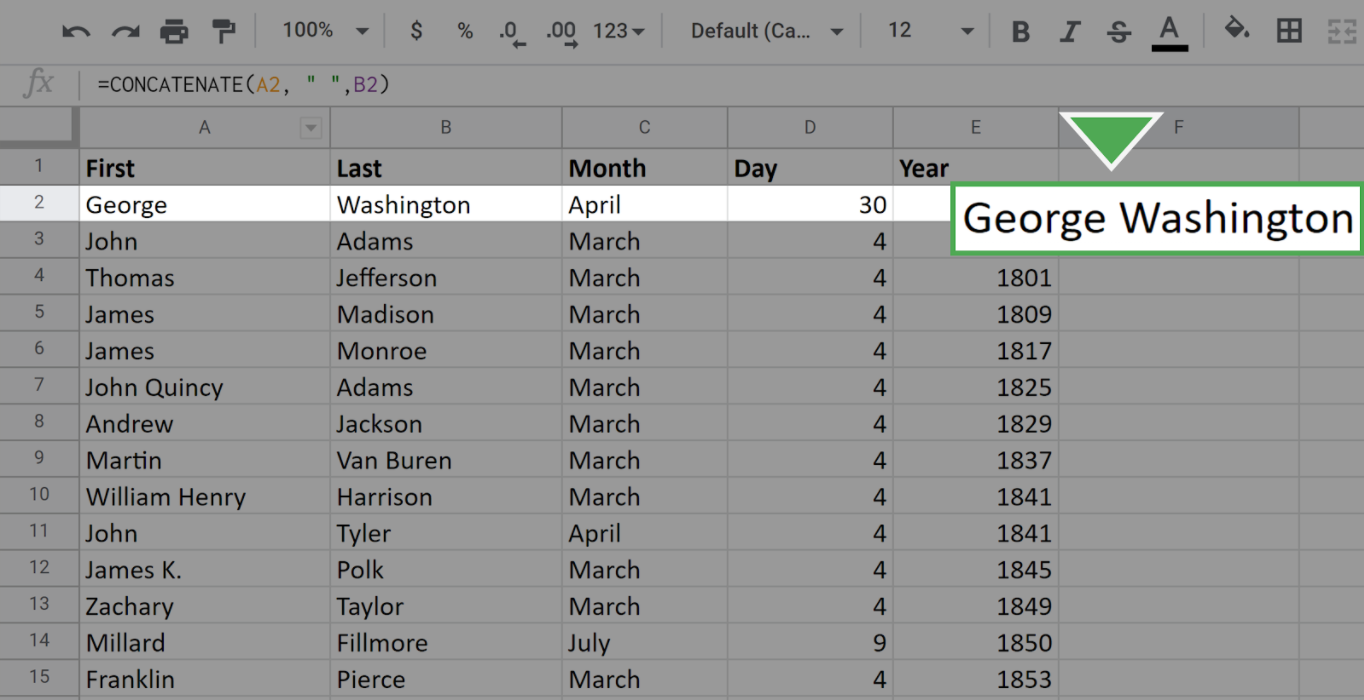
Once you press enter or return, your screen should appear like this:



Now there is a space between the first name and the last name.

3. Next, repeat this process for all the remaining cells in Column F. Of course, you don't want to do this manually for each cell. (Especially if the dataset were larger, it would be laborious to do this cell-by-cell.) Luckily, you can fill out the data in the column by using your mouse.

4. Click on the cell F2. Locate the small square in the lower-right corner of the highlighted boundary of the cell.



5. Click on this square, drag your mouse to the bottom of the column, and release. All the cells in the column should populate with the full name of the appropriate president.



Note: While it does not happen in this dataset, you may have extra spaces in your result after you **CONCAT**. If you notice you have extra spaces, you can use the **TRIM** function to remove them.

### 

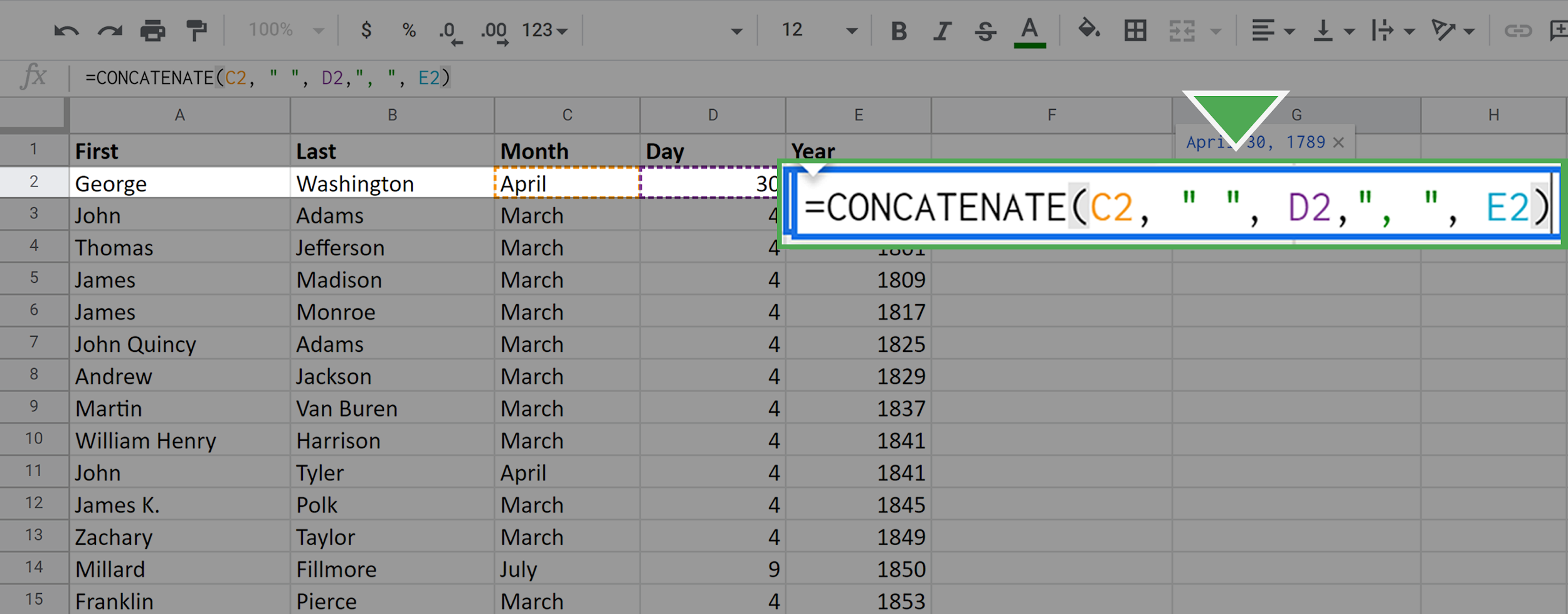
### Step 4: Combine data from three cells

The procedure for combining three pieces of data from different cells is almost identical to what you just did. The only difference is that you include a third cell in the full **CONCATENATE** argument.

