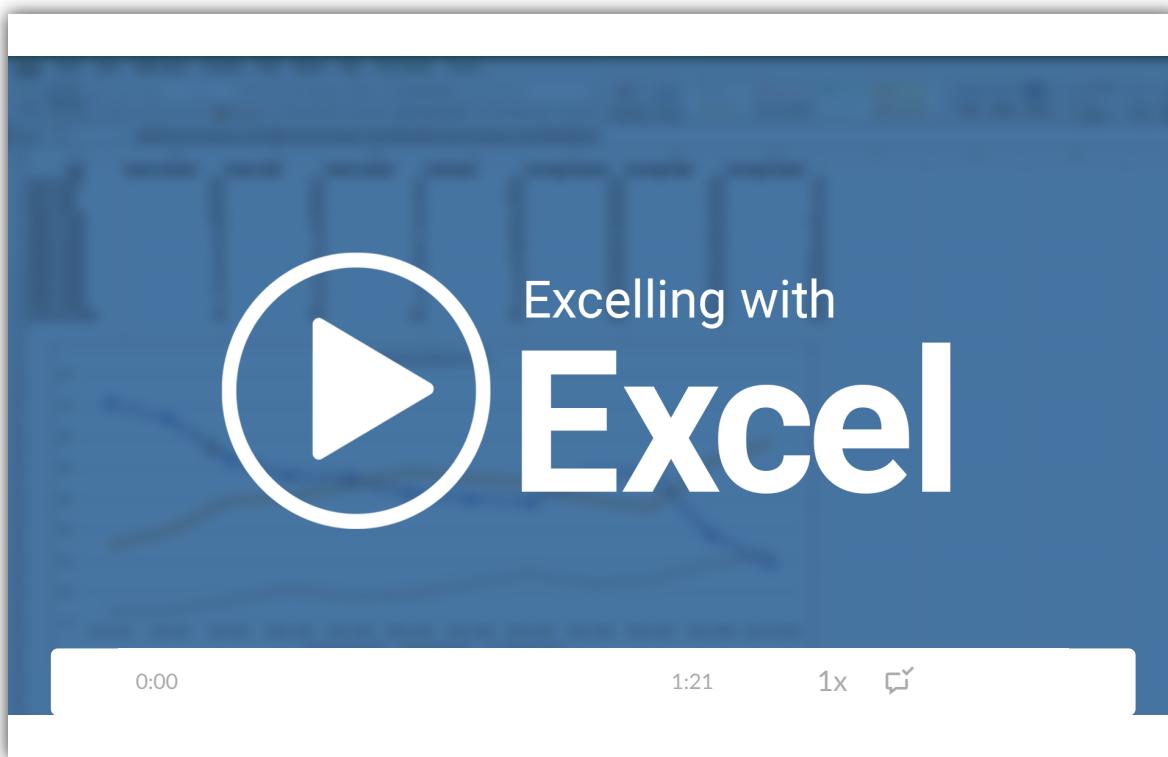


### 1.0.1 **Excelling with Excel**

Welcome to Module 1! We're excited to get started, but before we dive in, watch this video to learn more about how Excel fits into the data landscape.

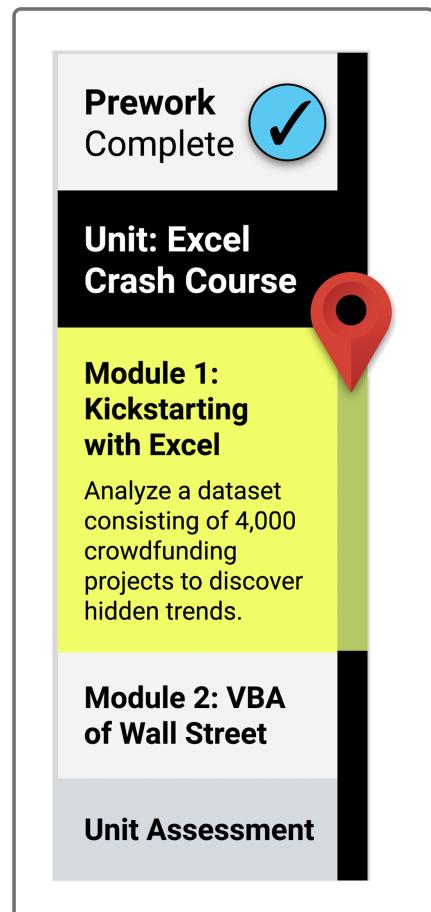


## 1.0.2 Module 1 Roadmap

### Looking Ahead

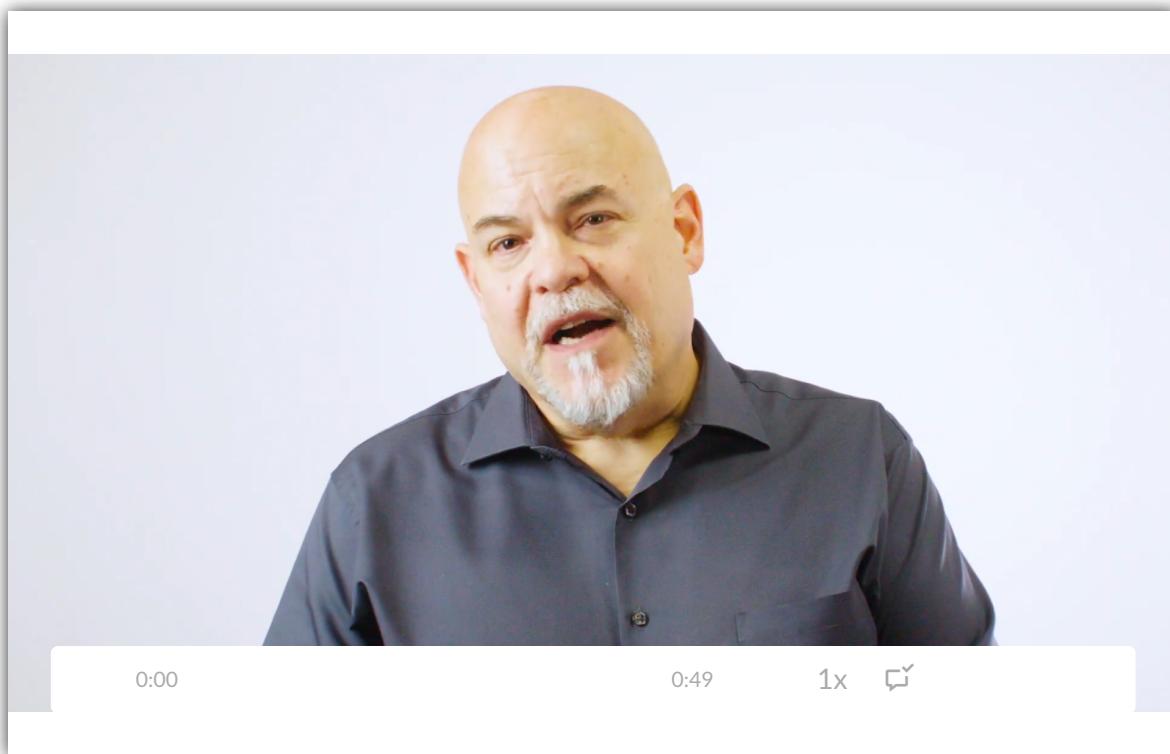
This module is a deep dive into Excel, a tool that can be used across all areas of life, from household budgeting to complex financial analysis. Learning the intricacies of Excel will draw on (and enhance) skills you may have already, like computer literacy, data literacy, and quantitative reasoning.

Many of the advanced Excel features you'll learn in this module—including formulas, charts, and pivot tables—set the stage for the data visualization languages and tools you'll encounter later in this course. Using Excel in the context of data analytics will help you start to think about data differently. What is data? How does it tell a story? How can data be used to answer specific questions, as well as uncover trends and patterns? Starting today and throughout the rest of the course, you'll learn how to find answers to these questions.



### 1.0.3 Kickstarting with Excel

Each module in this course is built around a project that mirrors a real-world scenario that would require data analysis and visualization. Watch the following video to learn about the background of the project you will be working on.



---

## What You Will Learn

By the end of this module, you should be able to:

- Import data into a table for analysis.
  - Apply filters, conditional formatting, and formulas.
  - Generate and interpret pivot tables.
  - Calculate summary statistics such as measures of central tendency, standard deviation, and variance.
  - Characterize data to identify outliers in datasets.
  - Perform an Excel analysis with visualizations.
- 

## Planning Your Schedule

Here's a quick look at the lessons and assignments you'll cover in this module. You can use the time estimates to help pace your learning and plan your schedule.

