**C2 – MODULE 3**

**WORK WITH SPREADSHEETS**

[**THE AMAZING SPREADSHEET**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/cS8lM/the-amazing-spreadsheet)

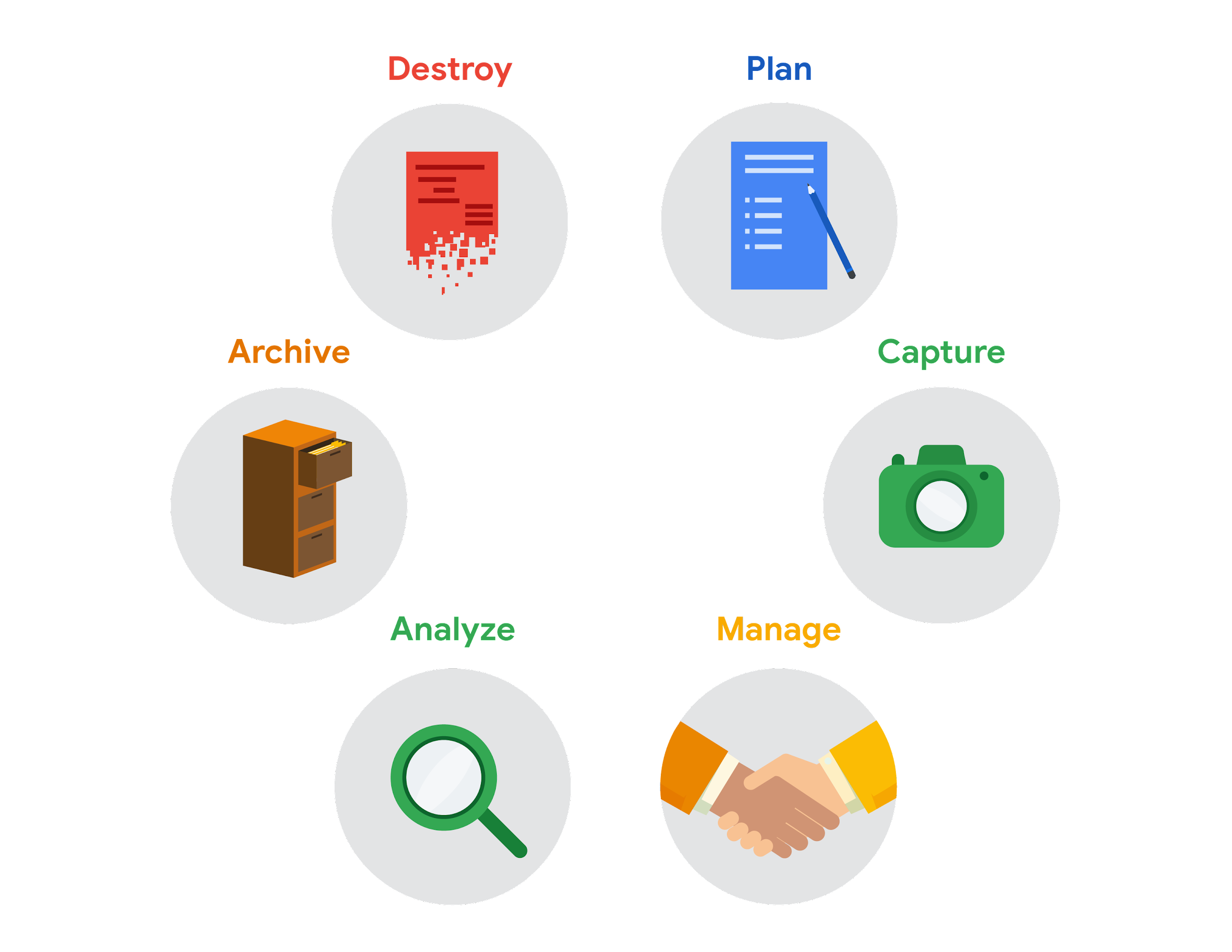
Spreadsheets are a powerful and versatile tool, which is why they're a big part of pretty much everything we do as data analysts. There's a good chance a spreadsheet will be the first tool you reach for when trying to answer data-driven questions. After you've defined what you need to do with the data, you'll turn to spreadsheets to help build evidence that you can then visualize, and use to support your findings. Spreadsheets are often the unsung heroes of the data world. They don't always get the appreciation they deserve, but as a data detective, you'll definitely want them in your evidence collection kit. I know spreadsheets have saved the day for me more than once. I've added data for purchase orders into a sheet, set up formulas in one tab, and had the same formulas do the work for me in other tabs. This frees up time for me to work on other things during the day.

[**GET TO WORK WITH SPREADSHEETS**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/182ab/get-to-work-with-spreadsheets)

Construction firm example: This will be different for each job, but you might start by organizing your data with the task you've been given.For example, you might put your data in a pivot table. Next, you might sort and filter the data in the pivot table. This lets you focus only on the data you'll need for your analysis. In our example, maybe you only need the expenses for a certain time frame, like the last three months. After you filtered your data, you could perform some calculations to learn more about it. Maybe you need to find out which construction projects ended up costing the most money. This is where formulas and functions are really handy.

[**SPREADSHEETS AND THE DATA LIFE CYCLE**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/ALVYZ/spreadsheets-and-the-data-life-cycle)

To better understand the benefits of using spreadsheets in data analytics, let’s explore how they relate to each phase of the data life cycle: plan, capture, manage, analyze, archive, and destroy.

****

* Plan for the users who will work within a spreadsheet by developing organizational standards. This can mean formatting your cells, the headings you choose to highlight, the color scheme, and the way you order your data points. When you take the time to set these standards, you will improve communication, ensure consistency, and help people be more efficient with their time.
* Capture data by the source by connecting spreadsheets to other data sources, such as an online survey application or a database. This data will automatically be updated in the spreadsheet. That way, the information is always as current and accurate as possible.
* Manage different kinds of data with a spreadsheet. This can involve storing, organizing, filtering, and updating information. Spreadsheets also let you decide who can access the data, how the information is shared, and how to keep your data safe and secure.
* Analyze data in a spreadsheet to help make better decisions. Some of the most common spreadsheet analysis tools include formulas to aggregate data or create reports, and pivot tables for clear, easy-to-understand visuals.
* Archive any spreadsheet that you don’t use often, but might need to reference later with built-in tools. This is especially useful if you want to store historical data before it gets updated.
* Destroy your spreadsheet when you are certain that you will never need it again, if you have better backup copies, or for legal or security reasons. Keep in mind, lots of businesses are required to follow certain rules or have measures in place to make sure data is destroyed properly.

## **Resources for more information**

Spreadsheet shortcuts can help you become more efficient with spreadsheets. If you’d like to learn more, you can explore the collection of [Google Sheets shortcuts](https://support.google.com/docs/answer/181110), or visit the [Microsoft Excel shortcuts](https://support.microsoft.com/en-us/office/keyboard-shortcuts-in-excel-1798d9d5-842a-42b8-9c99-9b7213f0040f) page if you are using Excel. Both of these resources contain a list of spreadsheet shortcuts you can save and reference as you work more with spreadsheets on your own.

**[HANDS-ON ACTIVITY: INTRODUCTION TO GOOGLE SHEETS](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/xMQnU/hands-on-activity-introduction-to-google-sheets)**

[**STEP-BY-STEP: BASIC SPREADSHEET TASKS**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/rhx5S/step-by-step-basic-spreadsheet-tasks)

[**BASIC SPREADSHEET TASKS**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/lpuHf/basic-spreadsheet-tasks)

**[LEARN MORE ABOUT SPREADSHEET BASICS](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/0vdZZ/learn-more-about-spreadsheet-basics)**

Below, you will find a list that covers two types of spreadsheet programs: Microsoft Excel and Google Sheets. The list includes quick-start guides, tutorials, and more. The examples in this course use Google Sheets, but you can follow along using Excel or any other spreadsheet application. The user interface might be a little different, but it should look and work similarly.