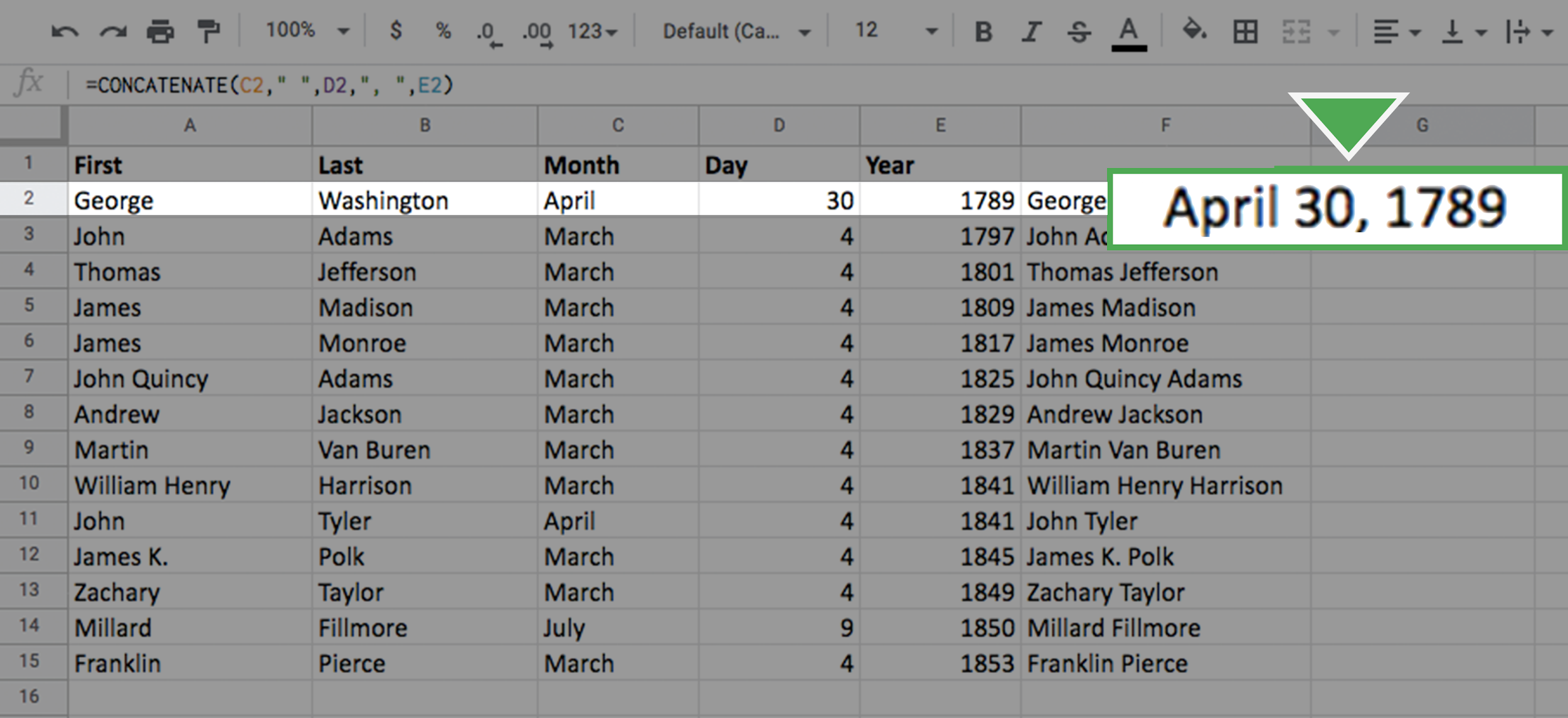
Now, combine the month, day, and year into a single data value: Date. This will occupy column G.

1. Click on the cell where you would like the new data to start. Here, this is cell G2.

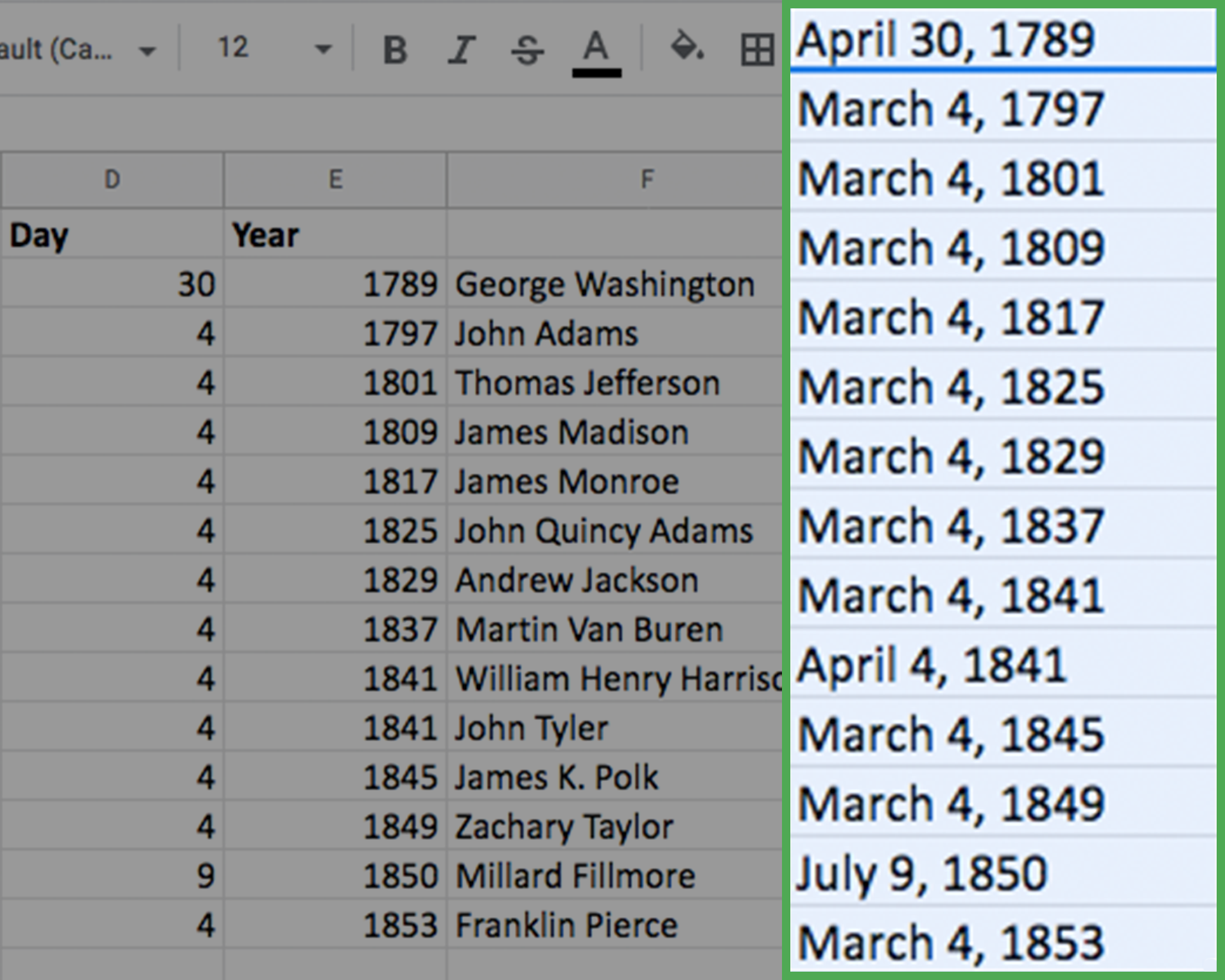
2. Enter the CONCAT command as **=CONCATENATE(C2," ",D2,", ",E2)**

****

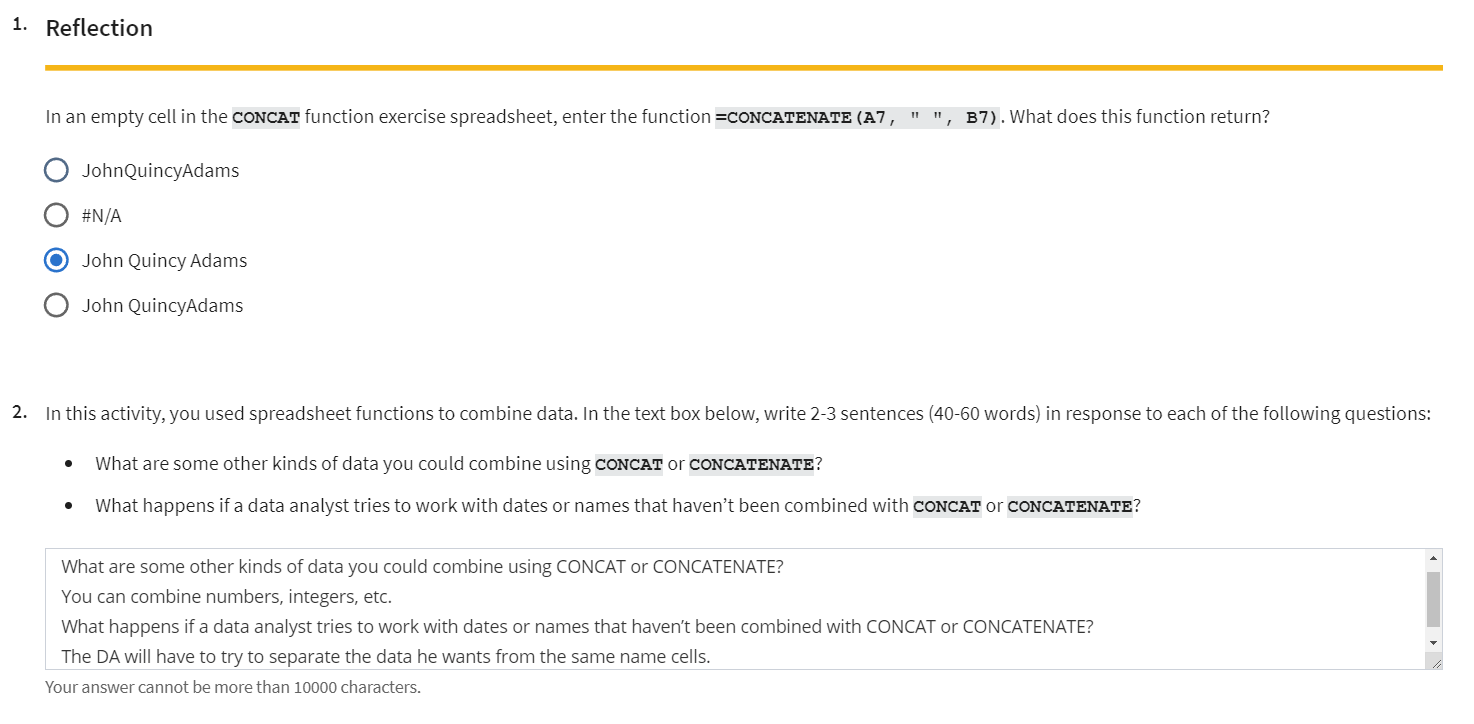
Pay particular attention to the extra strings you added between the month and the day, and between the day and the year. This is how you get the spaces and comma in your final result.