- Introduction (15 minutes)
- Lesson 1: Getting Started with Excel (15 minutes)
- Lesson 2: Viewing the Data (1 hour 15 minutes)
- Lesson 3: Using Pivot Tables and Pivot Charts (1 hour 30 minutes)
- Lesson 4: Using Filters and Searches to Deepen Your Analysis (30 minutes)
- Lesson 5: Applying Descriptive Statistics and Visualization (1 hour)
- Lesson 6: Sharing Your Findings (30 minutes)
- Application (5 hours)

### 1.1.1 Download and Install Excel

To help Louise, we need to have the right tools ready to go. Let's get started by installing Excel—if you don't have it already—and checking for updates.

#### Download Excel 2016 or Later

In preparation for the upcoming data work, we first need to be sure Excel is installed and ready to roll. It's best to use a recent version of Excel, namely, Excel 2016 or later.

If you are unable to purchase Excel 2016 or later, a free trial of the latest version is available. [This trial](https://products.office.com/en-us/excel) (<https://products.office.com/en-us/excel>) will give you access to the latest version of the software for one month.

#### CAUTION

It's important that you use Excel instead of Google Sheets. While Google Sheets has many capabilities similar to Excel (they even look the same!), Excel has more processing power and provides users an extended number of formulas and charts.

## First-Time Download

If you are downloading the program for the first time, refer to the [official Microsoft documentation](https://support.office.com/en-us/article/download-and-install-or-reinstall-office-365-or-office-2019-on-a-pc-or-mac-4414eaaf-0478-48be-9c42-23adc4716658) (<https://support.office.com/en-us/article/download-and-install-or-reinstall-office-365-or-office-2019-on-a-pc-or-mac-4414eaaf-0478-48be-9c42-23adc4716658>) for installation instructions.

## Update Excel 2016 or Later

If you already have Excel 2016 (or later) on your computer, open the program to check for updates to make sure you have all the latest features and capabilities. This will also update the software to address and fix any minor problems you may encounter.

Follow the directions below that correspond to your operating system.

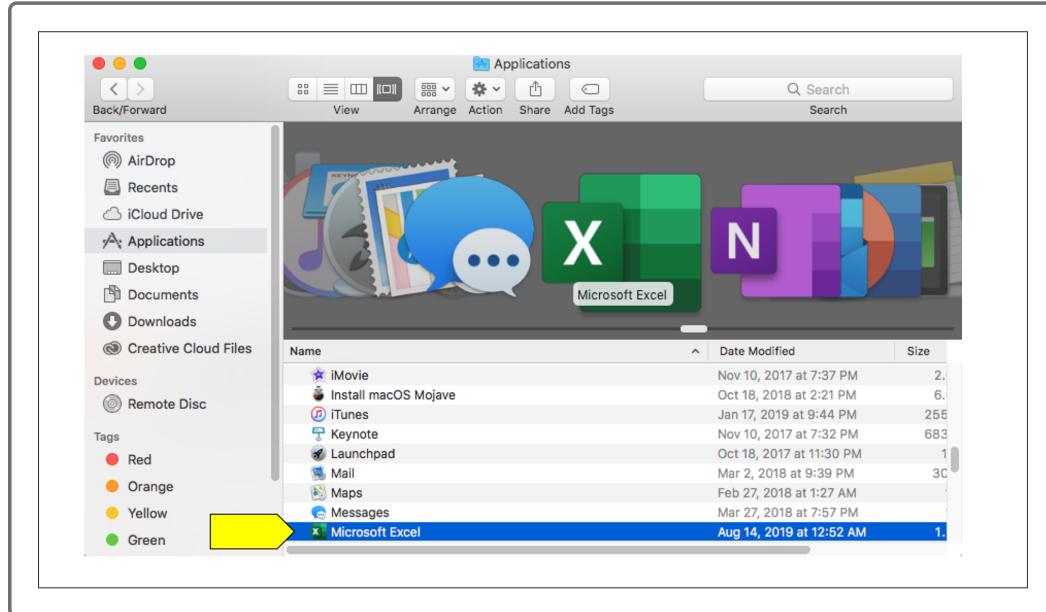
### macOS

Locate Excel by following these steps:

1. Click the Launchpad icon in your Dock.

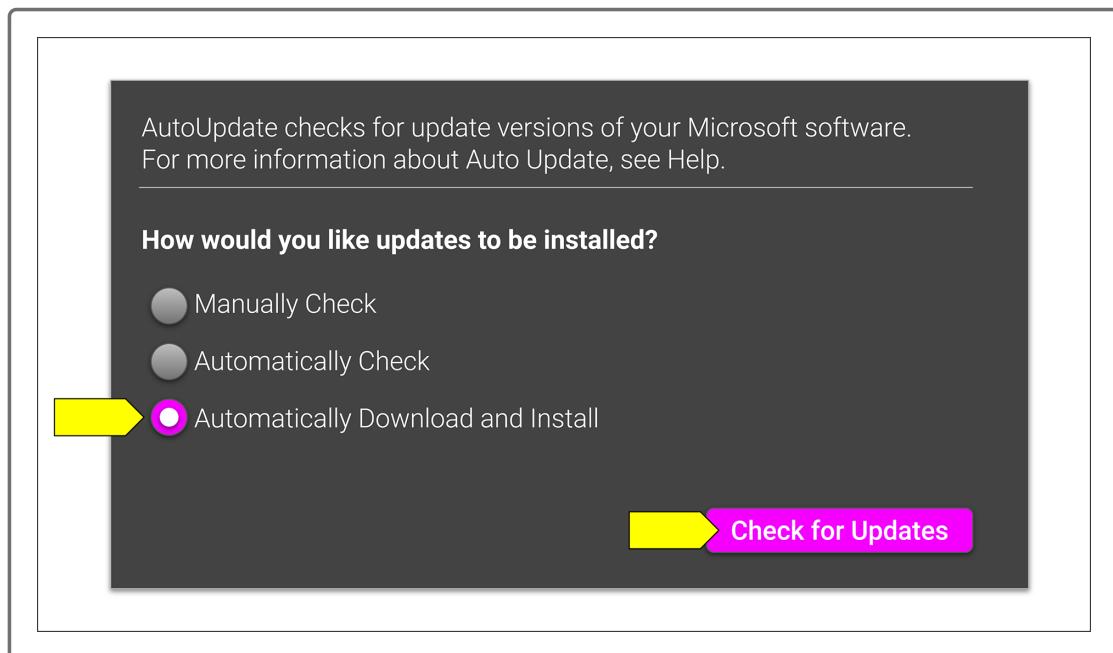


2. Scroll through the app icons until you find Microsoft Excel, and then click the icon.



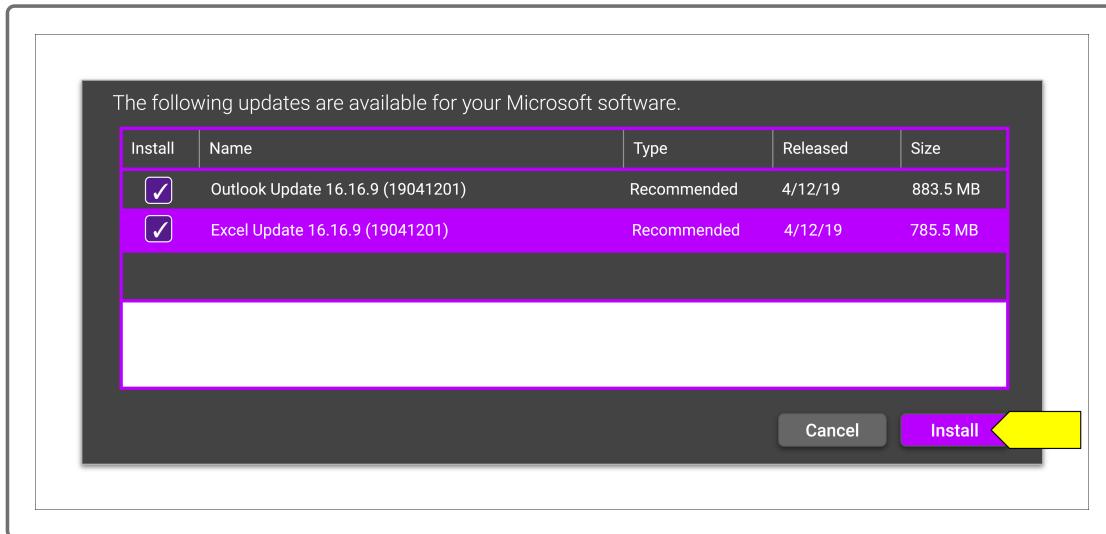
## Updating Excel on macOS

Navigate to the Help menu at the top of the screen, and then select “Check for Updates.” You will be automatically directed to an installation window asking how you would like the updates to be installed.