## **Microsoft Excel**

* [Office Quick Starts](https://support.microsoft.com/en-us/office/office-quick-starts-25f909da-3e76-443d-94f4-6cdf7dedc51e#ID0EAADAAA=At_work_or_school): Scroll down to the Downloadable guides section to download the Excel Quick Start Guide: This PDF guide begins with a labeled map of Excel that can guide you through the basic tasks you can accomplish in Excel. For tips on starting and opening Excel, this [Microsoft Support page](https://support.microsoft.com/en-us/office/create-a-new-workbook-ae99f19b-cecb-4aa0-92c8-7126d6212a83?wt.mc_id=otc_excel) will show you how to begin a new workbook.
* [Excel video training](https://support.microsoft.com/en-us/office/excel-for-windows-training-9bc05390-e94c-46af-a5b3-d7c22f6990bb?wt.mc_id=otc_home): This is a collection of step-by-step videos to use all sorts of Excel features, including adding and working within rows, columns, and cells; formatting; using formulas and functions; and adding charts and pivot tables.
* [Sort data in a range or table](https://support.microsoft.com/en-us/office/sort-data-in-a-range-or-table-62d0b95d-2a90-4610-a6ae-2e545c4a4654): This page guides you through all of the steps you will need to sort data by number, text, and color. You’ll also have the option to sort by custom list so that you can customize exactly what you want to sort.
* [Filter data in a range or table](https://support.microsoft.com/en-us/office/filter-data-in-a-range-or-table-01832226-31b5-4568-8806-38c37dcc180e): This article has step-by-step instructions on how to filter an Excel spreadsheet to show only the data you want to see. You can also use built-in comparison operators, such as “greater than” and “top 10” to reveal only the most relevant data.
* [Format a worksheet](https://support.microsoft.com/en-us/office/quick-start-format-a-worksheet-d70f75a2-23e6-4c92-83d6-2f219e4ad42e): The guide will help you select and format your Excel spreadsheet, then change the borders, shading, colors, and text. This can help improve your spreadsheet’s readability.

Pro tip: If you’re searching for information about using customizable options, check out Microsoft’s [Guidelines for organizing and formatting data on a worksheet](https://support.microsoft.com/en-us/office/guidelines-for-organizing-and-formatting-data-on-a-worksheet-90895cad-6c85-4e02-90d3-8798660166e3). This article provides clear methods for creating easy-to-read spreadsheets.

## **Google Sheets**

* [Google Sheets cheat sheet](https://support.google.com/a/users/answer/9300022): The cheat sheet puts all the basics of Sheets on a single page for easy reference. Here, you can learn about customizing your spreadsheet and the data inside; working with rows, columns, and cells; sharing your spreadsheet with others; creating different versions and copies of a spreadsheet; and more.
* [Get started with Sheets: Create and import files](https://support.google.com/a/users/answer/9300311?hl=en&ref_topic=9296423): This guide is a step-by-step guide for working with Sheets. You start by learning how to open a spreadsheet, then move on to adding data.
* [Sort and filter your data](https://support.google.com/docs/answer/3540681?co=GENIE.Platform%3DDesktop&hl=en): This resource can help you organize data in Sheets. Use this guide to sort part or all of a spreadsheet. You can sort by text, number, and color. Then, learn how to create filters to show only certain data while hiding the rest. Finally, the article includes information on creating, saving, and removing a filter view.
* [Edit and format a spreadsheet](https://support.google.com/docs/answer/46973?co=GENIE.Platform%3DDesktop&hl=en&oco=0): This will help you make easy-to-read spreadsheets. You will learn how to assign a color, customize borders around cells, and change the appearance of text. If you’d like to give your spreadsheet a theme, you can scroll to the bottom of the page and find how to apply it to parts of your spreadsheet.

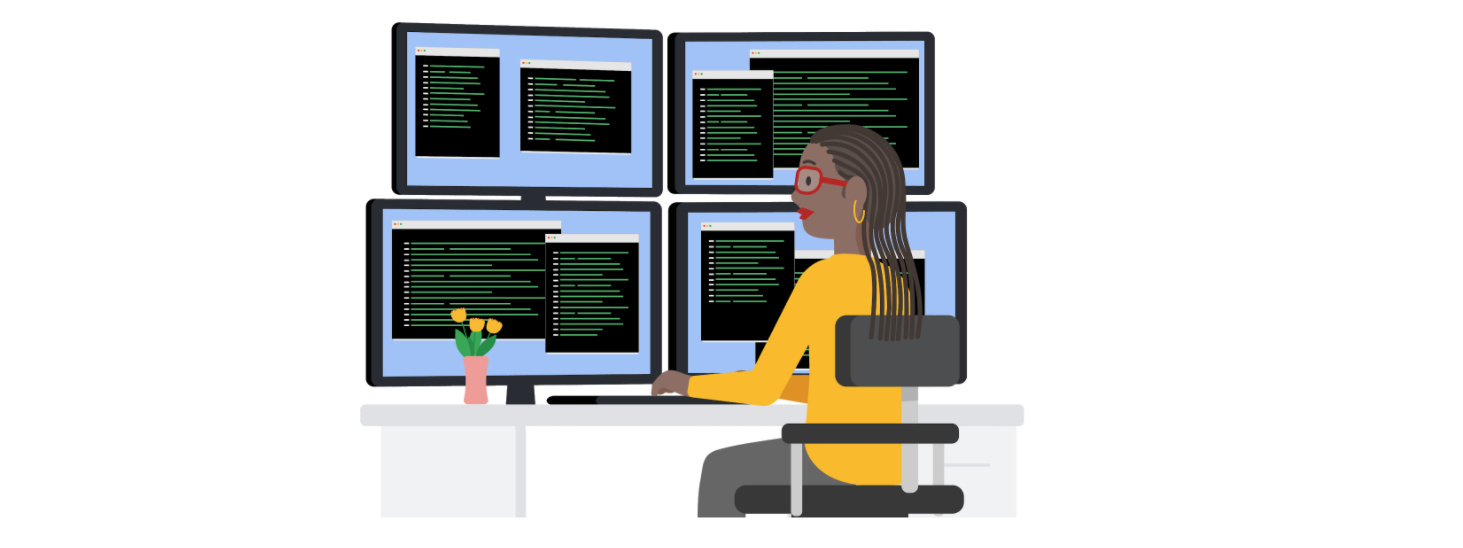
Tip: Microsoft Excel and Google Sheets are very similar in terms of calculations, formulas, functions, and many other features. But there are some differences, which can make it tricky to switch from one to the other. If you are moving between Excel and Google Sheets, find a quick list of the differences between the two kinds of spreadsheet applications in [Overview: Differences between Sheets and Excel](https://support.google.com/a/users/answer/9331278?hl=en).

**FORMULAS IN SPREADSHEETS**

[**STEP-BY-STEP: FORMULAS FOR SUCCESS**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/dr2GC/step-by-step-formulas-for-success)

[**FORMULAS FOR SUCCESS**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/s4RlB/formulas-for-success)

[**QUICK REFERENCE: FORMULAS IN SPREADSHEETS**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/ASfb7/quick-reference-formulas-in-spreadsheets)



You have been learning a lot about spreadsheets and all kinds of time-saving calculations and organizational features they offer. One of the most valuable spreadsheet features is a formula. As a quick reminder, a formula is a set of instructions that does a specific calculation using the data in a spreadsheet. Formulas make it easy for data analysts to do powerful calculations automatically, which helps them analyze data more effectively. Below is a quick-reference guide to help you get the most out of formulas.