3. Fill out the rest of the column using the same click-and-drag technique as before. Your screen should appear like this:



Congratulations! You’ve combined data in spreadsheets using the **CONCAT** and **CONCATENATE** functions.



### [STEP-BY-STEP: STRINGS IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/supplement/2Ulho/step-by-step-strings-in-spreadsheets)

This reading outlines the steps the instructor performs in the next video, [Strings in spreadsheets](https://www.coursera.org/learn/analyze-data/lecture/syKyK/strings-in-spreadsheets). In this video, the instructor demonstrates the **LEN**, **LEFT**, **RIGHT**, and **FIND** functions and discusses how you can use them to better understand your data.

Keep this step-by-step guide open as you watch the video. It can serve as a helpful reference tool if you need additional context or clarification while following the video steps. This is not a graded activity, but you can complete these steps to practice the skills demonstrated in the video.

**What you’ll need**

If you’d like to access the spreadsheet the instructor uses in this video, click the link to the dataset to create a copy. If you don’t have a Google account, you may download the data directly from the attachments below. Note that this is a larger database so it may take a moment or two to load.

[Citi Bike Trip Data](https://d3c33hcgiwev3.cloudfront.net/-eSTMJ-bSD2WQNV_8mQn2A_f258cf38d2ad4f90baf1ad10e7bf3be1_Citi-Bike-Trip-Data.xlsx?Expires=1712793600&Signature=OPzQJRPSl23N4wOgzMwiM5S9f4MY1vLbg6v9otATVCgNIiJmZfZRroDDdaA8Y9MZ7hjQUEHFg8pCfT~nThJ84vE78eHdB7XaGuG52sppRQwlfhlawu87qfoUqI7QSX5-CTkaj9NVzIn6ZcZjafvq6WvGRL6UEpGhdm1dZuxZflo_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

**Note:** If the directions in the video do not work for the version of Excel you have, visit the free online training center [Microsoft Excel for Windows Training](https://support.microsoft.com/en-us/office/excel-video-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb), and search for these functions to learn how to use them in Excel.

empty alt text

**Example 1: The LEN function**

The **LEN** function calculates a string’s length. Use this formula to check the length of the datetime strings in column C.

1. Open the [Citi Bike Trip Data](https://docs.google.com/spreadsheets/d/1ZetYs2n7csR1pI92hOuMfV_RXmxbOh1UOOxx1KdDv0E/template/preview?usp=sharing&resourcekey=0-qpI7g9md35n648x5jgNwaQ) spreadsheet.
2. In cell **B2**, enter the equals sign [**=**] to begin the function.
3. Enter **LEN**, followed by an open parenthesis [**(**].
4. Select cell **C2**. Then add a close parenthesis [**)**].
5. Press **Enter**. The result, 19, indicates the string in cell **C2** is 19 characters.

**Example 2: The FIND function**

The **FIND** function locates specific characters and substrings in a string. All of the start-time (column C) and stop-time (column D) strings in the spreadsheet have a space between the date and the time. Use the **FIND** function to determine where in the string this space is located.

1. Select cell **B3** and enter the equals sign[**=**].
2. Enter **FIND** followed by an open parenthesis [**(**].
3. Enter quotation mark, then space, then close quotation mark [**" "**] to specify that you want to find a space.
4. Select cell **C3** and add a close parenthesis [**)**].
5. Press **Enter** to run the formula. This function returns 11, indicating that the space is the 11th character in the string.

**Note: FIND** is case sensitive, so always make sure you input the substring correctly.

**Example 3: The RIGHT function**

Use the **RIGHT** function to select a specific number of characters on the right side of a cell. Here, you want to return the substring that represents time, which is contained within the eight characters to the right of the space.

1. Reopen the spreadsheet so you are working with an unaltered version of the document [Citi Bike Trip Data](https://docs.google.com/spreadsheets/d/1ZetYs2n7csR1pI92hOuMfV_RXmxbOh1UOOxx1KdDv0E/template/preview?usp=sharing&resourcekey=0-qpI7g9md35n648x5jgNwaQ).
2. In cell **B2**, enter the equals sign [**=**].
3. Enter **RIGHT** followed by an open parenthesis [**(**].
4. Select cell **C2** then enter a comma [**,**].
5. Enter **8** to specify that you want the function to return the eight rightmost characters in the string.
6. Add a close parenthesis [**)**] to complete the formula.
7. Press **Enter** to run the formula. The time stamp from the start-time data string has now been isolated in cell **B2**.
8. Double-click the fill handle in cell **B2** to fill the rest of the column.
9. Enter **Time** in cell **B1** to add a column header.

**Example 4: The LEFT function**

Use the **LEFT** function to select a specific number of characters on the left side of a cell. Here, you want to return the date substring, which is the 10 characters to the left of the space. These characters represent the start date.

1. Right-click cell B.
2. Select **Insert 1 column left** to create a new column **B** for the date substring.
3. Enter **Date** in cell **B1** to add a column header.
4. Enter the equals sign [**=**] in cell **B2**.
5. Enter **LEFT** followed by an open parenthesis [**(**].
6. Select cell **D2** then enter a comma [**,**]. **Note:** You added a column, so the start time column is now column **D**.
7. **Add a comma** [**,**] after **D2** in the formula.
8. Enter **10**, to indicate that you want to return the first 10 characters in the date string.
9. Add a close parenthesis [**)**] to complete the formula. **Note:** The instructor enters 11, which will return the date substring and the space.
10. Press **Enter**. The date from the start-time data string has now been isolated in the new cell **B2**.
11. Double-click the fill handle in cell **B2** to fill the rest of the column.

### [STRINGS IN SPREADSHEETS](https://www.coursera.org/learn/analyze-data/lecture/syKyK/strings-in-spreadsheets)

Previous topic video.

### [MANIPULATE STRINGS WITH SQL](https://www.coursera.org/learn/analyze-data/supplement/oitMs/manipulate-strings-with-sql)

# Manipulate strings with SQL

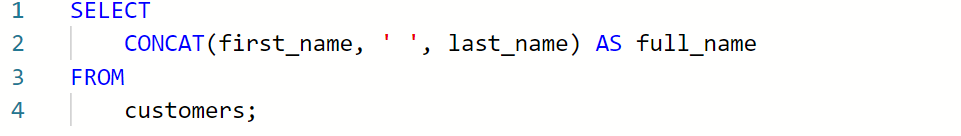
An important part of a data analyst’s job is knowing how to convert and manipulate data for analysis. One way data analysts manipulate strings is to concatenate them, which means to join together two or more text strings. Once strings are concatenated, they form a new, longer text string for analysis. In this reading, you'll learn about different SQL functions that can be used to concatenate strings.

## **CONCAT in action**

Here are some examples of how you might use **CONCAT** as you work with data.

### **CONCAT**

You’re working with the marketing team on an email campaign, and you need to generate full names from your database’s first and last name columns. SQL's **CONCAT** function allows you to join together two or more string values, simplifying this task, as follows:



In this example, **CONCAT** merges the **first\_name** and **last\_name** fields to create a new field called **full\_name**. The space (**' '**) separator ensures the full name appears properly.

### **CONCAT\_WS**

Now, you're tasked with creating a report that includes a website's URL components: the protocol (http), domain name (**your\_company**), and domain (**com**). You'd use **CONCAT\_WS**, which stands for **CONCAT With Separator**, to achieve this. It's similar to **CONCAT**, but it includes a separator, such as a space or period, between the strings.

SELECT CONCAT\_WS('.', 'www', 'your\_company', 'com') as website FROM web\_data;

Here, **CONCAT\_WS** adds a period (**'.'**) between each part of the website URL, ensuring the URL is in the correct, navigable format.

### **CONCAT with ||**

In BigQuery, you can use the **||** operator to concatenate strings. For instance, if you're working with a dataset of book information and want to create a full title by combining the book's name and its edition, you could use **||**, like so:

SELECT book\_name || ' - ' || edition AS full\_book\_title FROM library;

This script combines the book name and edition, separated by a hyphen for clarity, providing a complete, informative title for your records.