Select “Automatically Download and Install,” and then click the “Check for Updates” button in the bottom-right corner.

The next window shows which update you'll be installing. Make sure the Excel update is checked, and then click Install.



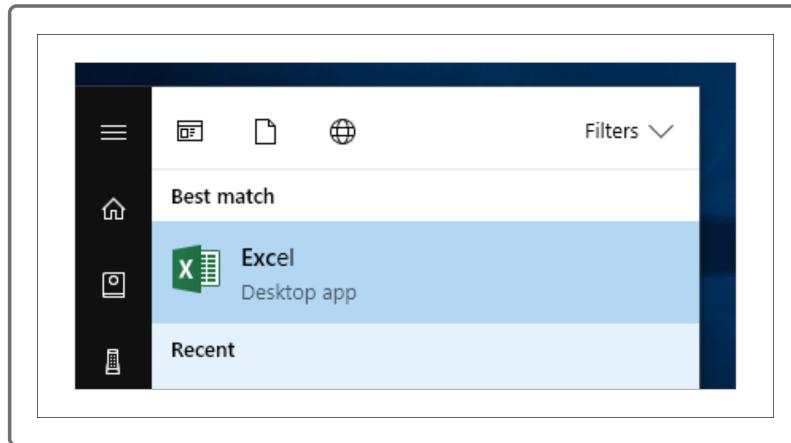
#### NOTE

More than one Microsoft Office product may be available to update. There's no harm in updating everything, but the updates are usually large and can take awhile to complete. If you want to get started with Excel quickly, then update only Excel and return to update the other programs when you're ready.

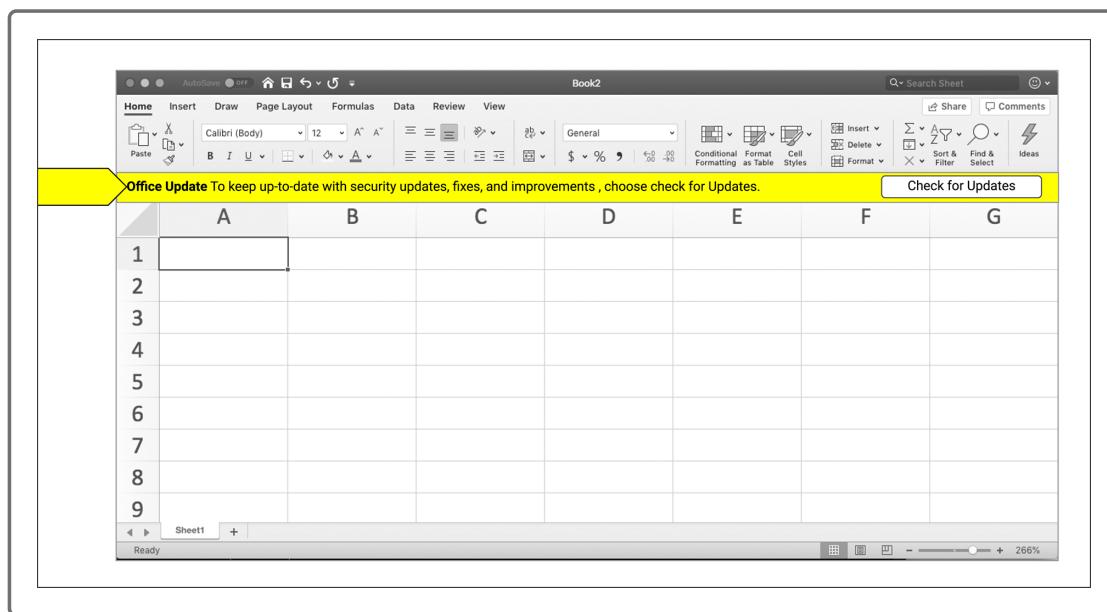
## Windows

Excel can be found by following these steps:

1. Open the Start menu by clicking the Windows icon in the lower-left corner of your screen, or by pressing the Windows key (between CTRL and ALT on the left side of your keyboard).
2. Type "Excel" in the search bar. You should see the Excel desktop app pop up as a suggestion. Click on the icon to open Excel.



If this is your first time opening Excel in a while, there's a good chance you'll need to update the program. Thankfully, Excel makes the update process easier by providing a yellow banner and a "Check for Updates" button at the top of the window.



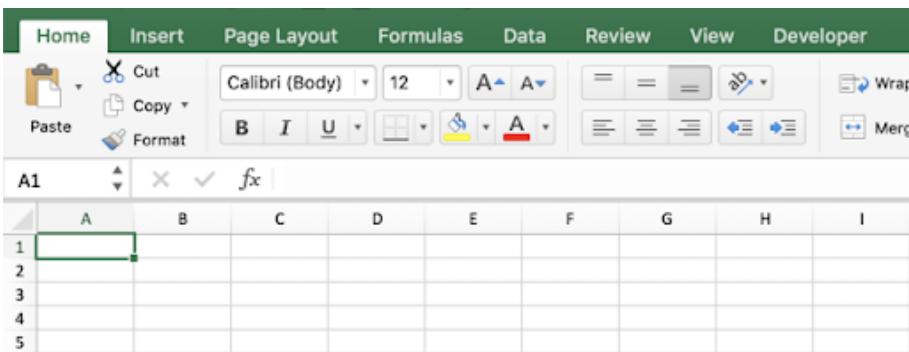
When you click the “Check for Updates” button, Excel automatically connects to the internet and searches for the most recent version of the software. Depending on how much time has passed since the previous update, there may be multiple updates to install.

## Updating Excel on Windows

On Windows, Excel automatically checks for updates when you open the program. If there's an update available, you will see a yellow banner along the top of the screen prompting you to install the update. Read the [official documentation](https://support.office.com/en-ie/article/update-office-with-microsoft-update-f59d3f9d-bd5d-4d3b-a08e-1dd659cf5282) (<https://support.office.com/en-ie/article/update-office-with-microsoft-update-f59d3f9d-bd5d-4d3b-a08e-1dd659cf5282>) with steps to update your software.

When your Excel software is ready for use, the “Check for Updates” banner will disappear.

Does an update banner appear in your Excel window?



The screenshot shows the Microsoft Excel ribbon with the Home tab selected. Below the ribbon is a standard Excel toolbar with icons for Cut, Copy, Paste, and Format. The worksheet area shows a grid from A1 to I5. Cell A1 is currently selected and contains a small green square handle indicating it is selected. The formula bar above the grid shows cell A1 and the fx button.