## **FORMULAS**

### **The basics**

* When you enter a formula in math, it generally ends with an equal sign (2 + 3 = ?). But with formulas, they always start with one instead (=A2+A3). The equal sign tells the spreadsheet that what follows is part of a formula, not just a word or number in a cell.
* After you enter the equal sign, most spreadsheet applications will display an autocomplete menu that lists valid formulas, names, and text strings. This is a great way to create and edit formulas while avoiding typing and syntax errors.
* A fun way to learn new formulas is just by typing an equal sign and a single letter of the alphabet. Choose one of the options that pops up and you will learn what that formula does.

### **Mathematical operators**

* The mathematical operators used in spreadsheet formulas include:
* Subtraction – minus sign ( - )
* Addition – plus sign ( + )
* Division – forward-slash ( / )
* Multiplication – asterisk ( \* )

### **Auto-filling**

The lower-right corner of each cell has a fill handle. It is a small *green square* in Microsoft Excel and a small *blue circle* in Google Sheets.

* Click the fill handle square or circle for a cell and drag it down a column to auto-fill other cells in the column with the same value or formula in that cell.
* Click the fill handle square or circle for a cell and drag it across a row to auto-fill other cells in the row with the same value or formula in that cell.
* If you want to create a numbered sequence in a column or row, do the following: 1) Fill in the first two numbers of the sequence in two adjacent cells, 2) Select to highlight the cells, and 3) Drag the fill handle square or circle to the last cell to complete the sequence of numbers. For example, to insert 1 through 100 in each row of column A, enter 1 in cell A1 and 2 in cell A2. Then, select to highlight both cells, click the fill handle square or circle in cell A2, and drag it down to cell A100. This auto-fills the numbers sequentially so you don't have to enter them in each cell.

### **Absolute referencing**

* Absolute referencing is marked by a dollar sign ($). For example, =$A$10 has absolute referencing for both the column and the row value
* Relative references (which is what you normally do, e.g. “=A10”) will change anytime the formula is copied and pasted. They are in relation to where the referenced cell is located. For example if you copied “=A10” to the cell to the right it would become “=B10”. With absolute referencing “=$A$10” copied to the cell to the right would remain “=$A$10”. But if you copied $A10 to the cell below, it would change to $A11 because the row value isn't an absolute reference.
* Absolute references will not change when you copy and paste the formula in a different cell. The cell being referenced is always the same.
* To easily switch between absolute and relative referencing in the formula bar, highlight the reference you want to change and press the F4 key; for example, if you want to change the absolute reference, $A$10, in your formula to a relative reference, A10, highlight $A$10 in the formula bar and then press the F4 key to make the change.

### **Data range**

* When you click into your formula, the colored ranges let you see which cells are being used in your spreadsheet. There are different colors for each unique range in your formula.
* In a lot of spreadsheet applications, you can press the F2 (or Enter) key to highlight the range of data in the spreadsheet that is referenced in a formula. Click the cell with the formula, and then press the F2 (or Enter) key to highlight the data in your spreadsheet.

### **Combining with functions**

* COUNTIF() is a formula and a function. This means the function runs based on criteria set by the formula. In this case, COUNT is the formula; it will be executed IF the conditions you create are true. For example, you could use =COUNTIF(A1:A16, “7”) to count only the cells that contained the number 7. Combining formulas and functions allows you to do more work with a single command.

[**HANDS-ON ACTIVITY: DATA ANALYSIS AND FORMULAS: BAKERY SALES INSIGHTS**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/Yxy6K/hands-on-activity-data-analysis-and-formulas-bakery-sales-insights)

[**SPREADSHEET ERRORS AND FIXES**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/VbM4p/spreadsheet-errors-and-fixes)

Sometimes data analysts encounter a problem with our formulas and we get an error.

One **error you may encounter** is the **DIV error**. The DIV error happens when a formula is trying to divide a value in a cell by zero or by an empty cell.**To avoid this problem**, we can have this spreadsheet automatically enter not applicable whenever a cell in column A contains a zero that would cause the error. To do this, we'll use the **IFERROR function**. If it encounters a DIV error caused by a cell that contains the zero, the phrase "Not applicable" will be inserted.   
  
Another common error is the **#ERROR!** tells us the formula can't be interpreted as it is input. This is AKA a **parsing error**. Say we want to tally the number of total tasks in column B and C, we use the SUM function, but the formula equal sum B2 to B6, C2 to C6 causes an error. Examining it more closely, we see that a comma is missing between the cell ranges B2 to B6 and C2 to C6. We can fix this by **inserting a comma between the cell ranges to indicate the end of each data item**. This is called a **delimiter**.

The **N/A error** tells you that the data in your formula can't be found by the spreadsheet. Generally, this means the data doesn't exist. This error most often occurs when using functions such as **VLOOKUP**, which searches for a certain value in a column to return a corresponding piece of information. Here, we see a master list of nuts and their prices. Using VLOOKUP, the spreadsheet finds prices in the list, then calculates the prices for each store using the assigned markup. But we have a N/A error in cells B49 and C49. The VLOOKUP formula is correct, so what's going on? Well, if we look carefully at the name of the nut, "almond" has no match in the lookup table, the lookup table uses the plural "almonds" instead. So we change “almond” to almonds, and with that typo fixed, the right prices are filled in. Speaking of typos, sometimes a typo can cause a **NAME error**. A NAME error can happen when a formula's name isn't recognized or understood. Suppose we see a NAME error in the nut prices spreadsheet. If we look carefully, the VLOOKUP function in cell B21 is spelled incorrectly, it has one extra O; this causes a NAME error for both the price and the resulting markup calculation for the store. To fix this error, we can delete the extra O in VLOOKUP.

Sometimes an error is caused by inconsistent or wrong data.

For instance, the **NUM error** tells us that a formula's calculation can't be performed as specified by the data. The data doesn't make sense for that calculation. Here's what I mean. Suppose we're working on a large construction project using a spreadsheet to track how many months it takes to reach key milestones. We can use the **DATEDIF function to calculate the number of months between start and end dates**. The function requires the start date to be in the first cell referenced and the end date to be in the second cell referenced. In our case, cells B2 and C2 respectively. The M represents months, as we want this spreadsheet to calculate the number of months between our start and end dates. But we get a NUM error in cell D6. We notice that the end date comes before the start date, so the **DATEDIF function** can't calculate the number of months between. It's likely the start and end dates were interchanged by accident. We can request verification of the data to make sure. In the meantime, let's reverse the order of the cells in the formula to temporarily get around the error. Now, the result is nine months. What if the client's name was accidentally inserted into the start date in the spreadsheet? You guessed it, we get an error. The VALUE error can indicate a problem with a formula or referenced cells. It's often not clear right away what the problem is, so this error might take a little more effort to fix. In this case, John Welty was input as the start date, making the calculation impossible for the DATEDIF function in the cell D6. We just replaced the text, John Welty, with the correct start date of September 1st, 2016.