**Note:** In some other SQL environments, you cannot use the **||** operator to concatenate strings. You must use **+** instead. For example, to concatenate the strings **'Google'** and **'.com'** in Microsoft SQL server, you would use:

SELECT 'Google' + '.com'

Always ensure you're using the correct syntax for the specific SQL environment you're working in!

## **Concatenate strings with SQL**

Review the table below as a summary of the **CONCAT** function and its variations in SQL.

| **Function/ operator** | **Use** | **Example** | **Result** |
| --- | --- | --- | --- |
| **CONCAT** | Concatenate strings to create new text strings | **CONCAT('Google', '.com')** | **Google.com** |
| **CONCAT\_WS** | Concatenate two or more strings together with a separator between each string | **CONCAT\_WS(' . ', 'www', 'google', 'com')** | **www.google.com** |
| **||** | Concatenate two or more strings together with the **||** operator | **'Google' || '.com'** | **Google.com** |

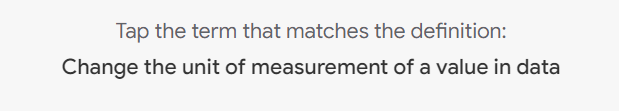
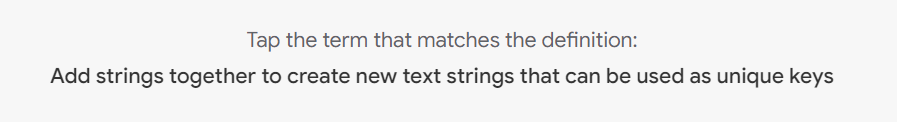
## 

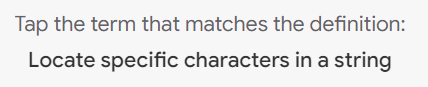
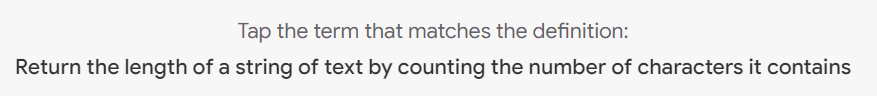
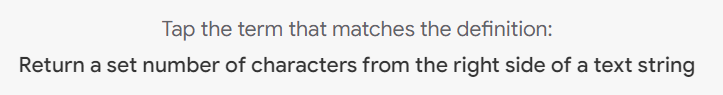
## **Key takeaways**

In SQL, **CONCAT** is a function that joins strings together to create new text strings. This is useful for creating new variables or features for data analysis, as well as more readable and informative output. In this way, **CONCAT** can simplify your data analysis and make you more efficient.

### 

### [SQL QUERY SYNTAX](https://www.coursera.org/learn/analyze-data/ungradedWidget/e1EZD/sql-query-syntax)





### [TEST YOUR KNOWLEDGE ON COMBINING MULTIPLE DATASETS](https://www.coursera.org/learn/analyze-data/quiz/btTha/test-your-knowledge-on-combining-multiple-datasets)

### 

## GET SUPPORT DURING ANALYSIS

### [WHEN YOU GET STUCK](https://www.coursera.org/learn/analyze-data/lecture/Vqiu7/when-you-get-stuck)

Data analysts spend a lot of time problem-solving, and that means **there's going to be times when you get stuck**, but the trick is knowing what to do when that happens. In this video, we'll talk about the importance of knowing how to get help, whether that means asking someone else for help or searching the internet for answers.

Asking other people about a problem you're having can help you find new solutions that move a project forward. **It's always a good idea to reach out to your peers and mentors, especially if they're working with you on that project**. Your team members have valuable knowledge and insight that can help you find the solution you need to get unstuck. Sometimes we spend a lot of time spinning our wheels saying, "I can do this myself," but we can be way more productive if we engage with other people, find new resources to lean on and try to get as many voices as we can.

For example, let's say you're working with the bike trip time data from the previous videos. Maybe you're trying to find the average time between bike rides in a given month. Calculating the difference between bike rides before midnight is easy, but you can run into a problem if the elapsed time crosses into the next day. If someone went on a bike ride at 11:00 PM, but the next ride wasn't until 06:00 AM, your formula would return a negative number because the end time is less than the start time. You know that you can add one minus the start time if two bike rides start and end on different days, but that formula won't work on times that happened in the same day, and it's pretty inefficient to scroll through every bike ride to pinpoint these special cases.

You need to find a way to build a conditional formula, but you aren't sure how.

You decide to check in with other analysts working on your team to see if they have any ideas. You could send them a quick email, or stop by their desk, to find out if they have a minute to talk it over with you. Turns out they had a similar problem on a previous project, and they're able to show you a conditional formula that you could use to speed up your calculations. Great! They suggest using an IF formula like this. This basically says that, "if the end time is larger than the start time, replace the standard end time minus start time formula with one minus start time plus end time." Now it's also possible that your team members don't have an answer; that's okay too.

There's definitely someone else with the same problem asking the same questions online. **Knowing how to find solutions online is an incredibly valuable problem-solving tool for data analysis**. There's also all kinds of forums where spreadsheet users can ask questions, and you never know what you can turn up with just a basic search.

For example, let's say you look at "calculate number of hours between times" spreadsheets and find a helpful walk-through for a more complicated formula using MOD. This flips the negative values into positive ones, solving your calculation problem. Whether you're asking someone you know or searching the internet for answers, reaching out for help can give you some really interesting solutions and new ways to solve problems for future analysis.

### 

### 

### [ADVANCED SPREADSHEET TIPS AND TRICKS](https://www.coursera.org/learn/analyze-data/supplement/6Vp5U/advanced-spreadsheet-tips-and-tricks)

Like a lot of the things you’re learning in this program, spreadsheets will get easier the more you practice. This reading provides you with a list of resources that may help advance your knowledge and experience with spreadsheet functions and functionality. The goal is to provide you with access to a variety of advanced tips and tricks that will help make you more efficient and effective when working with spreadsheets to perform data analysis. Review the description of each resource below, click the links to learn more, and save or bookmark any links that are useful to you. You can immediately start practicing anything that you learn to increase the chances of your understanding and to build your familiarity with spreadsheets. This reading provides a range of resources, so feel free to explore the ones that are applicable to you and skip the ones that aren’t.