Yes  
 No

Check Answer

Finish ►

### 1.1.2 Set Up Your Folder Structure

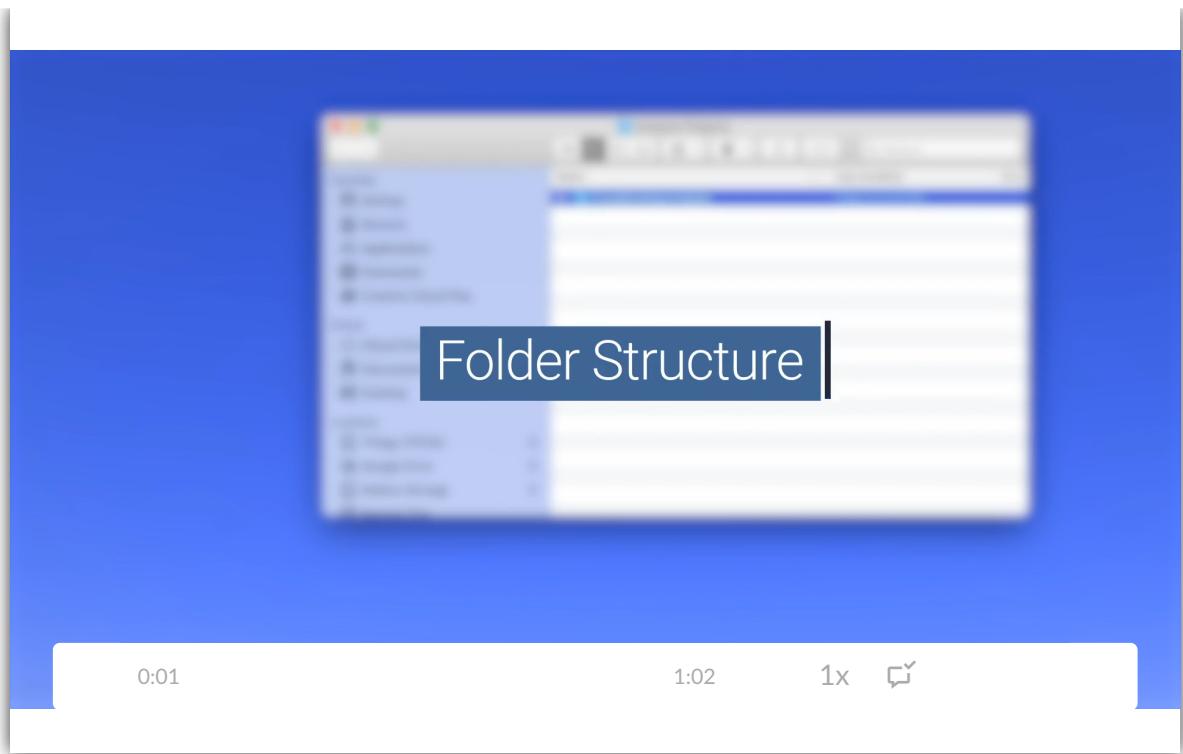
Now that we have downloaded and updated Excel, let's think about how to organize the files we'll need for our project. An organized folder structure will allow us to find the information we need more quickly.

The next step in the setup process is to create an organized folder structure that will allow us to easily locate our projects and files as we work. By staying organized in this way, we can avoid searching for files as well as potentially losing or duplicating work.

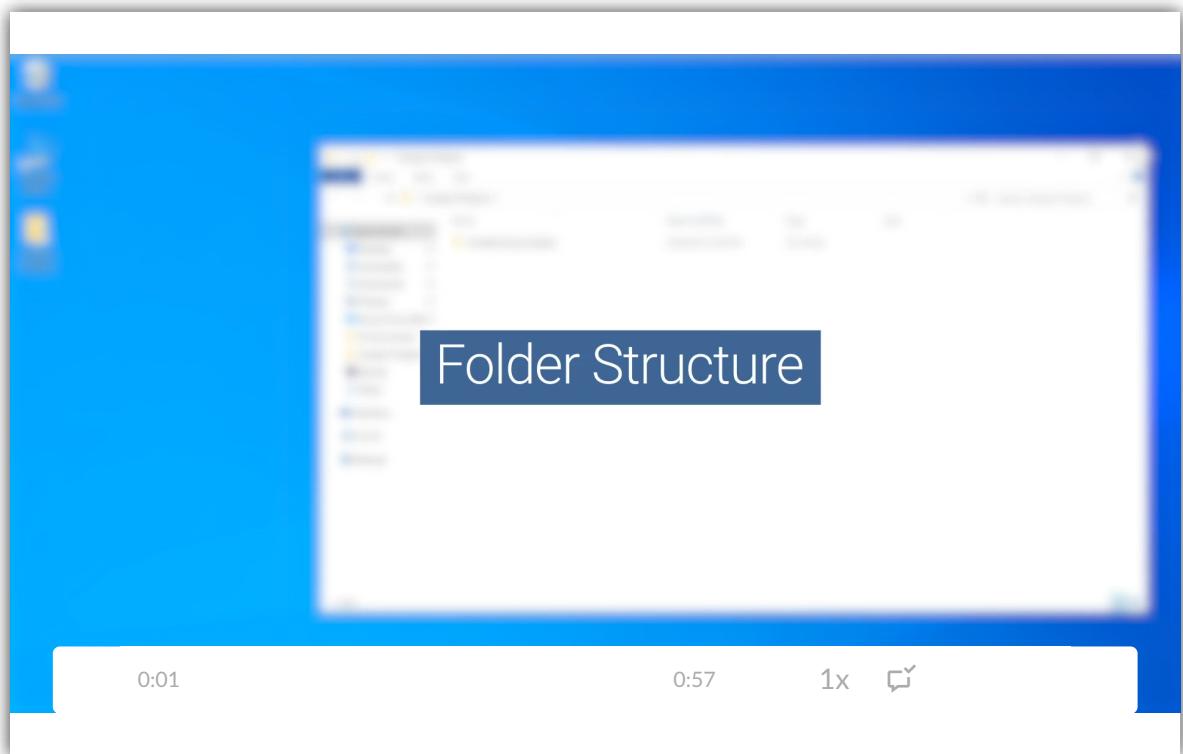
Let's create a folder on our desktop to hold data and other project files. Choose a name that is appropriate for the type of work you'll be saving there, such as Classwork or Analysis Projects. We'll refer to this folder as the "class folder" from here on out. This will serve as our master folder, a central location for course projects. Within this folder, we'll create subfolders for individual projects.

The following video walks you through the steps for setting up the master project folder and project subfolders.

**macOS**



## Windows



### 1.1.3

## Download the Kickstarter Data

Now it's time to retrieve the Kickstarter data we need to help Louise with her project campaign. Then we'll make the data more readable and searchable with the help of Excel tools.

Let's get the Kickstarter data! Click the link below to download the dataset. Once downloaded, the data will be a **local file** that is accessible from your computer, which is necessary in order to work on it in Excel.

### [Download the dataset](#)

([https://courses.bootcampspot.com/courses/138/files/16126/download?  
wrap=1](https://courses.bootcampspot.com/courses/138/files/16126/download?wrap=1)) 

([https://courses.bootcampspot.com/courses/138/files/16126/download?  
wrap=1](https://courses.bootcampspot.com/courses/138/files/16126/download?wrap=1))

### NOTE

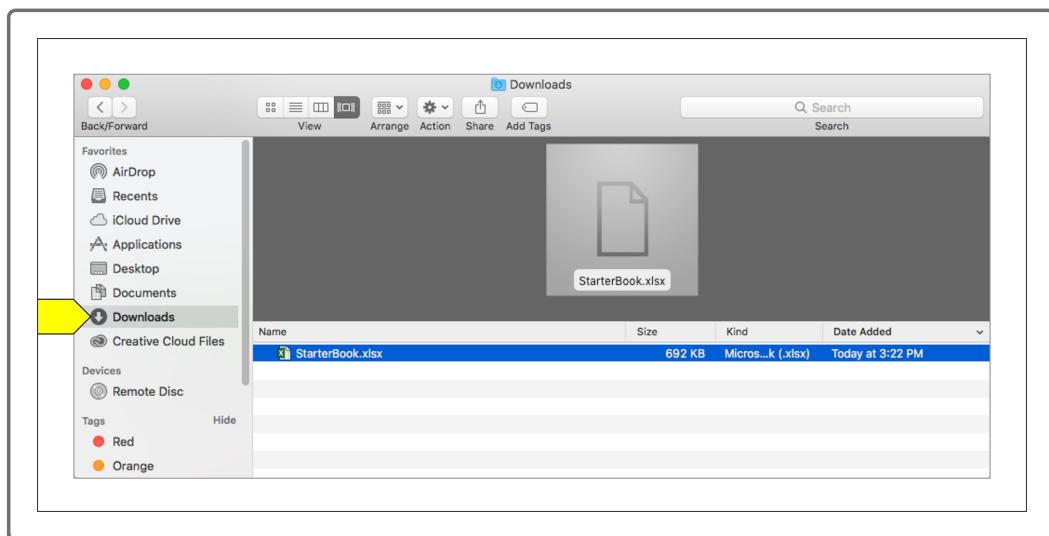
A **hosted file** is a file that is stored online, often with the intention to share via download.

A **local file** is one that is physically stored on your computer. Our file needs to be stored locally so that Excel can open and interact with it.