Last is the **REF error**, which often comes up when cells being referenced in a formula have been deleted, thus making the formula unable to perform the calculation. Here's a spreadsheet used to calculate the number of seats available for a company lunch. Let's say the company decided not to run the second floor, so we delete row 4. This results in a REF error when calculating the total seats available in cell B5. To fix this, we can change the formula to add the values in cells B2 and B3. Also, in this case, we could have prevented the REF error by using the SUM function and a range of cells instead of adding the cell value by direct reference. Now, if we delete row 10, the SUM function calculates the total seats available. There you go. We've now fixed some of the most common spreadsheet errors. When you see them again, you'll know what they mean.

Troubleshooting is a big part of data analysis, so being able to find solutions is a key skill for data analysts.

[**MORE ABOUT SPREADSHEET ERRORS AND FIXES**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/1sjGm/more-about-spreadsheet-errors-and-fixes)

When you are new to data analytics—and sometimes even when you aren't—spreadsheet struggles are real. It never feels good when you enter in what you are sure is a perfect formula or function, only to get an error message. Understanding errors and how to fix them is a big part of keeping your data clean, so it’s important to know how to deal with issues as they come up, and more importantly, not to get discouraged.

Even the most advanced spreadsheet users come across problems from time to time.



As a follow-up to what you learned in the previous video, here are a few best practices and helpful tips. These strategies will help you avoid spreadsheet errors to begin with, making your life in analytics a whole lot less stressful:

1. Filter data to make your spreadsheet less complex and busy.
2. Use and freeze headers so you know what is in each column, even when scrolling.
3. When multiplying numbers, use an asterisk (**\***) not an **X**.
4. Start every formula and function with an equal sign (**=**).
5. Whenever you use an open parenthesis, make sure there is a closed parenthesis on the other end to match.
6. Change the font to something easy to read.
7. Set the border colors to white so that you are working in a blank sheet.
8. Create a tab with just the raw data, and a separate tab with just the data you need.

Now that you have learned some basic ways to avoid errors, you can focus on what to do when that dreaded pop-up does appear. The following table is a reference you can use to look up common spreadsheet errors and examples of each. Knowing what the errors mean takes some of the fear out of getting them.

| **Error** | **Description** | **Example** |
| --- | --- | --- |
| **#DIV/0!** | A formula is trying to divide a value in a cell by 0 (or an empty cell with no value) | **=B2/B3**, when the cell B3 contains the value 0 |
| **#ERROR!** | (Google Sheets only) Something can’t be interpreted as it has been input. This is also known as a parsing error. | **=COUNT(B1:D1 C1:C10)** is invalid because the cell ranges aren't separated by a comma |
| **#N/A** | A formula can't find the data | The cell being referenced can't be found |
| **#NAME?** | The name of a formula or function used isn't recognized | The name of a function is misspelled |
| **#NUM!** | The spreadsheet can't perform a formula calculation because a cell has an invalid numeric value | **=DATEDIF(A4, B4, "M")** is unable to calculate the number of months between two dates because the date in cell A4 falls after the date in cell B4 |
| **#REF!** | A formula is referencing a cell that isn't valid | A cell used in a formula was in a column that was deleted |
| **#VALUE!** | A general error indicating a problem with a formula or with referenced cells | There could be problems with spaces or text, or with referenced cells in a formula; you may have additional work to find the source of the problem. |

If you are working with Microsoft Excel, an interactive page, [How to correct a #VALUE! error](https://support.microsoft.com/en-us/office/how-to-correct-a-value-error-15e1b616-fbf2-4147-9c0b-0a11a20e409e), can help you narrow down the cause of this error. You can select a specific function from a drop-down list to display a link to tips to fix the error when using that function.

### PRO TIP: SPOTTING ERRORS IN SPREADSHEETS WITH CONDITIONAL FORMATTING

**Conditional formatting can be used to highlight cells a different color based on their contents.** This feature can be extremely helpful when you want to locate all errors in a large spreadsheet. For example, using conditional formatting, you can highlight in yellow all cells that contain an error, and then work to fix them.

### CONDITIONAL FORMATTING IN MICROSOFT EXCEL

To set up conditional formatting in Microsoft Excel to highlight all cells in a spreadsheet that contain errors, do the following:

1. Click the gray triangle above row number 1 and to the left of Column A to select all cells in the spreadsheet.
2. From the main menu, click **Home**, and then click **Conditional Formatting** to select **Highlight Cell Rules > More Rules**.
3. For **Select a Rule Type,** choose **Use a formula to determine which cells to format**.
4. For Format values where this formula is true, enter **=ISERROR(A1)**.
5. Click the **Format** button, select the **Fill** tab, select yellow (or any other color), and then click **OK**.
6. Click **OK** to close the format rule window.

To remove conditional formatting, click **Home** and select **Conditional Formatting,** and then click **Manage Rules.** Locate the format rule in the list, click **Delete Rule,** and then click **OK.**

### CONDITIONAL FORMATTING IN GOOGLE SHEETS

To set up conditional formatting in Google Sheets to highlight all cells in a spreadsheet that contain errors, do the following:

1. Click the empty rectangle above row number 1 and to the left of Column A to select all cells in the spreadsheet. In the [Step-by-step in spreadsheets](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/lpuHf/step-by-step-in-spreadsheets) video, this was called the **Select All** button.
2. From the main menu, click **Format** and select **Conditional Formatting** to open the Conditional format rules pane on the right.
3. While in the **Single Color** tab, under Format rules, use the drop-down to select **Custom formula is,** enter **=ISERROR(A1)**, select yellow (or any other color) for the formatting style, and then click **Done**.

To remove conditional formatting, click **Format** and select **Conditional Formatting,** and then click the **Trash** icon for the format rule.

## 

### SPREADSHEET ERROR RESOURCES

To learn more and read about additional examples of errors and solutions, explore these resources:

* [**Microsoft Formulas and Functions**](https://support.microsoft.com/en-us/office/formulas-and-functions-294d9486-b332-48ed-b489-abe7d0f9eda9?ui=en-US&rs=en-US&ad=US#id0eaabaaa=errors)**:** This resource describes how to avoid broken formulas and how to correct errors in Microsoft Excel. This is a useful reference to have saved in case you run into a specific error and need to find solutions quickly while working in Excel.
* [**When Your Formula Doesn’t Work: Formula Parse Errors in Google Sheets**](https://www.benlcollins.com/spreadsheets/formula-parse-error/): This resource is a guide to finding and fixing some common errors in Google Sheets. If you are working with Google Sheets, you can use this as a quick reference for solving problems you might encounter working on your own.

With some practice and investigative determination, you will become much more comfortable handling errors in spreadsheets. Each error you catch and fix will make your data clearer, cleaner, and more useful.

**[HANDS-ON ACTIVITY: RESOLVE SPREADSHEET ERRORS](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/37ePh/hands-on-activity-resolve-spreadsheet-errors)**

[**TEST YOUR KNOWLEDGE ON USING FORMULAS IN SPREADSHEETS**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/QwB0w/test-your-knowledge-on-using-formulas-in-spreadsheets)

**FUNCTIONS IN SPREADSHEETS**

**[STEP-BY-STEP: FUNCTIONS 101](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/MAdSM/step-by-step-functions-101)**

[**FUNCTIONS 101**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/BO2Ql/functions-101)

In the world of spreadsheets a **function is a preset command that automatically performs a specific process or task using the data.**

[**QUICK REFERENCE: FUNCTIONS IN SPREADSHEETS**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/EaHLl/quick-reference-functions-in-spreadsheets)

As a quick refresher, a function is a preset command that automatically performs a specific process or task using the data in a spreadsheet. Functions give data analysts the ability to do calculations, which can be anything from simple arithmetic to complex equations. Use this reading to help you keep track of some of the most useful options.