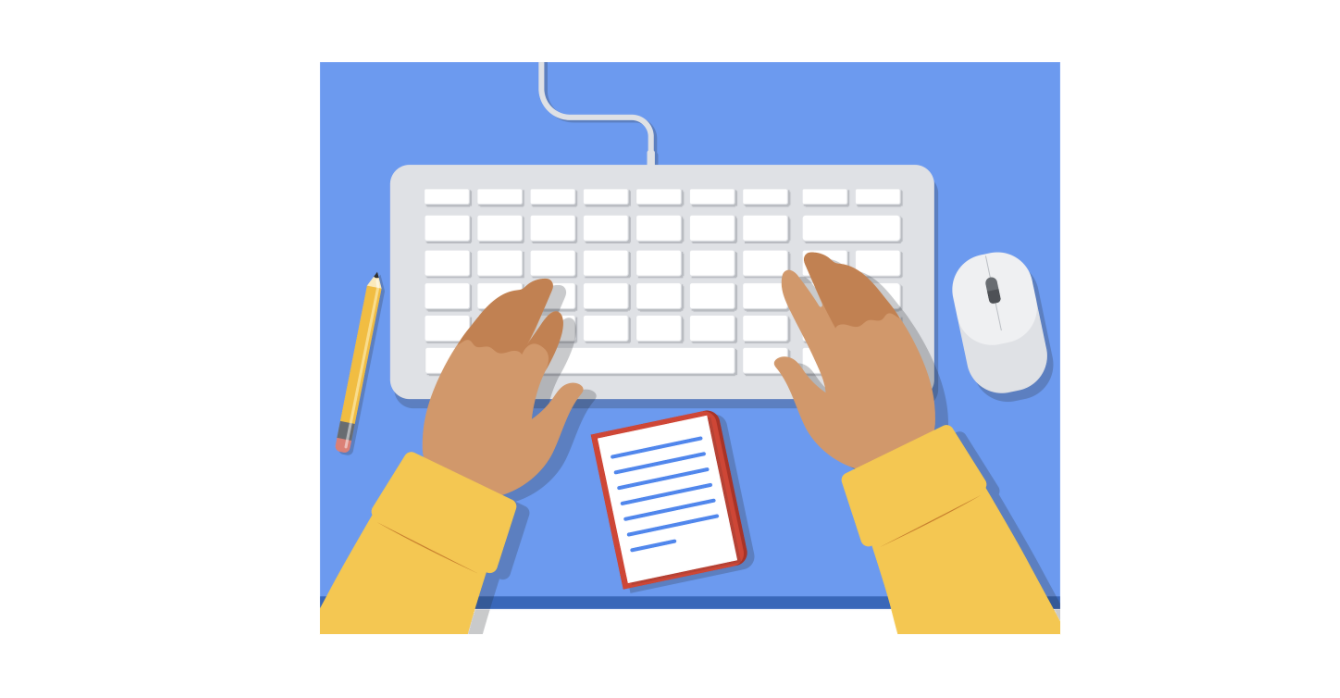
### **Google Sheets**

* [**Keyboard shortcuts for Google Sheets:**](https://support.google.com/docs/answer/181110)This is a great resource for quickly learning a range of keyboard shortcuts that can make regular tasks quicker and easier, like navigating your spreadsheet or accessing formulas and functions. This list contains shortcuts for the desktop and mobile versions of Google Sheets so that you can apply them to your work no matter what device you are using.
* [**List of Google Sheets Functions**](https://support.google.com/docs/table/25273?hl=en)**:** This is a comprehensive list of the Google Sheets functions and syntax. Each function is listed with a link to learn more.
* [**23 Must-Know Google Sheet Formulas**](https://blog.golayer.io/google-sheets/google-sheets-formulas)**:** This blog article from *Layer* summarizes and describes 20 of the most useful Google Sheets formulas.
* [**18 Google Sheets Formula Tips and Techniques:**](https://www.benlcollins.com/spreadsheets/google-sheets-formulas-techniques/) These are tips for using Google Sheets shortcuts when working with formulas.

### **Excel**

* [**Keyboard shortcuts in Excel:**](https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f?ui=en-US&rs=en-US&ad=US)Earlier in this list, you were provided with a resource for keyboard shortcuts in Google Sheets. Similarly, this resource provides a list of keyboard shortcuts in Excel that will make performing regular spreadsheet tasks more efficient. This includes keyboard shortcuts for both desktop and mobile versions of Excel, so you can apply them no matter what platform you are working on.
* [**222 Excel shortcuts:**](https://exceljet.net/keyboard-shortcuts)A compilation of shortcuts includes links to more detailed explanations about how to use them. This is a great way to quickly reference keyboard shortcuts. The list has been organized by functionality, so you can go directly to the sections that are most useful to you.
* [**List of spreadsheet functions:**](https://exceljet.net/excel-functions)This is a comprehensive list of Excel spreadsheet functions with links to more detailed explanations. This is a useful resource to save so that you can reference it often; that way, you’ll have access to functions and examples that you can apply to your work.
* [**List of spreadsheet formulas:**](https://exceljet.net/formulas)Similar to the previous resource, this comprehensive list of Excel spreadsheet formulas with links to more detailed explanations and can be saved and referenced any time you need to check out a formula for your analysis.
* [**Essential Excel Skills for Analyzing Data:**](https://learntocodewith.me/posts/excel-skills/)This blog post includes more advanced functionalities of some spreadsheet tools that you have previously learned about, like pivot tables and conditional formatting. These skills have been identified as particularly useful for data analysis. Each section includes a how-to video that will take you through the process of using these functions step-by-step, so that you can apply them to your own analysis.
* [**Advanced Spreadsheet Skills:**](https://www.slideshare.net/markjhonoxillo/advanced-spreadsheet-skills)Mark John C. Oxillo’s presentation starts with a basic overview of spreadsheet but also includes advanced functions and exercises to help you apply formulas to actual data in Excel. This is a great way to review some basic concepts and practice the skills you have been learning so far.

There are lots of resources online about advanced spreadsheet tips and tricks. You'll probably discover new resources and tools on your own, but this list is a great starting point as you become more familiar with spreadsheets.



### [LAYLA: ALL ABOUT THE ANALYZE STAGE](https://www.coursera.org/learn/analyze-data/lecture/eA2sw/layla-all-about-the-analyze-stage)

Hi. My name is Layla, and I'm an analytical lead at Google. An analytical lead is someone who helps advertisers understand the value of their advertising dollars. We also help them understand if they were to spend another dollar on ads, where they should spend that dollar, and what they could expect to get from it.

The skill set involved in this kind of role has all to do with being able to look at a dataset and make sense of it, and then tell a story to people who maybe don't have that same experience level with data.

**What's going on in the data? What's driving growth for your client or your company? What could they do more or less to drive more of what they want to happen?**

*The Analyze stage is like preparing a fabulous meal. You have done all the cleaning and the preparing and the cooking, and you're finally able to take a bite and to see if what you're originally hoping to happen or what you were expecting, to see if that is really the case.* Is it delicious? Is it exactly like you expected? Or is the consistency a little off and you need to add a little bit more salt?

**The Analysis stage begins once you've prepped and cleaned your data**. You don't want to have those blank fields that will throw you off or duplicate entries that will enlarge your dataset beyond what's actually true. The Analyze stage is where you become the expert about your dataset. Here, you're going to understand all of the different fields. You're going to understand their averages, potentially the median of the data. You're going to understand how different rows in your data differ from each other.

And it's where you're going to gain the confidence to be able to explain your findings to an audience that maybe does not have the same level of expertise with data that you have.

When I analyze data, I often like to use SQL and spreadsheets. You can use these tools to, for example, sort your data and understand which entries are larger than others.

Or to understand how many times something happens by selecting the distinct entries.

Here, you can also filter out data that you are specifically interested in analyzing, or in a spreadsheet, use conditional formatting to show which entries show a more positive outcome and which ones are maybe more negative.

Using SQL and spreadsheets to help you through the analyze phase are absolutely crucial. With these tools, you can format your dataset in a way that is digestible and then begin to tell a story with the data.

My favorite part of working in a spreadsheet is when you finally have that data that you want cleaned up and exported from your SQL query. Then you get to turn it into a pivot table and chart out exactly the cut of data that you were interested in looking at in the first place and explore the trends that are happening there. When you get to do that, you basically unlock a whole world of information and you get to pick what story that you actually want to tell with your data without just saying, "This is the largest number; that's the smallest number." Here, you have to show what is happening over time potentially or what you should expect to see in the future.

[RUNNING INTO CHALLENGES? NOT TO WORRY!  
  
Earlier, we talked a little bit about finding resources online to help you figure out solutions to problems during analysis.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[The internet has so much knowledge and advice to offer, but you need to know how to find it. In this video, we'll talk more about finding answers online.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**You might think that great data analysts don't regularly rely on outside resources, but that's a myth.**](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[The best data analysts know that finding answers to their problems online can be empowering and give them new knowledge for the future. Being able to find new ideas and combine them with what you already know can help you come up with some amazing things.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**Don't be afraid to turn to the internet to find your answers.** It's a great resource that lots of analysts use. Me included.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[But let's talk more about how you can make sure you're using web resources in the best way possible.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**There's a combination of best practices that you can use to guide your search for answers online**. By practicing the thinking skills we've learned in this program, using the right data analytics terms, and your basic knowledge of analysis tools, you have everything you need to find answers and apply them to your own work.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[And it starts with how you approach a problem mentally.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[You've learned about different kinds of thinking skills and how to practice them in your data analysis work.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**From analytical, to mathematical, to structured thinking**. This helps build your mental model, or your thought process, and the way you approach a problem. **Data analysts use these thinking skills to approach a problem logically and break it into smaller parts**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[Building this into your own problem-solving process can help you pinpoint specific questions, which you can use to find resources more easily.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[For example, maybe you keep running into an error in your analysis. You narrow it down to two possibilities: your formula or the data itself. You double check your formula, and you see that it's correct. So now you know that you need to make sure that data has been entered correctly. You consider the problem logically and track it down to the source, using your mental model.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[Next, it's important to use the right terms when searching for solutions. **Knowing how to frame data analytics questions with the same language other analysts are using will help you get more search results, and it'll help you understand what other analysts are saying**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[For example, maybe you need to use the left four characters of a string for a column in SQL. How would you search for this? Searching for "four characters in a column" is a little vague and might not bring up specific resources. But "left string query SQL" **uses some keywords that other data analysts are also using to talk about these things**. **On top of being able to use the right terms to search online, you also need to be familiar with basic tools**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[That way, when an online resource is walking you through a new function and a tool that you've used before, you'll know how those tools work.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[For example, if you find a spreadsheet formula online, you need to understand how formulas work to apply it to your own spreadsheet, or maybe the dataset you're working with is too large for a single spreadsheet, and you'll need to switch to SQL.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**Having a variety of tools in your toolkit is important as a data analyst, but just as important is knowing when to use them**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[**If you find yourself stuck on a problem, it can be a good idea to take a step back and reconsider how you're approaching a task**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[We've covered a lot of tools that you can use as a data analyst throughout this program.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[We'll talk about R more later on, but here's a sneak peak to get you excited for it. **R is another programming language, but it's not a database language like SQL. It's a programming language frequently used for statistical analysis, visualization, and other data analysis. R is a little different from other tools we've been working with, but it's a great complement for the tools you're already using, and it will give you more potential solutions when you run into problems**.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[Using the thinking skills we've learned throughout this program, the right terms, and your understanding of different analysis tools, we'll get you ready for the next steps of this process: actually searching for answers online.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[There's a lot of resources like program support websites and forums where other data analysts are asking and answering questions. In an earlier video, we ran into a problem trying to calculate the time elapse between bike rides and the bike sharing data. Maybe our first search, "Calculate time in spreadsheets," didn't turn up the answers we needed.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[By thinking about our specific question and how other data analysts might be asking it, we could change that search into "Conditional formula for calculating elapsed time in spreadsheets." Now, we have more specific solutions to our problem.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[Finally, being able to modify example code to fit your own needs is so useful. Understanding the syntax of formulas and functions for different tools will allow you to take what you learned online and make it work for you, and maybe even build on it to create a whole new solution.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