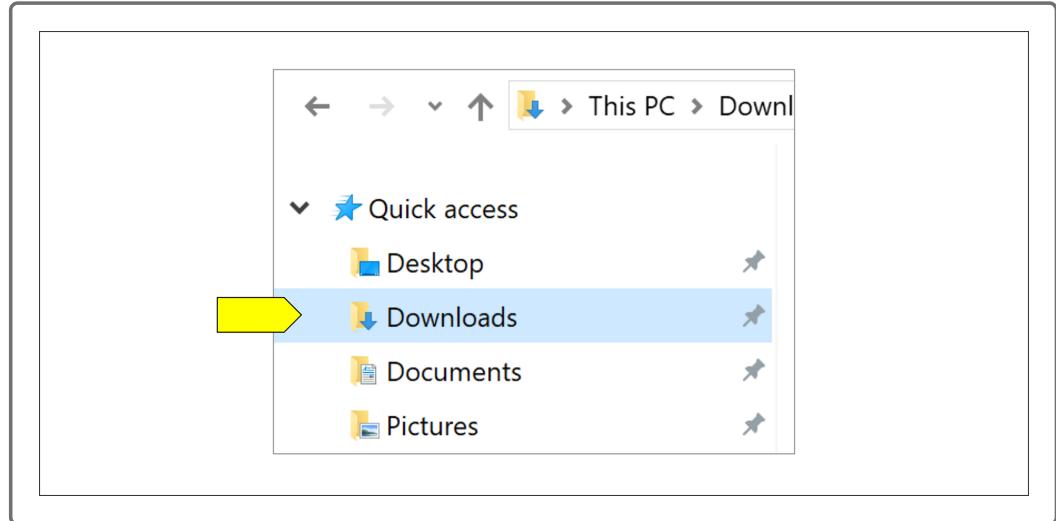
By default, downloaded files are saved to your computer's Downloads directory. Since we have a folder set up specifically for this dataset and our analysis, let's move the downloaded file to that folder.

1. Open the Downloads folder.

(macOS) Click the Finder icon, and then select the Downloads directory.



(Windows) Open a new File Explorer window, and select Downloads from the Quick Access menu on the left-hand side.



2. Left-click and hold your cursor on the data file to drag it from the Downloads folder to the Crowdfunding Analysis subfolder. (Later on in this module, we'll also call this folder "project folder.")

Now let's open the file to see what we're working with by double-clicking it. The file will automatically open in Excel because it has an Excel-specific file extension (.xlsx).

1.2.1

## Get Situated in the Big Picture

We have Excel up and running, set up our project folders, and downloaded the data file. But before we can dive in and analyze the data for Louise, we need to familiarize ourselves with the software.

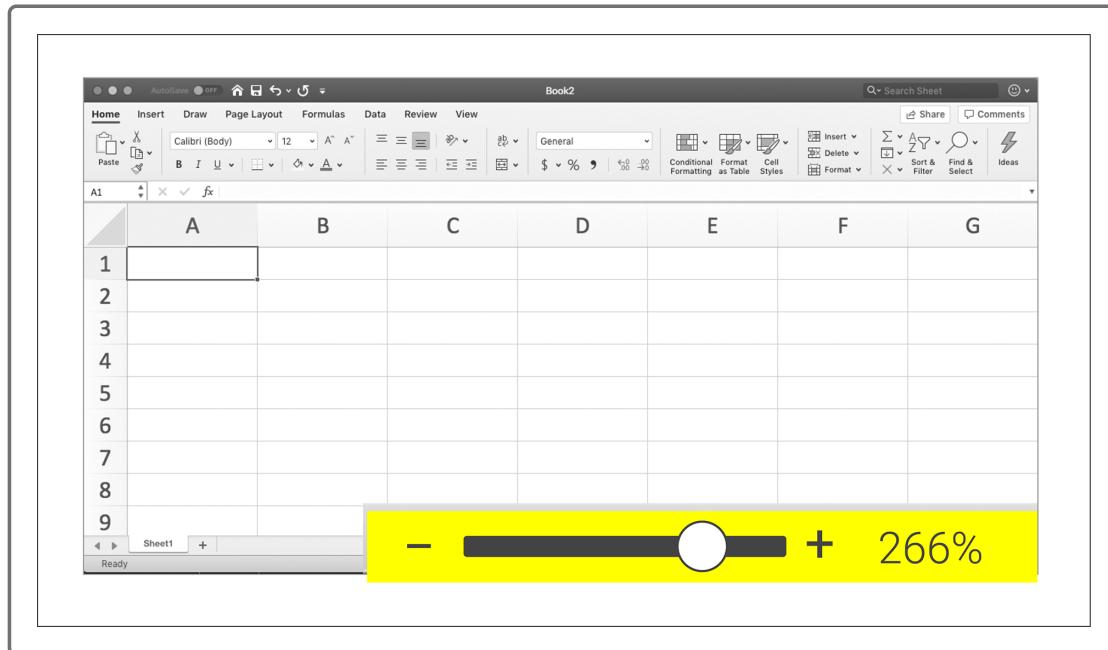
We'll be performing data analysis on several thousand crowdfunding projects to uncover any hidden trends. To do this, we'll need to get acquainted with Excel a bit more. Specifically, we need to know how to view and read data in an Excel file. Let's discuss some of the visual features of Excel.

---

## Resize the Data

When you open the data file, one thing you'll probably notice right away is the small text size. Because there are so many columns, Excel automatically resizes the sheet to fit everything on one screen, but that often makes the text too small to read. Let's adjust the zoom to make the text legible.

To do this, look at the bottom right of the screen, where you'll see a bar with a slider—a negative (minus) sign on the left and a positive (plus) sign on the right. To the right of the plus sign is a percentage.



This percentage indicates the magnification of the text. The text can be enlarged by dragging the small white circle to the right, or by clicking the “+” sign until the text is a comfortable size.

## Tabular Data

When you open the data file in Excel, what you're looking at is a **worksheet**, also referred to as a sheet. At a glance, we can see that the data is arranged in rows and columns. Data in this format is called **tabular data**.

|   | A  | B   | C   | D     | E       | F          |
|---|----|---|---|-------|---------|------------|
| 1 | id | name  | blurb   | goal  | pledged | state      |
| 2 | 0  | GIRLS STATE a new musical comedy TV project   | In this new TV show "All Politics is Vocal" as high school girls campaign, sing and cheer to be elected Governor of their summer camp.  | 8500  | 11633   | successful |
| 3 | 1  | FannibalFest Fan Convention   | A Hannibal TV Show Fan Convention and Art Collective  | 10275 | 14653   | successful |
| 4 | 2  | Charlie teaser completion<br>Unsure/Positive: A Dramedy Series About Life<br>3 with HIV | Completion fund for post-production for teaser of British crime/drama tv series about a girl who sells morals for money<br><br>We already produced the *very* beginning of this story. Help us to see it through? | 500   | 525     | successful |
| 5 |    |   |   | 10000 | 10390   | successful |

## IMPORTANT

**Tabular data** is data that is displayed in a column and row format. This format isn't limited to Excel spreadsheets; any data displayed as a table is considered to be tabular. This includes digital tables on a website and printed tables in a textbook.

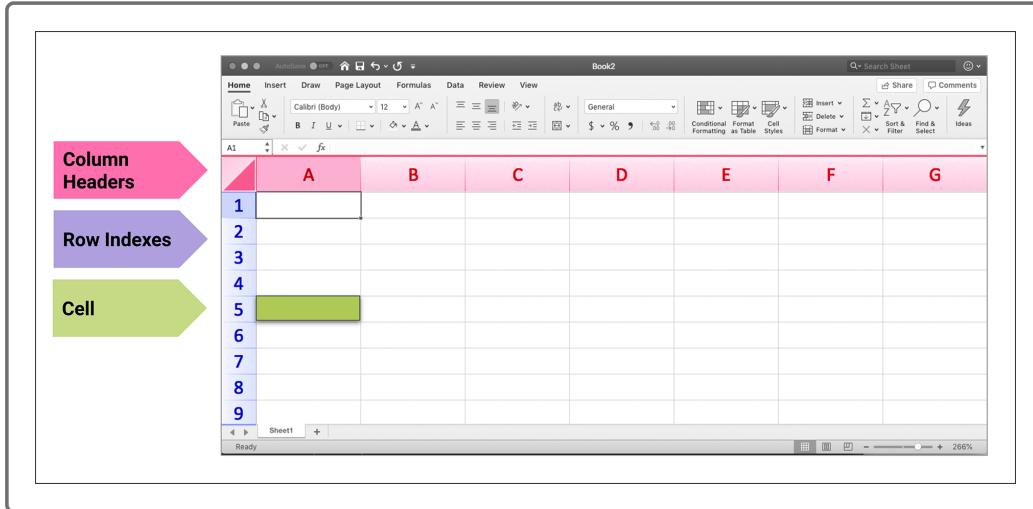
## Anatomy of a Worksheet

You have a worksheet of tabular data that you have resized to make it readable. Now let's break down the worksheet a bit more to understand what it's comprised of.