## **Functions**

### **The basics**

* Just like formulas, start all of your functions with an equal sign; for example **=SUM**. The equal sign tells the spreadsheet that what follows is part of a function, not just a word or number in a cell.
* After you enter the equal sign, most spreadsheet applications will display an autocomplete menu that lists valid functions, names, and text strings. This is a great way to create and edit functions while avoiding typing and syntax errors.
* A fun way to learn new functions is by simply typing an equal sign and a single letter of the alphabet. Choose one of the options that pops up and learn what that function does.

### **Difference between formulas and functions**

* A formula is a set of instructions used to perform a calculation using the data in a spreadsheet.
* A function is a preset command that automatically performs a specific process or task using the data in a spreadsheet.

### **Popular functions**

A lot of people don’t realize that keyboard shortcuts like cut, save, and find are actually functions. These functions are built into an application and are amazing time-savers. Using shortcuts lets you do more with less effort. They can make you more efficient and productive because you are not constantly reaching for the mouse and navigating menus. Use these links to discover the most popular shortcuts, for [Chromebook](https://support.google.com/chromebook/answer/183101?hl=en), [PC](https://support.microsoft.com/en-us/windows/keyboard-shortcuts-in-windows-dcc61a57-8ff0-cffe-9796-cb9706c75eec), and [Mac](https://support.apple.com/en-us/HT201236).

### **Auto-filling**

The lower-right corner of each cell has a fill handle. It is a small green square in Microsoft Excel and a small blue circle in Google Sheets.

* Click the fill handle for a cell and drag it down a column to auto-fill other cells in the column with the same formula or function used in that cell.
* Click the fill handle for a cell and drag it across a row to auto-fill other cells in the row with the same formula or function used in that cell.

### **Relative, absolute, and mixed references**

* Relative references (cells referenced without a dollar sign, like **A2**) will change when you copy and paste the function into a different cell. With relative references, the location of the cell that contains the function determines the cells used by the function.
* Absolute references (cells fully referenced with a dollar sign, like **$A$2**) will not change when you copy and paste the function into a different cell. With absolute references, the cells referenced always remain the same.
* Mixed references (cells partially referenced with a dollar sign, like **$A2** or **A$2**) will change when you copy and paste the function into a different cell. With mixed references, the location of the cell that contains the function determines the cells used by the function, but only the row or column is relative (not both).
* In spreadsheets, you can press the **F4** key to toggle between relative, absolute, and mixed references in a function. Click the cell containing the function, highlight the referenced cells in the formula bar, and then press **F4** to toggle between and select relative, absolute, or mixed referencing.

### **Data ranges**

* When you click a cell that contains a function, colored data ranges in the formula bar indicate which cells are being used in the spreadsheet. There are different colors for each unique range in a function.
* Colored data ranges help prevent you from getting lost in complex functions.
* In spreadsheets, you can press the **F2** key to highlight the range of data used by a function. Click the cell containing the function, highlight the range of data used by the function in the formula bar, and then press **F2**. The spreadsheet will go to and highlight the cells specified by the range.

### **Data ranges evaluated for a condition**

**COUNTIF** is an example of a function that returns a value based on a condition that the data range is evaluated for. The function counts the number of cells that meet the criteria. For example, in an expense spreadsheet, use **COUNTIF** to count the number of cells that contain a reimbursement for "airfare."

For more information, refer to:

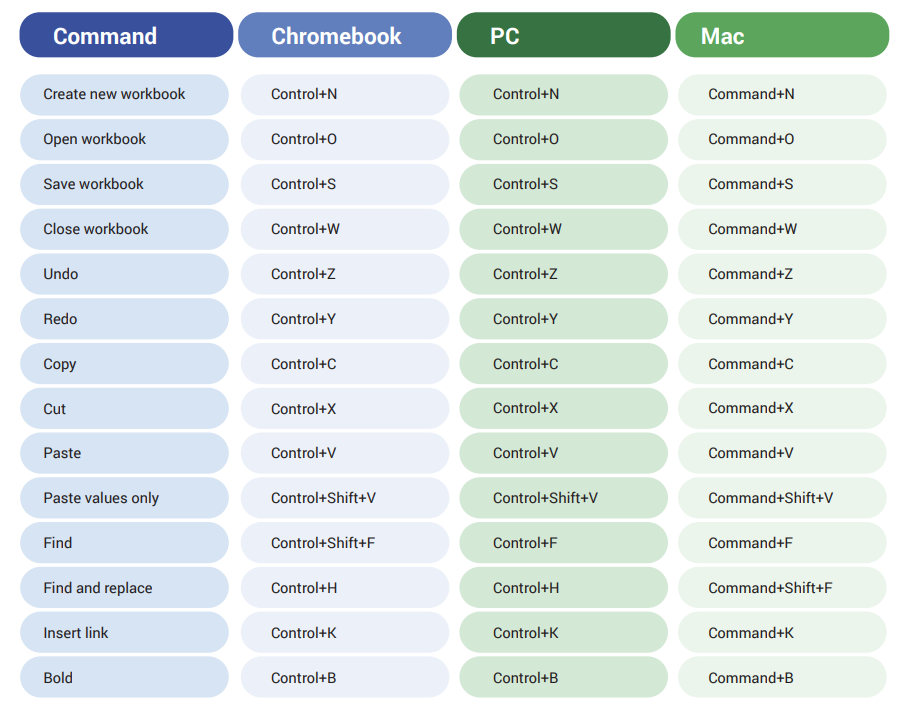
* [Microsoft Support's page for COUNTIF](https://support.microsoft.com/en-us/office/countif-function-e0de10c6-f885-4e71-abb4-1f464816df34)
* [Google Help Center's documentation for COUNTIF](https://support.google.com/docs/answer/3093480?hl=en) where you can copy a sheet with [COUNTIF examples](https://docs.google.com/spreadsheets/d/1PYoKCYZAkWSaMBsiTyvxZzCCt2WQ-QKOC763RWHMB7c/template/preview) (click "Use Template" if you click the COUNTIF link provided on this page)

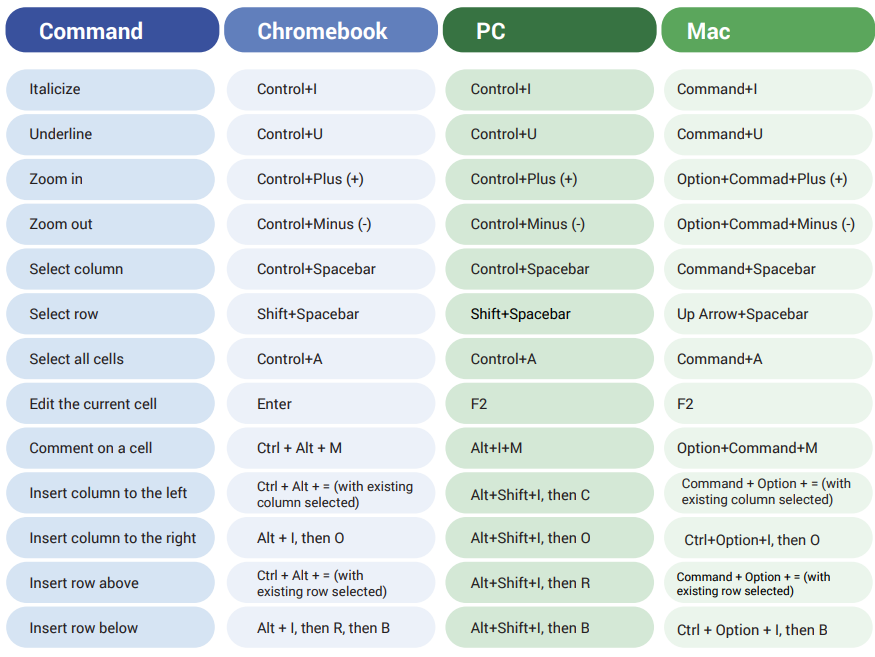
## **Key takeaways**

There are a lot more functions that can help you make the most of your data. This is just the start. You can keep learning how to use functions to help you solve complex problems efficiently and accurately throughout your entire career.