[For example, the MOD formula we built to account for trips that started and ended on different days in our bike sharing data. The MOD formula we found online wasn't created for the data we were working with. But because we are familiar with spreadsheet tools, we were able to apply it to our data and use it as a solution to our problem.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)

**[Great data analysts know how to find and use resources online to help them build new solutions to problems they face. By using the thinking skills you've already learned in this program and using your knowledge of data analytics tools and terms, you can, too. Once you've found some answers to your problems, you can build them into your analysis work to get past any challenge you might face.](https://www.coursera.org/learn/analyze-data/lecture/5KRtk/running-into-challenges-not-to-worry)**

### [WHEN TO USE WHICH TOOL](https://www.coursera.org/learn/analyze-data/lecture/IlvEB/when-to-use-which-tool)

**Sometimes you find yourself stuck on a problem during your data analysis. That might mean it's time to reconsider which tool you're using for the job**.

For example, if you're working with a **simple spreadsheet**, maybe five to ten rows and a few columns, **then pivot tables are a great way to visualize that data**. But if that **spreadsheet is more than a million rows**, **it'll start to crash, making a pivot table hard to complete**.

When you find yourself **working with a huge spreadsheet that keeps crashing, you might switch to SQL to pull the data you need from different locations in a database instead of from a single spreadsheet**. You might remember that SQL can handle trillions of rows of data and is now a standard language for working with database programs. SQL is great for querying, updating, and optimizing data. **But trying to analyze your data with only SQL can get complicated**.

As you continue to progress as a data analyst, you might find yourself spending a lot of time building long, nested queries and then debugging them. It might be time to consider another tool, R.

R is a new tool that you'll work with later on, but for now, I'll tell you a little bit about it so that you can start getting excited. **R is another programming language, but it's not a database language like SQL. It's a programming language frequently used for statistical analysis, visualization, and other data analysis.**

R is a little different from other tools we've been working with, but it's a great complement for the tools you're already using. With R, you'll be able to analyze and visualize data in all kinds of new ways. We'll talk about R more later on, but I hope this sneak peek gives you an exciting first look.

Having a variety of tools in your tool kit is important as a data analyst, but just as important is knowing when to use them. If you find yourself stuck on a problem, it can be a good idea to take a step back and reconsider how you're approaching a task.

**Do you have too much data for a single spreadsheet? Switch to SQL. Are you spending more time debugging queries than actually analyzing data? Maybe you should consider R.** You also know how to find answers online now. So if you ever run into a problem and need to try a different tool, a quick search can be really helpful.

There might be resources online, or someone else may have had the same problem and posted about it. This is great if you start feeling stuck on a problem, and you might even find a new way to use a tool you're already familiar with. That brings us to the end of this module. Great job. We've covered a lot of information. We learned about converting and formatting data, how to combine multiple pieces of data, and how to search for help when you need support during your analysis.

### [SELF-REFLECTION: STACK OVERFLOW](https://www.coursera.org/learn/analyze-data/quiz/Opk0B/self-reflection-stack-overflow)



## **Activity overview**

**Empty alt text.**

In this activity, you’ll learn how to ask the data analytics community questions about the tools and programming languages you’ve been using in this course. You’ll examine best practices for writing questions, which can help you refine the questions you ask. Knowing where to find help and how to ask questions makes you a more efficient data analyst.

### 

### Step-By-Step Instructions

Follow the instructions to complete each step of the activity. Then answer the question at the end of the activity before going to the next course item.

### 

### **Step 1: Identify Stack Overflow as a platform to get help**

All data professionals get stuck at times and have questions. Fortunately, there are many resources available and a dynamic community to help. One such resource is Stack Overflow.

Stack Overflow is an online platform where data analysts and programmers ask code-related questions, and peers suggest answers. It provides an outlet for quick problem-solving and access to a wide-ranging knowledge base with networking and career development opportunities. You can ask questions about tools, programming languages, and much more.