## Headers and Indexes

The letters along the top of the columns represent the **column headers**, and the numbers to the left of the rows are the **row indexes**. The headers and indexes help us identify where each data point is located.

In the following image, the cell is located at the intersection of column B, row 3, so we refer to it as B3.



B19 is a cell reference pointing to the cell located at the intersection of the second column, B, and the 19th row. What value appears in B19?

- ArtMoose TV Series
- Brouhaha (an Original Sitcom)
- 17
- Humble Pie

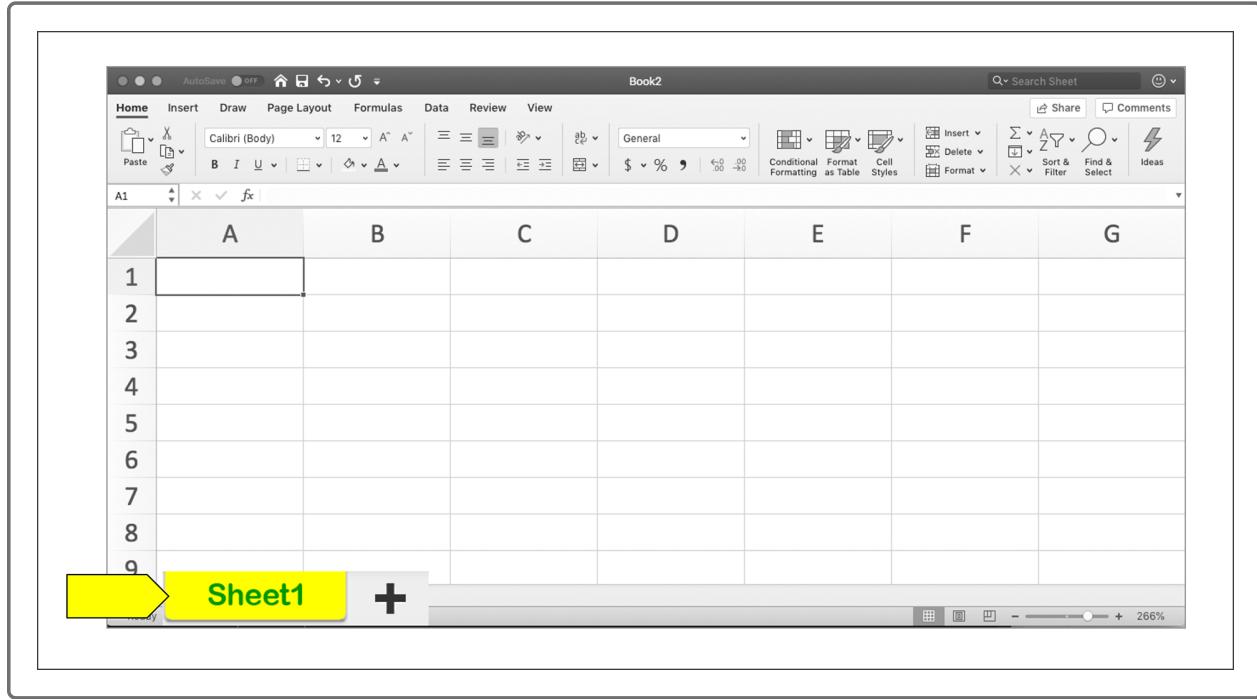
Response-specific feedback:

Correct. Humble Pie appears in B19.

[Finish ►](#)

## Worksheet Names

At the bottom of the sheet is a tab labeled "Sheet1." This is our current, or active, worksheet.



To create a new worksheet, click the plus sign (+). When multiple sheets are being used, the left and right arrows allow us to navigate between them. To help us more easily identify our data, let's rename "Sheet1" to "Kickstarter" by following these steps:

1. Right-click the sheet name (currently Sheet1).
2. Select Rename from the pop-up menu.
3. With the current name highlighted, type the new name and press Enter.

## 1.2.2 Initial Look at the Data

We're almost ready to start analyzing the Kickstarter data file! The first step of an analysis is to take an initial look at the data. The goal of this step is essentially to size up the data, familiarize yourself with it, and get a sense of what you're working with. Let's do that now.

How do you approach a new experience? For instance, think about when you're eating in a restaurant for the first time. You probably peruse the menu to familiarize yourself with the food and drink options. Or consider what you do when you're in a new city. You probably take some time to survey your surroundings, taking in the sights and sounds of your new environment.

A similar process needs to take place for data analysts when they look at a dataset for the first time. They need to size it up and get a feel for what they'll be working with, which is exactly what you'll do next with the Kickstarter data.

Here are a few things to keep in mind during this initial review:

- How many columns and rows are there?

- What types of data are present?
- Is the data readable, or does it need to be converted in some way?

Let's investigate each of these questions more thoroughly.

---

## How Many Columns and Rows Are There?

Using a mouse or laptop touchpad, scroll left and right and up and down to get a feel for how large the worksheet is. That's a lot of scrolling, right? Because of its size, manually scrolling through the worksheet will take some time; so instead we'll use a few shortcuts to get an idea of how many rows and columns we're working with.

To quickly view the last column containing data, press the Command and right arrow keys (Mac), or the CTRL and right arrow keys (Windows). This will bring the cursor to the final column. Each column's header is a letter of the alphabet, which keeps track of how many columns are in place.

How many data columns are there in this worksheet?

14 ✓

4,115

12

15

Response-specific feedback:  
Correct. Columns A–N contain data.

Finish ►

The length of a spreadsheet can be viewed using a similar shortcut: press the Command and down arrow keys (Mac), or the CTRL and down arrow keys (Windows), to bring the cursor to the very last row of data. The

number to the left tells us how many rows of data are in the worksheet. Note that there are many keyboard shortcuts like these that can be used in Excel as well as other programs.

How many rows are in the worksheet?

- 4,115 ✓
- 14
- 4,113
- 4,000

Response-specific feedback:  
Correct. There is a lot of data to work with!

[Finish ▶](#)

Smaller datasets (for example, a dataset with 5 columns and 10 rows) can be manipulated manually, which makes manual calculations within the worksheet a more doable task. However, since we are working with a substantial dataset, we'll need to use built-in formulas and formatting to perform our calculations, which we'll get to later on.