## **Keyboard shortcuts**

*You can save these functions for future reference. Feel free to download a .****pdf version*** *of the functions below:*

**

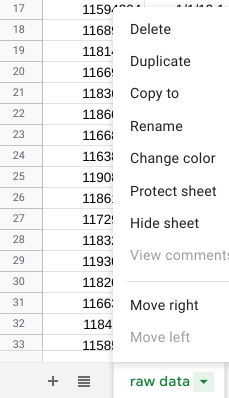
**

[**HANDS-ON ACTIVITY: CREATE A CUSTOM DATA TABLE**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/bB5EY/hands-on-activity-create-a-custom-data-table)

**1- SORT YOUR DATA**

Because you want to answer questions based on a specific timeframe (in this case, applications received per month in 2023), start by sorting the data by date. Sorting involves arranging data into a meaningful order to make it easier to understand, analyze, and visualize. Considering the order in which each application was received can help you discover trends in data analytics job applications.

1. First, rename your spreadsheet. Select Untitled Spreadsheet and enter a new name. Use data\_analyst\_jobs\_2023 or a similar name that clearly describes the data your spreadsheet contains.
2. When you are working in a spreadsheet, you can have multiple sheets open. Currently, your spreadsheet contains one sheet labeled 2023\_data\_analyst\_job. Rename this sheet by selecting the sheet tab and choosing Rename from the menu. Then, enter raw data.



3. Make the columns Job Title (C) and Job Location (D) wider by dragging the right boundary of the column headings.

4. **Select all the data in the spreadsheet by selecting the cell where the rows and columns intersect.**

**5. From the menu bar, select Data > Sort range > Advanced range sorting options.**

**6. In the pop-up window, select the Data has header row box.**

**7. In the Sort by dropdown, choose the header Date. Then, select A to Z to sort in ascending order.**

**8. Finally, select Sort.**

Your spreadsheet now displays job applications received in chronological order.

**2- CREATE A CUSTOM DATA TABLE**

Now that you’ve sorted your data, you’re ready to create a custom data table to help you answer each of your questions. Your table will clearly summarize the data. Plus, if you want to share your results, your table will be well-organized and easy to understand.

Count the number of applications received each month

First, use spreadsheet functions to help you find the total number of applications received in each month.

1. To start, select the Add sheet icon (the plus sign) in the menu bar to add a new sheet to your spreadsheet. You’ll create your data table in this sheet

2. Rename the new sheet. Select the sheet tab and choose Rename on the menu. Then, enter summary data.

3. Next, add column headers to your table. In cell A1 of your summary data sheet, enter Month. In cell B1, enter Applications.

4. Underneath the Month label in cell A2, enter January. Press Enter.

5. Now, use autofill to add the rest of the months of the year. Select cell A2 again. The fill handle will appear in the cell. Select on the fill handle and drag it down to cell A13 to autofill all the months of the year.

6. Next, convert the number values in the Date column in the raw data sheet into text. Select the raw data tab to return to the raw data sheet. In cell G1, enter Month.

7. The **TEXT** function converts a number into text according to a specified format. In this case, list them in which an application was received. Use the format “mmmm” for the full name of the month. In cell G2, enter the following code (do not copy+paste):

**=TEXT(B2,"mmmm")**

The first entry **B2** refers to the cell you want to convert. The second entry **("mmmm")** refers to the specific format you want to use. Press Enter.

(Note: It is *very important* that you manually enter all formulas and functions. They should not be copied and pasted from the activity, as this will result in an error message.)

8. If a box pops up with the option to autofill the column, select the check mark or enter Ctrl + Enter (Windows) or Cmd + Return (Mac). If this box does not pop up, select cell G2. Then, double-click on the fill handle to copy the function down the column. This will populate all the cells in the column with the corresponding month.

9. Now you're ready to total the applications by month. You could do this manually, by filtering the data and counting the number of entries for each month, but this would take a long time and be prone to errors. Instead, use the **COUNTIF** function.

10. The **COUNTIF** function quickly counts how many items in a range of cells meet a given criterion. First, select the summary data tab to return to your summary data sheet. Then, in cell B2, enter **=COUNTIF('raw data'!G:G,A2)**. The first entry **'raw data'!G:G** refers to the range where you are counting the data. The range is located on your raw data sheet **'raw data'!** and includes all column G **G:G** entries. This column contains the data for months. The second entry **A2** refers to the criterion you want to count. In this case, it’s “January,” the value in cell A2 of your summary data sheet. The function calculates how many times January (the criterion) appears in the Month column (the range).

11. Press Enter. You’ll notice the value 2387 appears in cell B2. This means that 2,387 job applications were submitted in January.

12. Select cell B2. Double-click the fill handle to copy the function down through cell B13.

Now your table lists the total number of applications submitted for each month:

| **Month** | **Applications** |
| --- | --- |
|  |  |
| January | 2387 |
| February | 2312 |
| March | 2536 |
| April | 2544 |
| May | 2954 |
| June | 2990 |
| July | 3138 |
| August | 2969 |
| September | 2865 |
| October | 2751 |
| November | 2508 |
| December | 2642 |

**3- FIND THE TOTAL NUMBER OF APPLICATIONS RECEIVED**

Now that you’ve calculated the number of applications received in each month, use spreadsheet formulas to calculate the total number of applications received.

1. Label the cell in which you’ll calculate your result. In cell A14, enter Total.
2. In cell B14, enter **=SUM(B2:B13)**. This function calculates the number of applications received from January through December.
3. Cell B14 contains the total number of applications, 32596.

**4- FIND THE MONTHS WITH THE LOWEST AND HIGHEST NUMBER OF APPLICATIONS RECEIVED**

Use the **MIN** and **MAX** functions to calculate this information.

1. First, make labels for your results. In cell A16, enter Min. In cell A17, enter Max.
2. The **MIN** function returns the minimum value in a numeric range. In cell B16, enter **=MIN(B2:B13)**. The result, 2312, is the lowest number of applications received in any month in 2023.
3. The **MAX** function returns the maximum value in a numeric dataset. In cell B17, enter **=MAX(B2:B13)**. The result, 3138, is the highest number of applications received in any month in 2023.

**5- FIND THE AVERAGE NUMBER OF APPLICATIONS RECEIVED PER MONTH**

Use the **AVERAGE** function to calculate this information.

1. First, make labels for your results. In cell A18, enter Avg.
2. The **AVERAGE** function returns the average value in a numeric dataset. In cell B18, enter **=AVERAGE(B2:B13)**. The result, 2716.33, is the average number of monthly applications received in 2023.

Your work will help your team discover important trends and patterns in the agency’s data and generate insights for optimizing the application process. For example, because your findings reveal that February was the slowest month, the agency can devote more of its advertising and outreach budget to February and less to the peak month of July. This is the strategic impact of data analysis.