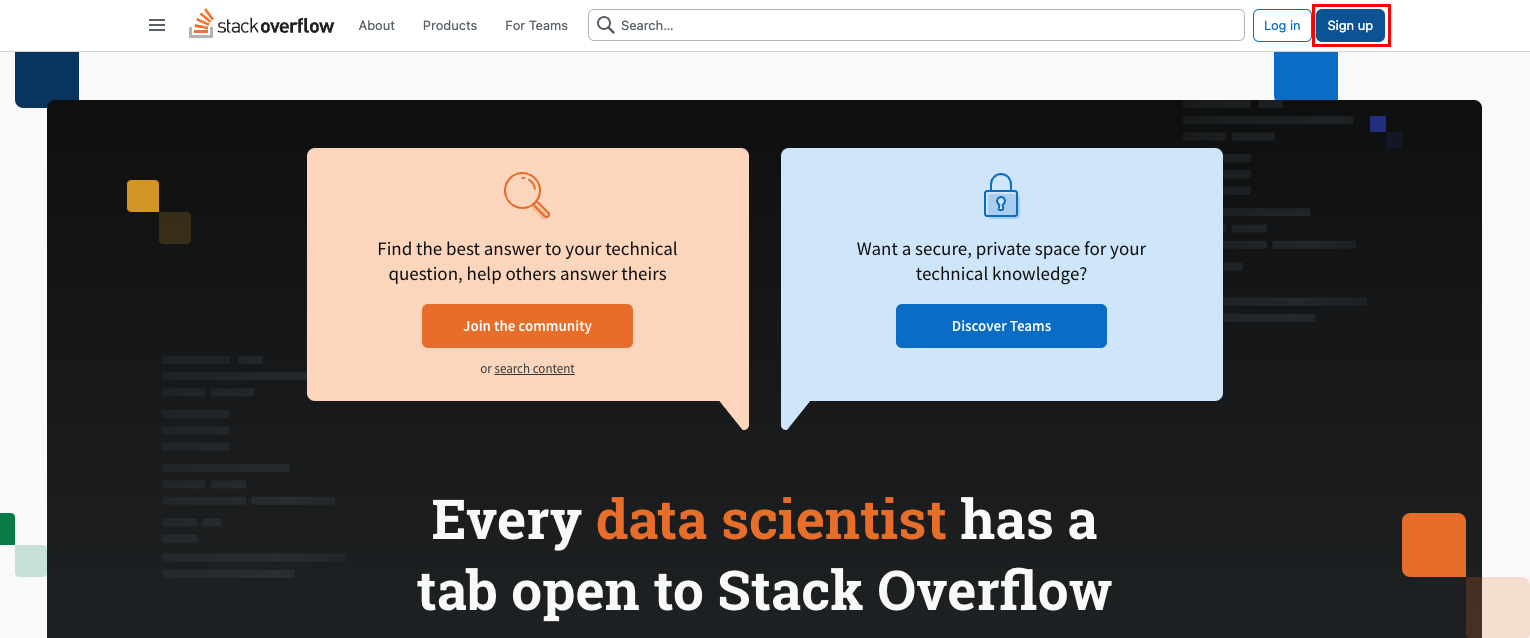
### 

### **Step 2: Sign up for a Stack Overflow account**

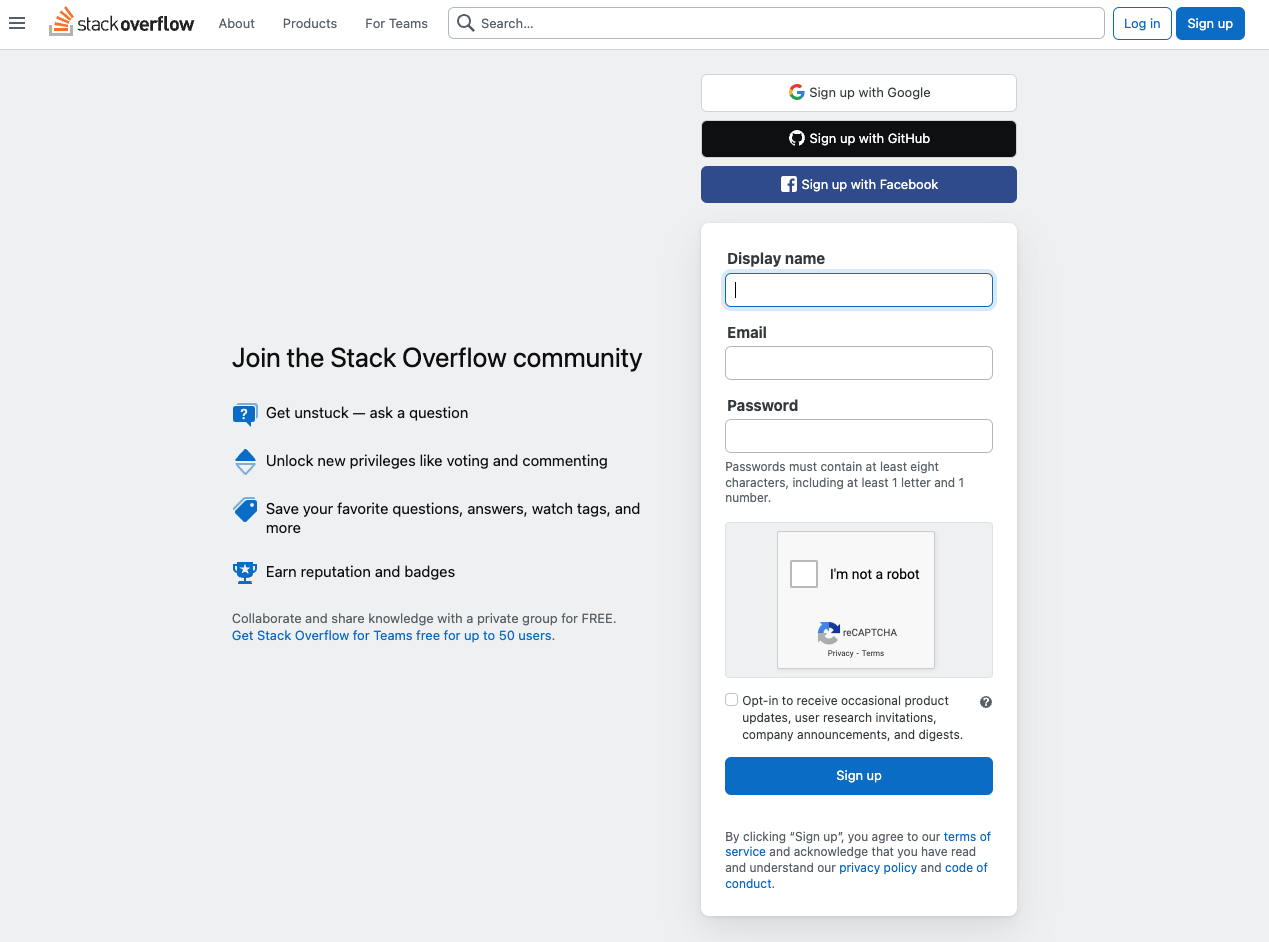
If you do not have an account, complete the following steps to sign up for Stack Overflow.

1. Navigate to the [Stack Overflow](https://stackoverflow.com/) website.

2. Select the Sign up button.



3. Enter the display name you want to use, your email address, a password, and any other required information.



Manual fields include Display name, Email, Password, and a check box that indicates you are not a robot. The Sign up button is below the check box.

4. Select Sign up.

5. Once you’ve signed up for your account, Stack Overflow will send you an email to complete your registration. Select the Complete your registration link.

6. You’ll be prompted to upload a new photo and add any tags about technology that you use or are interested in. Once you’ve made any desired updates, select Create my account.

7. You’ll be redirected to a welcome screen that provides you with an overview of the features and capabilities on Stack Overflow, as well as some how-to instruction.

### 

### **Step 3: Explore Stack Overflow**

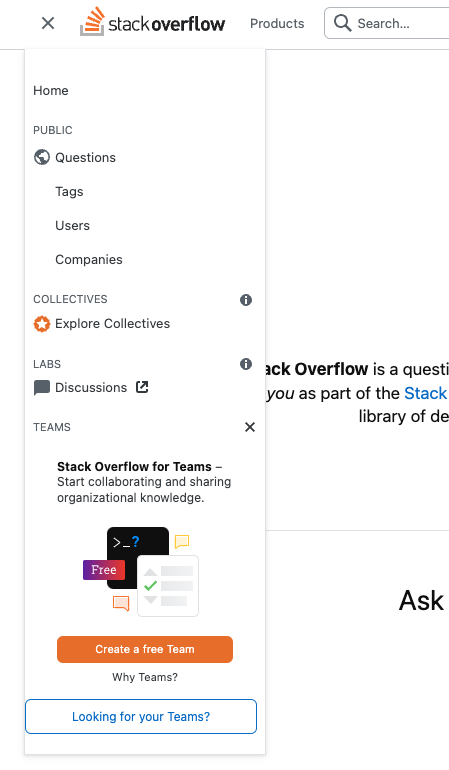
Stack Overflow has several resources you can use to search for information, including a Questions feature, a Tags feature, and a Search bar to help you find answers to the questions you might have.

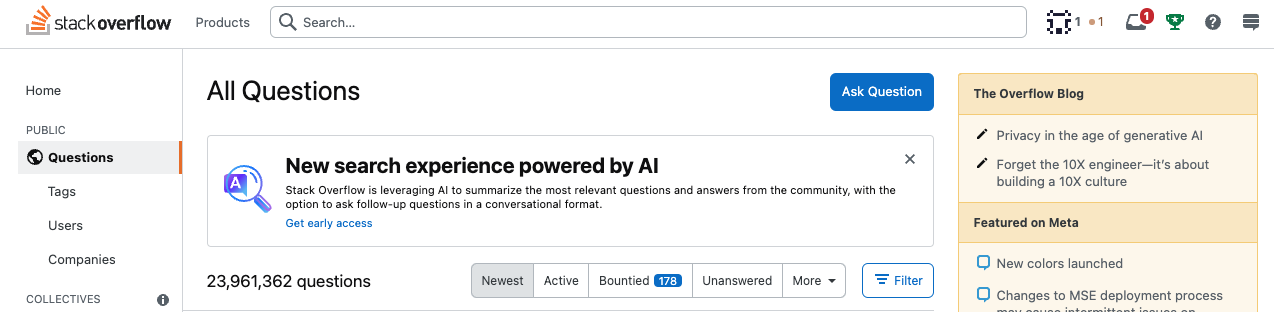
Questions

The Questions page provides different categories of questions for you to choose from. Some examples include Newest, Active, and Unanswered. To view all questions:

1. Select the Navigation menu (an icon of three stacked horizontal lines).
2. Select Questions.

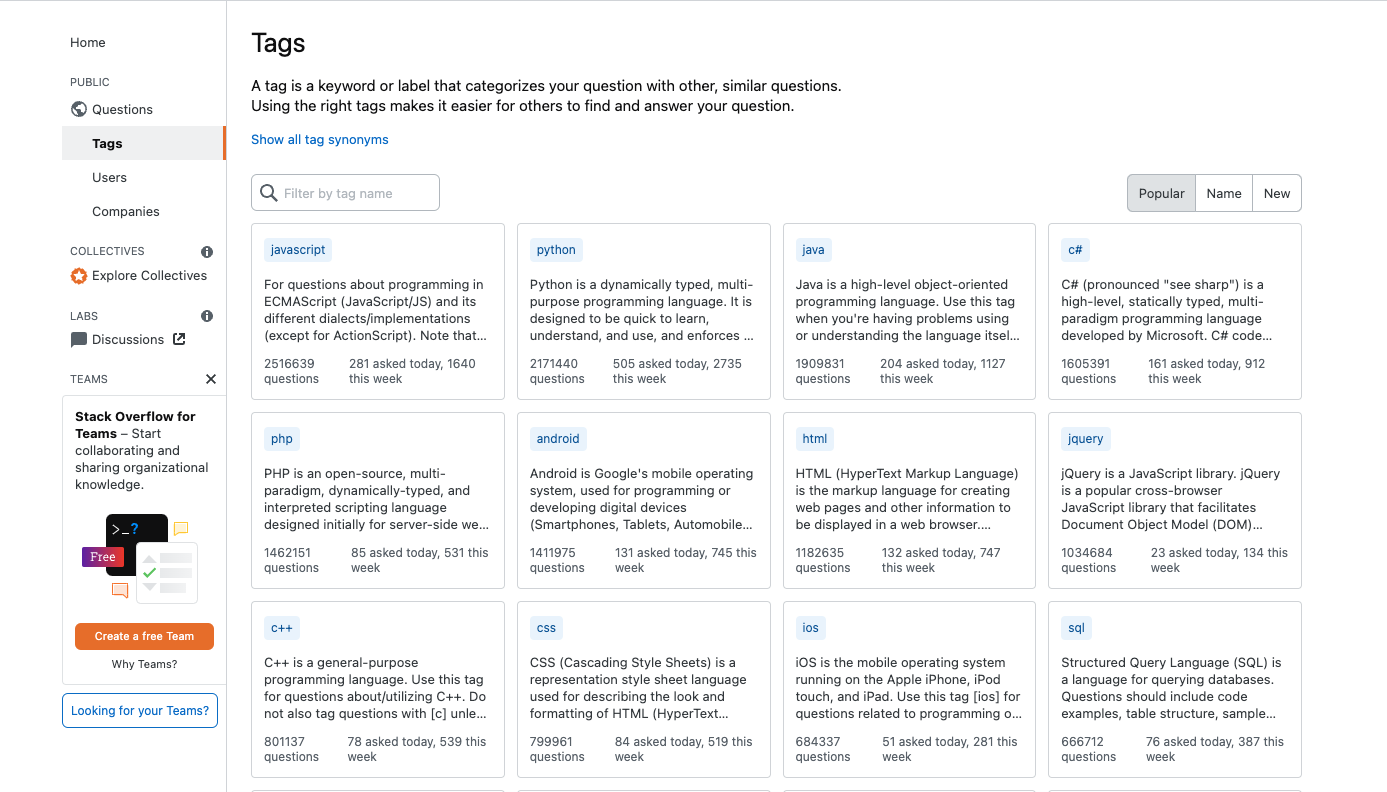
Explore the types of questions listed under the different categories.