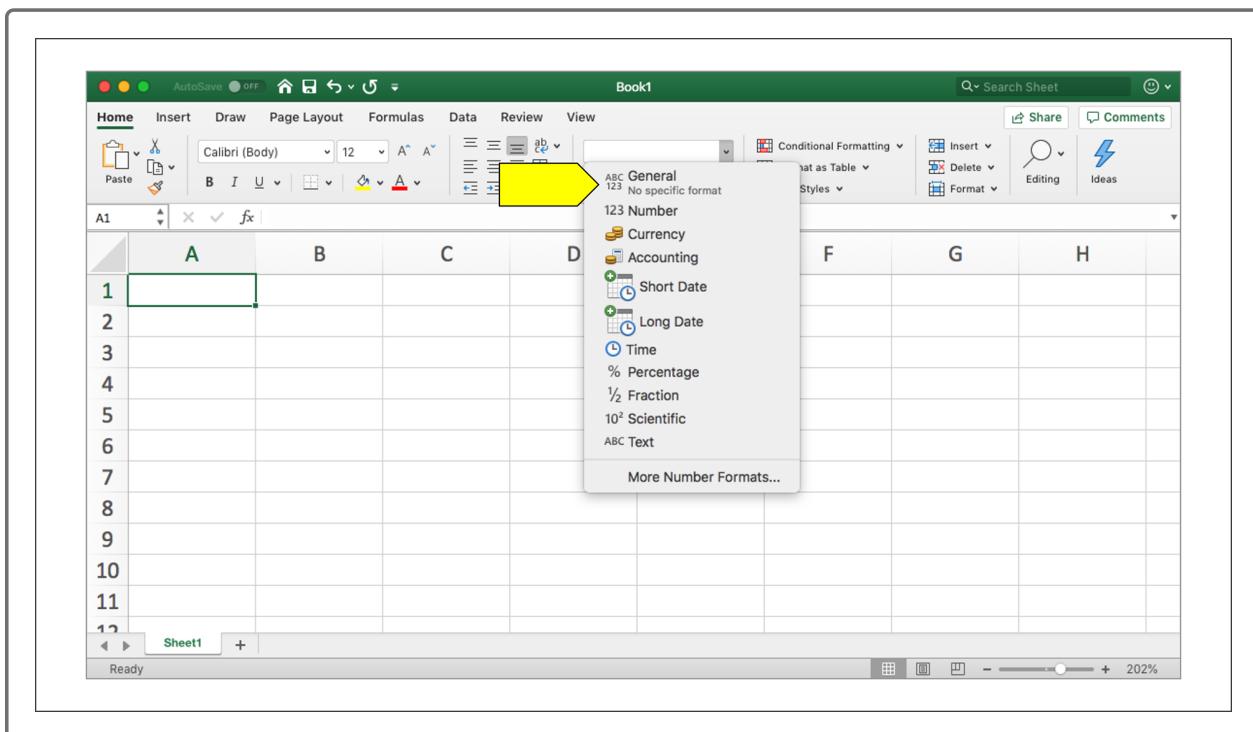
---

## What Types of Data Are Present?

Another reason to take an initial look at your data is to familiarize yourself with the types of data you'll be working with. Looking at the dataset, you'll notice that some of the columns contain text while other columns contain numbers. And not all numbers are in the same format; there are monetary units present as well as dates.

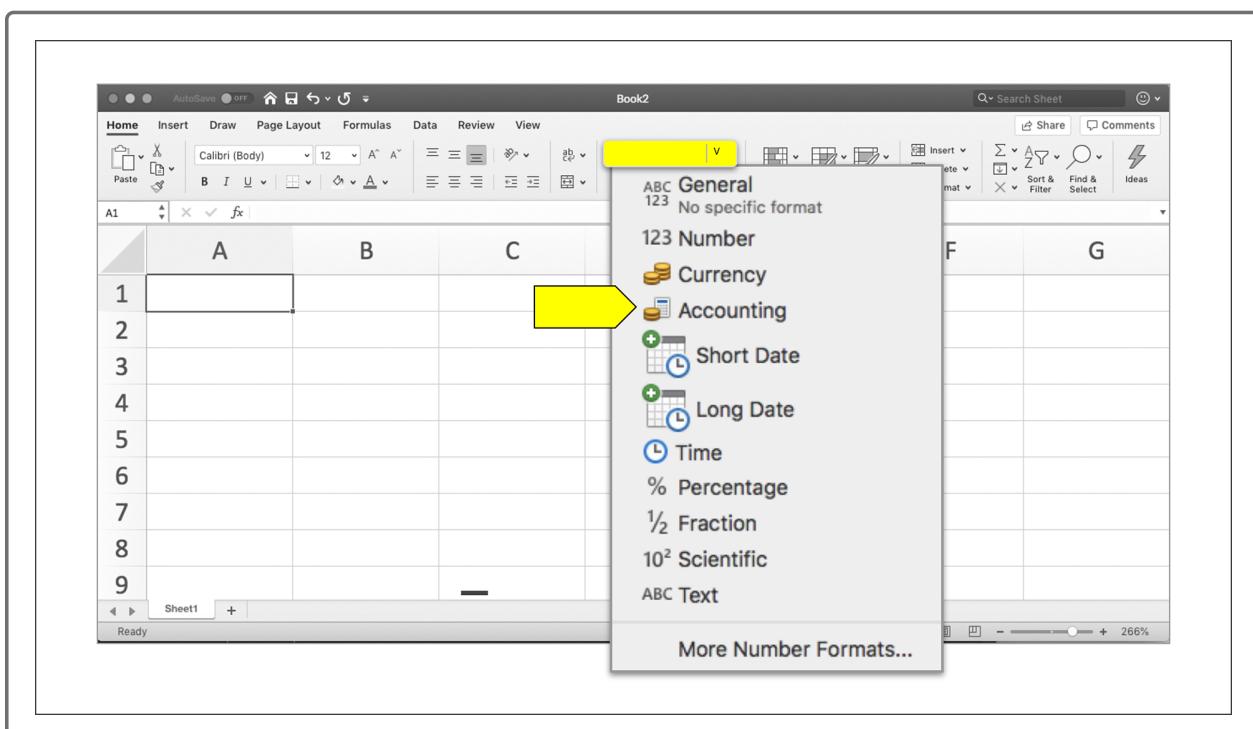
You can determine the data types in a column by clicking on a cell. For example, click on C2. You'll see that it's a description of the Kickstarter campaign, and that Excel has assigned it a specific data type. Look at the

top toolbar and note the dropdown menu that has General selected. This is the default format of the column's data.



A screenshot of Microsoft Excel showing the 'Home' tab selected. The ribbon bar at the top includes Home, Insert, Draw, Page Layout, Formulas, Data, Review, and View. The 'Font' group on the Home tab shows 'Calibri (Body)', '12', and bold/italic/underline options. The 'Number' group shows a dropdown arrow pointing to a context menu. This menu is titled 'ABC General' and '123 Number'. Other options listed include Currency, Accounting, Short Date, Long Date, Time, Percentage, Fraction, Scientific, and Text. A yellow arrow points from the left towards the 'General' option. The main worksheet area shows columns A through H and rows 1 through 12. Cell A1 is selected and contains the value '1'. The status bar at the bottom right shows '202%'. The bottom navigation bar includes tabs for 'Sheet1' and 'Ready'.

Now let's check the next column over, D2. The value has changed to Accounting because this cell contains monetary data.



A screenshot of Microsoft Excel showing the 'Home' tab selected. The ribbon bar at the top includes Home, Insert, Draw, Page Layout, Formulas, Data, Review, and View. The 'Font' group on the Home tab shows 'Calibri (Body)', '12', and bold/italic/underline options. The 'Number' group shows a dropdown arrow pointing to a context menu. This menu is titled 'ABC General' and '123 Number'. Other options listed include Currency, Accounting, Short Date, Long Date, Time, Percentage, Fraction, Scientific, and Text. A yellow arrow points from the left towards the 'Accounting' option. The main worksheet area shows columns A through G and rows 1 through 9. Cell A1 is selected and contains the value '1'. The status bar at the bottom right shows '266%'. The bottom navigation bar includes tabs for 'Sheet1' and 'Ready'.

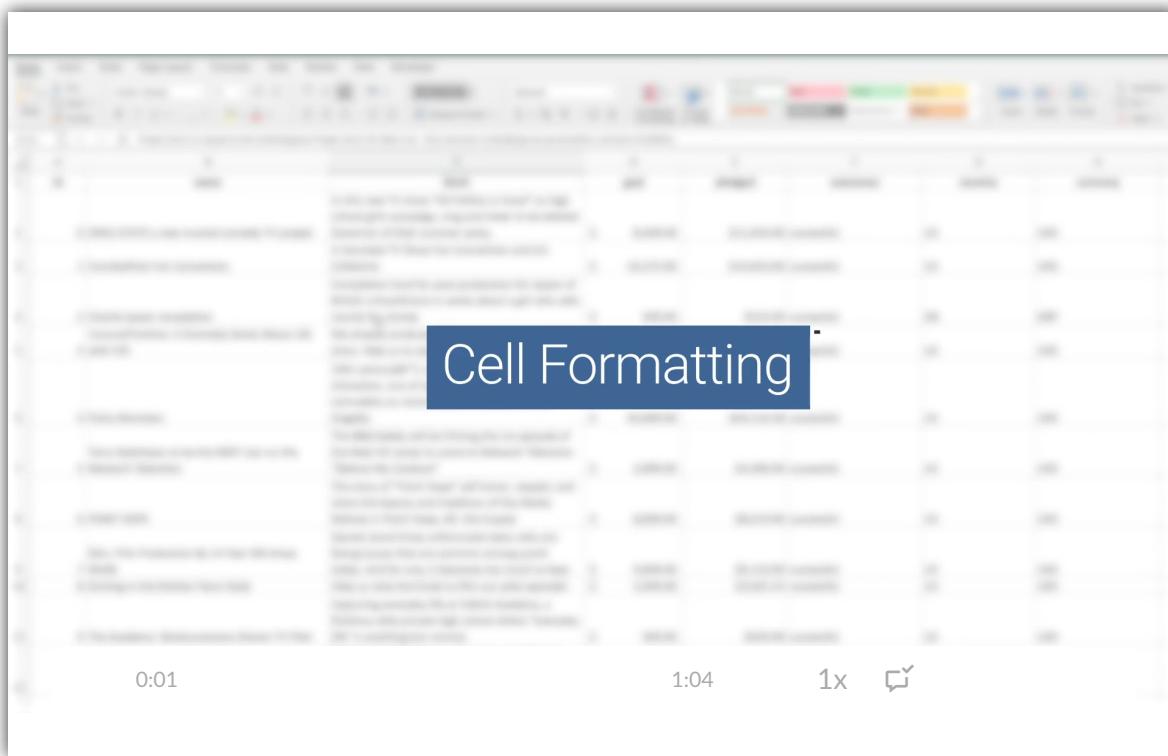
Be aware that sometimes the data isn't classified correctly, so it's a good idea to double-check them. Let's test this by changing the Goal column's format to General.