**B**[**EST YOUR KNOWLEDGE ON USING FUNCTIONS IN SPREADSHEETS**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/c3vT8/test-your-knowledge-on-using-functions-in-spreadsheets)

**SAVE TIME WITH STRUCTURED THINKING**

[**BEFORE SOLVING A PROBLEM, UNDERSTAND IT**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/F50J1/before-solving-a-problem-understand-it)

**Albert Einstein** once said**," If I were given one hour to save the planet, I would spend 59 minutes defining the problem and one minute resolving it."** Now, that might seem extreme, but it does show us just how important it is to define the problems before trying to solve them.

**STRUCTURED APPROACH IN DEFINING THE PROBLEM DOMAIN**

In the data world, we call this first piece the **problem domain**: the specific area of analysis that encompasses every activity affecting or affected by the problem.

**Jigsaw puzzle example** =>Data analysts face the same kinds of challenges too. You might remember that **data analysts aren't always given the complete picture at the start of a project.** **A big part of their job is to develop a structured approach and use critical thinking to find the best solution.** That starts with understanding the problem domain. This is where **structured thinking** comes into play. To successfully solve a problem as a data analyst, you need to train your brain to think structurally.

[**SCOPE OF WORK AND STRUCTURED THINKING**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/S6uVY/scope-of-work-and-structured-thinking)

**Structured thinking** is the process of recognizing the current problem or situation, organizing available information, revealing gaps and opportunities, and identifying the options. It's having a clear list of what you are expected to deliver, a timeline for major tasks and activities, and checkpoints so the team knows you're making progress. Structured thinking helps us save time and effort, but also makes our job as data analysts easier because it allows us to better understand the work we are doing

**Structured thinking** will help you understand problems at a high level so that you can identify areas that need deeper investigation and understanding.

The starting place for **structured thinking** is the **problem domain** (the specific area of analysis that encompasses every activity affecting or affected by the problem). Once you know the specific area of analysis, you can set your base and lay out all your requirements and hypotheses before you start investigating.

Another way that you can practice **structured thinking** and avoid mistakes is by using a **scope of work**. A **scope of work or SOW** is an agreed-upon outline of the work you're going to perform on a project. For many businesses, this includes things like work details, schedules, and reports that the client can expect.

As a data analyst, your **scope of work** will be a bit more technical and include those basic items we just mentioned, but you'll also focus on things like **data preparation, validation, analysis of quantitative and qualitative datasets, initial results**, and maybe even some **visuals** to really get the point across.

Be sure to differentiate **statement of work** from **scope of work**, which are both abbreviated as **SOW**. Although they help define deliverables and a timeline, they aren't the same and shouldn't be used interchangeably.

A **statement of work** is a document that clearly identifies the products and services a vendor or contractor will provide to an organization. It includes objectives, guidelines, deliverables, schedule, and costs.

A **scope of work** is project-based and sets the expectations and boundaries of a project.

A **scope of work may be included in a statement of work** to help define project outcomes. With a **solid scope of work**, you'll be able to address any confusion, contradictions, or questions about the data up- front and make sure these sneaky setbacks don't stand in your way.

As a junior data analyst, it's more typical to be asked to create a scope of work than a statement of work.

**[CREATING A SCOPE OF WORK](https://www.coursera.org/learn/ask-questions-make-decisions/ungradedWidget/FqQb9/creating-a-scope-of-work)**

[**HANDS-ON ACTIVITY: CREATE A SCOPE OF WORK**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/IEJTe/hands-on-activity-create-a-scope-of-work)

**UNDERSTAND WHAT YOU NEED TO KNOW ABOUT AN SOW**

As a data analyst, it’s hard to overstate the importance of an SOW document. A well-defined SOW keeps you, your team, and everyone involved with a project on the same page. It ensures that all contributors, sponsors, and stakeholders share the same understanding of the relevant details.

### **Why do you need an SOW?**

****

The point of data analysis projects is to complete business tasks that are useful to the stakeholders. Creating an SOW helps to make sure that everyone involved, from analysts and engineers to managers and stakeholders, shares the understanding of what those business goals are, and the plan for accomplishing them.

**Clarifying requirements and setting expectations** are two of the most important parts of a project. Recall the first phase of the Data Analysis Process—asking questions.

As you ask more and more questions to clarify requirements, goals, data sources, stakeholders, and any other relevant info, an SOW helps you formalize it all by recording all the answers and details. In this context, the word “ask” means two things. Preparing to write an SOW **is about asking questions to learn the necessary information about the project**, but **it’s also about clarifying and defining what you’re being asked to accomplish, and what the limits or boundaries of the “ask” are**. After all, if you can’t make a distinction between the business questions you are and aren’t responsible for answering, then it’s hard to know what success means!

### **What is a good SOW?**

****

**There’s no standard format for an SOW.** They may differ significantly from one organization to another, or from project to project. **However, they all have a few foundational pieces of content in common.**

* **Deliverables:** What work is being done, and what things are being created as a result of this project? When the project is complete, what are you expected to deliver to the stakeholders? Be specific here. Will you collect data for this project? How much, or for how long?

**Avoid vague statements.** For example, “fixing traffic problems” doesn’t specify the scope. This could mean anything from filling in a few potholes to building a new overpass. **Be specific! Use numbers and aim for hard, measurable goals and objectives.** For example: “Identify top 10 issues with traffic patterns within the city limits, and identify the top 3 solutions that are most cost-effective for reducing traffic congestion.”

* **Milestones:** **This is closely related to your timeline.** What are the major milestones for progress in your project? How do you know when a given part of the project is considered complete?

**Milestones** can be identified by you, by stakeholders, or by other team members such as the Project Manager. Smaller examples might include incremental steps in a larger project like “Collect and process 50% of required data (100 survey responses)”, but may also be larger examples like ”complete initial data analysis report” or “deliver completed dashboard visualizations and analysis reports to stakeholders”.

* **Timeline: Your timeline will be closely tied to the milestones you create for your project.** The **timeline** is a way of mapping expectations for how long each step of the process should take. The **timeline** should be specific enough to help all involved decide if a project is on schedule. When will the deliverables be completed? How long do you expect the project will take to complete? If all goes as planned, how long do you expect each component of the project will take? When can we expect to reach each milestone?
* **Reports: Good SOWs also set boundaries for how and when you’ll give status updates to stakeholders.** How will you communicate progress with stakeholders and sponsors, and how often? Will progress be reported weekly? Monthly? When milestones are completed? What information will status reports contain?

**At a minimum, any SOW should answer all the relevant questions in the above areas.** Note that these areas may differ depending on the project. But at their core, the SOW document should always serve the same purpose by containing information that is specific, relevant, and accurate. If something changes in the project, your SOW should reflect those changes.

### 

### 

### 

### **What is in and out of scope?**

****

SOWs should also contain information specific to what is and isn’t considered part of the project. The scope of your project is everything that you are expected to complete or accomplish, defined to a level of detail that doesn’t leave any ambiguity or confusion about whether a given task or item is part of the project or not.

Notice how the previous example about studying traffic congestion defined its scope as the area within the city limits. This doesn’t leave any room for confusion — stakeholders need only to refer to a map to tell if a stretch of road or intersection is part of the project or not. Defining requirements can be trickier than it sounds, so it’s important to be as specific as possible in these documents, and to use quantitative statements whenever possible.

For example, assume that you’re assigned to a project that involves studying the environmental effects of climate change on the coastline of a city: How do you define what parts of the coastline you are responsible for studying, and which parts you are not?

In this case, it would be important to define the area you’re expected to study using GPS locations, or landmarks. Using specific, quantifiable statements will help ensure that everyone has a clear understanding of what’s expected.