Tags: Tags are keywords or labels that can help you find questions. To view the Tags page:

1. Open the Navigation menu, then select Tags.



To search for a specific tag:

1. Enter a tag name in the search bar, then press enter.
2. Select any tag to view a list of questions that have that particular tag.

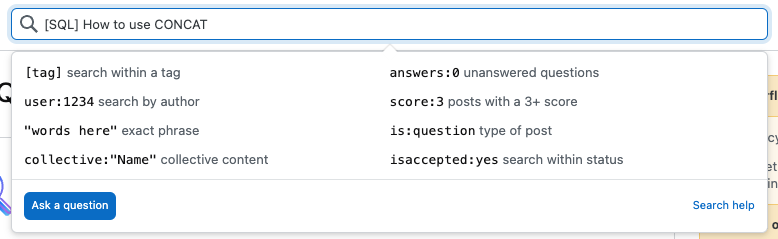
Different types of tags that you can search for include:

* Product lines
* Projects
* Teams
* Specific technologies
* Specific languages

You can also examine this [list of existing tags](https://stackoverflow.com/tags) for examples of which tags to use.

**Search bar**

Use the Search bar to search for keywords and questions. To view only questions with a specific tag, include the tag name in brackets within your search. For example, if you want to only find questions that have the tag “SQL,” then type [SQL] in the search field along with your keywords or question.



To learn more about searching, read these [instructions on how to search](https://stackoverflow.com/help/searching). For a quick guide on how different syntax structures can help you search more efficiently, go to this list of [search types and search syntax](https://stackoverflow.com/search).

### 

### **Step 4: Follow best practices for writing questions**

The following best practices can help set you up for success when asking questions on Stack Overflow:

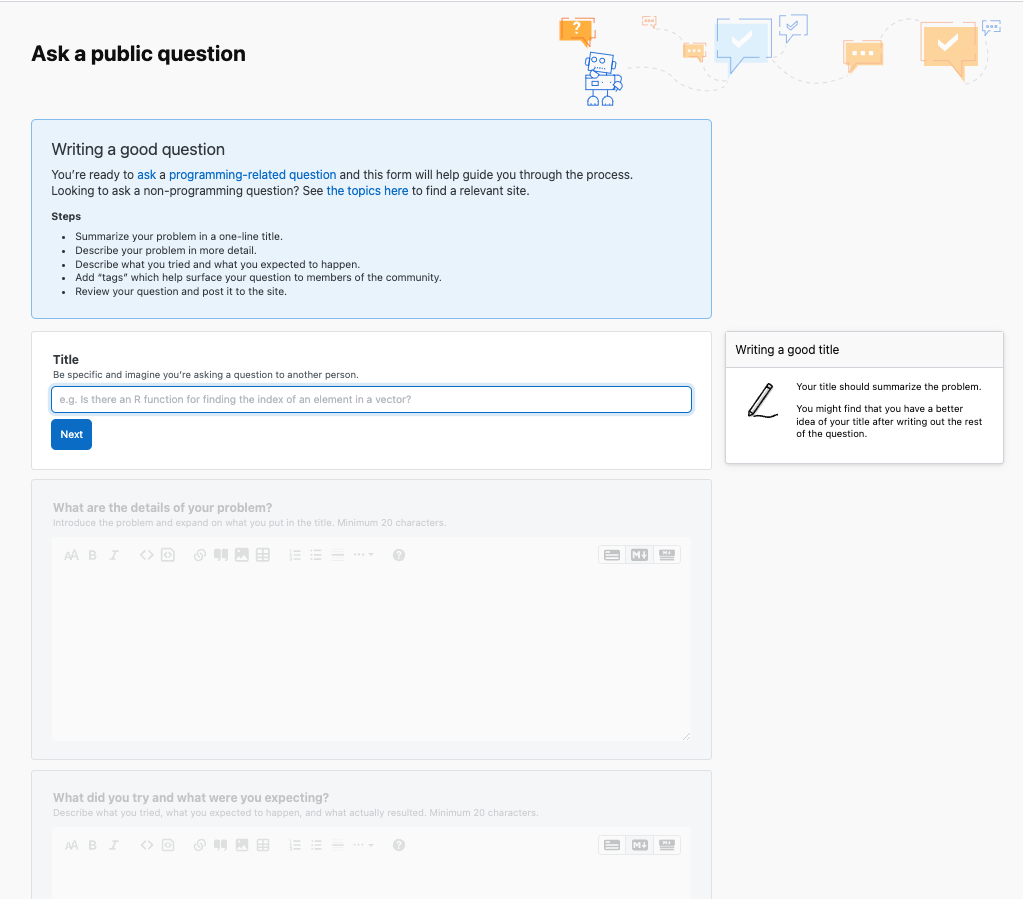
* Keep it specific.
* Don’t use Stack Overflow to ask questions with opinion-based answers. For example:
  + “Which SQL function can I use to add two numbers together?” is an appropriate question.
  + “Which SQL function is your favorite?” is not.
* Before asking a question, search the Stack Overflow website first—someone may have already asked it. This reduces the number of redundant questions on the site and saves you time.
* Write clear and concise questions in complete sentences. People are more likely to understand what you are asking and can give you more specific or helpful answers.

### 

### **Step 5: Ask your own question**

To ask your own question, follow these steps:

1. Navigate to the Stack Overflow homepage. Select Home from the Navigation menu or the Stack Overflow icon.
2. Select Ask Question. Note: If you need additional help, Stack Overflow displays some helpful tips for writing a good question and provides external links for additional resources.
3. Enter a title for your question. Make sure it’s specific! Then, select Next.
4. Enter the details of the problem you’re experiencing. Introduce the problem and explain the circumstances in which you encountered it. Then, select Next.
5. Add a description about what you tried, what you expected to happen, and what actually happened. If applicable, add a minimal, reproducible example of the code you’ve written so that someone can try to reproduce the issue. To insert a section of code, select Ctrl+K on your keyboard. Then, enter your code. When you have completed your description, select Next.
6. Add up to five tags that help describe your question. These tags will help to make sure your question is answered by the right people. Then, select Next.
7. Stack Overflow provides you with the ability to review questions that have already been asked to see if your question is a duplicate. The Stack Overflow community discourages repeat questions, so make sure to check previous questions thoroughly. If you find your question in the results listed, select that question and the hyperlink will redirect you to that question’s page in a new tab. You can also select Discard draft at the bottom of the form on the Ask a public question page.
8. If you don’t see your question listed, check the box to confirm that none of the posts answer your question. Then, select Review your question.
9. Review all the fields where you entered information about your question. If you need to make any changes, you can do so in the provided text boxes. If everything is correct, select Post your question.



Note: Stack Overflow is a public forum. Do not post any private information or confidential details.

## MODULE 2 CHALLENGE

### [GLOSSARY TERMS FROM MODULE 2](https://www.coursera.org/learn/analyze-data/supplement/slcjy/glossary-terms-from-module-2)

### 

### [MODULE 2 CHALLENGE](https://www.coursera.org/learn/analyze-data/exam/BWocw/module-2-challenge)