[**STAYING OBJECTIVE**](https://www.coursera.org/learn/ask-questions-make-decisions/lecture/YQw5K/staying-objective)

**Data** doesn't live in a vacuum, **it needs context**. Earlier, we learnt that context is the condition in which something exists or happens. Actions can be appropriate in some context, but inappropriate in others, for example, yelling is rude in one context, if your friend is standing in front of the TV, but it's entirely appropriate in another, if that friend is about to get hit by a kid on a tricycle. Do you see the difference? In the world of data, numbers don't mean much without context.

To really understand what the data is about, you have to think through who, what, where, when, how and why. It's good to ask yourself questions like, who collected the data? And what is it about? What does the data represent in the world, and how does it relate to other data? When was the data collected? Data collected awhile ago may have certain limitations, given the present day situation.

[**THE IMPORTANCE OF CONTEXT**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/VijfZ/the-importance-of-context)

# 

# **Context** in data analytics is the condition and circumstances that surround and give meaning to the data. Context is important in data analytics because it helps make disorganized data accessible and understood. The fact is, data has little value if it is not paired with context.



Understanding the context behind the data can help us make it more meaningful at every stage of the data analysis process. For example, you might be able to make a few guesses about what you're looking at in the following table, but you couldn't be certain without more context.

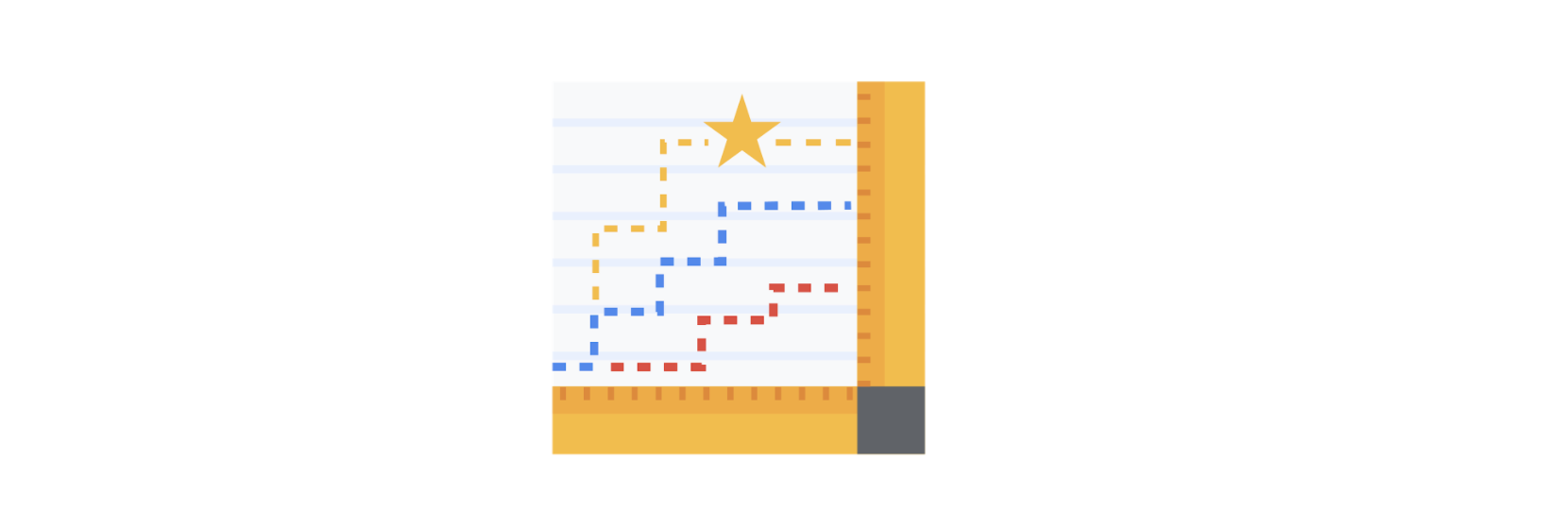
| 2010 | 28000 |
| --- | --- |
| 2005 | 18000 |
| 2000 | 23000 |
| 1995 | 10000 |

On the other hand, if the first column was labeled to represent the years when a survey was conducted, and the second column showed the number of people who responded to that survey, then the table would start to make a lot more sense. Take this a step further, and you might notice that the survey is conducted every 5 years. This added context helps you understand why there are five-year gaps in the table.

| **Years (Collected every 5 years)** | **Respondents** |
| --- | --- |
| 2010 | 28000 |
| 2005 | 18000 |
| 2000 | 23000 |
| 1995 | 10000 |

**Context can turn raw data into meaningful information.** It is very important for data analysts to contextualize their data. This means giving the data perspective by defining it. To do this, you need to identify:

* Who: The person or organization that created, collected, and/or funded the data collection
* What: The things in the world that data could have an impact on
* Where: The origin of the data
* When: The time when the data was created or collected
* Why: The motivation behind the creation or collection
* How: The method used to create or collect it



Understanding and including the context is important during each step of your analysis process, so it is a good idea to get comfortable with it early in your career. For example, when you collect data, you’ll also want to ask questions about the context to make sure that you understand the business and business process. During organization, the context is important for your naming conventions, how you choose to show relationships between variables, and what you choose to keep or leave out. And finally, when you present, it is important to include contextual information so that your stakeholders understand your analysis.

**[SELF-REFLECTION: WORK WITH DATASETS](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/rSHf2/self-reflection-work-with-datasets)**

[**TEST YOUR KNOWLEDGE ON STRUCTURED THINKING**](https://www.coursera.org/learn/ask-questions-make-decisions/quiz/VA5ym/test-your-knowledge-on-structured-thinking)

**MODULE 3 CHALLENGE**

[**GLOSSARY TERMS FROM MODULE 3**](https://www.coursera.org/learn/ask-questions-make-decisions/supplement/aZ5ss/glossary-terms-from-module-3)

## Terms and definitions for Course 2, Module 3

AVERAGE: A spreadsheet function that returns an average of the values from a selected range

Borders: Lines that can be added around two or more cells on a spreadsheet

Cell reference: A cell or a range of cells in a worksheet typically used in formulas and functions

COUNT: A spreadsheet function that counts the number of cells in a range that meet a specific criteria

Equation: A calculation that involves addition, subtraction, multiplication, or division (also called a math expression)

Fill handle: A box in the lower-right-hand corner of a selected spreadsheet cell that can be dragged through neighboring cells in order to continue an instruction

Filtering: The process of showing only the data that meets a specified criteria while hiding the rest

Header: The first row in a spreadsheet that labels the type of data in each column

Math expression: A calculation that involves addition, subtraction, multiplication, or division (also called an equation)

Math function: A function that is used as part of a mathematical formula

MAX: A spreadsheet function that returns the largest numeric value from a range of cells

MIN: A spreadsheet function that returns the smallest numeric value from a range of cells

Open data: Data that is available to the public

Operator: A symbol that names the operation or calculation to be performed

Order of operations: Using parentheses to group together spreadsheet values in order to clarify the order in which operations should be performed

Problem domain: The area of analysis that encompasses every activity affecting or affected by a problem

Range: A collection of two or more cells in a spreadsheet

Report: A static collection of data periodically given to stakeholders

Return on investment (ROI): A formula that uses the metrics of investment and profit to evaluate the success of an investment

Revenue: The total amount of income generated by the sale of goods or services

Scope of work (SOW): An agreed-upon outline of the tasks to be performed during a project

Sorting: The process of arranging data into a meaningful order to make it easier to understand, analyze, and visualize

SUM: A spreadsheet function that adds the values of a selected range of cells

**[MODULE 3 CHALLENGE](https://www.coursera.org/learn/ask-questions-make-decisions/exam/Oocv7/module-3-challenge)**