---

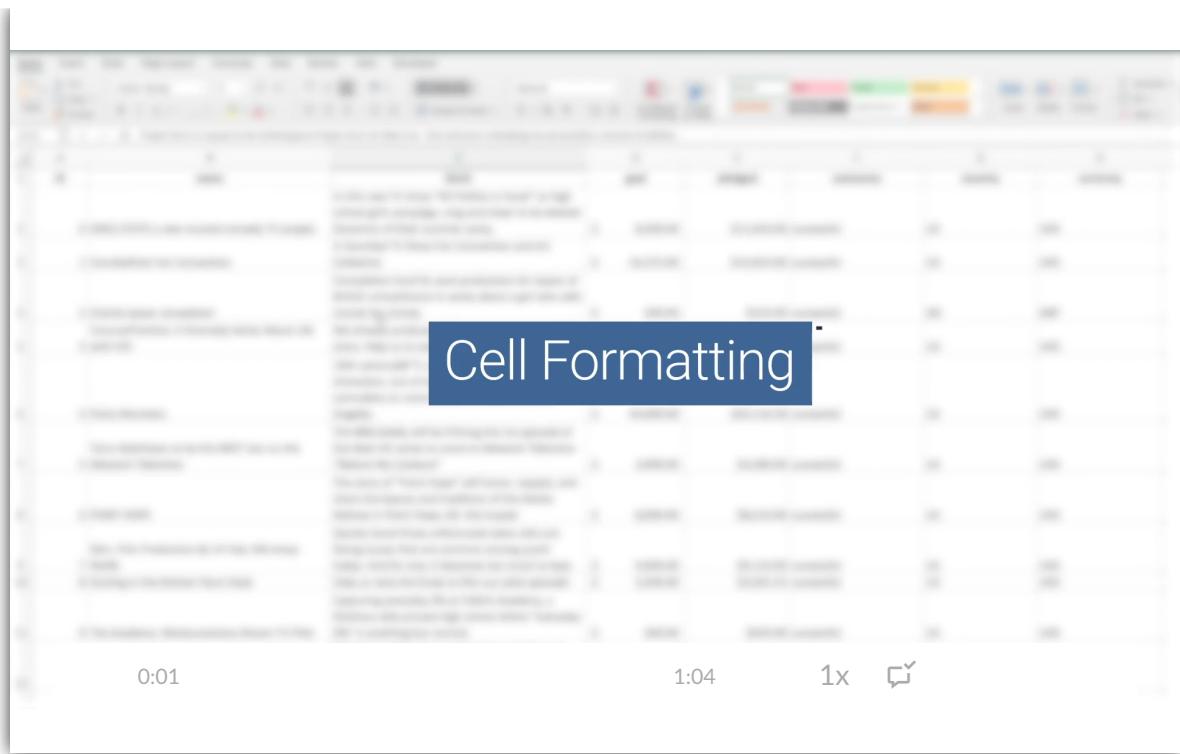
## Changing Data Formats

Watch the following video to explore changing data formats, as not all data is the same. Putting data in the correct format unlocks your ability to perform categorization, sorting, statistics, and other operations.

### macOS



### Windows



Let's format another column by changing a text column to a numeric column. Select all of column A and change the format to Accounting.

Notice anything different? Not really. This is because Excel knows that letters can't be monetary values; instead, it checks every cell in that column for numerical values and changes those. For example, a cell named "1234" would be changed to "\$1,234.00," while cells containing only text would remain unchanged. Cells containing a combination of text and numbers—for example, "1234 go!"—would also not be altered.

Now revert column A to the General format.

---

## Is the Data Readable?

Turn your attention back to the top of the dataset and scroll to view columns I and J. You may notice that the data in these columns look like dates, but they're not in a readable format. This is a common issue in data analysis; you'll often encounter data that needs to be converted to make it readable. Right now, the Deadline and Launched\_at columns contain Unix timestamps rather than dates in a standard format.

How do we know the data is Unix timestamps and not random numbers? We already know that the data is supposed to represent a date, and timestamps like these are common. But to be sure that the data are timestamps, we can use a [timestamp converter](https://www.epochconverter.com/) (<https://www.epochconverter.com/>) tool.

#### NOTE

Learn more about the use of [Unix timestamps](https://websiteseochcker.com/blog/what-is-timestamp/) (<https://websiteseochcker.com/blog/what-is-timestamp/>)..

The converter tool confirms that we're working with Unix timestamps, so we know that we'll need to convert those timestamps into a format that's readable. This task will require an advanced formula, which we'll cover as we take a more in-depth look at the Kickstarter dataset.

### 1.2.3 Filtering, Formatting, and Freezing

Now we've taken an initial look at the data to determine how many rows and columns are in the worksheet, what data types we're working with, and whether we need to convert the data to make it readable. The next step is to organize the data so that it can be more easily understood and, therefore, generate insights that will help Louise's project. We'll accomplish this task by using filters and formatting, and by freezing specific columns and rows.

Now we can see that our data is stored in columns A through N and rows 1 through 4115—that's a lot of data! Let's break down some of the information we're working with:

- The Goal column tells us how much money each campaign will need to succeed.
- The Pledged column tells us how much each campaign actually made.
- The Outcomes column tells us if the campaign met its goal.
- The Country column lists the country in which the campaign was started.

But even with this data breakdown, the current worksheet is still a little overwhelming, right? Let's make it more user-friendly by applying filters to the columns. **Filters** allow us to display only the specific data that we want to focus on.

Let's focus first on the money raised by various campaigns. Louise estimates that her play will cost \$12,000, so we can use data from the Pledged column to research projects with a similar monetary goal. We'll use filters to apply this customization.

---

## Filter the Data

Filters are commonly used in Excel because they make the data far more user-friendly. For instance, a dataset can be filtered to omit data we aren't interested in, or to single out specific data we want to learn more about. It helps negate some of the extra "noise"—information that is unnecessary to a project.

Follow these steps to add filters to your Excel worksheet.

1. Click the Data tab at the top of the screen.
2. Click Filter.
3. Click the down rows on the first row of data to reveal the sorting and filtering menu.

A screenshot of Microsoft Excel showing a data table. The table has three columns: C, D, and E. Column C contains a blurb, column D contains a goal amount, and column E contains a pledged amount. A yellow arrow points to the 'Descending' button in the filter dropdown menu for column E.

|   | C   | D            | E              |
|---|---|--------------|----------------|
| 1 | blurb   | goal         | pledged        |
| 2 | It's a pen that can draw in the air! 3Doodler is the 3D printing pen you can hold in your hand. Lift your imagination off the page!   | \$ 30,000.00 | \$2,344,134.67 |
| 3 | The world's most advanced jacket for SMARTPHONE USER, for WORKOUT, for TRAVEL, for OUTDOOR /Bluetooth Charging Wearable Apparel castAR: bridging the physical world with the virtual worlds; 3D holographic like projections in AR, fully | \$ 5,000.00  | \$1,076,751.05 |

Now we easily sort or filter the data as we see fit. Let's test this tool by sorting the goal amount from highest to lowest.

1. Click the down arrow on the Pledged column (column E) to reveal the Filter menu.
2. Click "Descending," and then close the window. **Note:** "Ascending" sorts values from lowest to highest; "Descending" sorts values from highest to lowest, which is what we want.

A screenshot of the 'Sort' dialog box in Microsoft Excel. The 'Descending' button is highlighted with a yellow arrow. The 'Choose One' dropdown shows several filter options for goal amounts.

Sort

By color: None

Filter

By color: None

Choose One

(Select All)

\$1.00

\$2.00

\$5.00

\$10.00

\$13.00

\$20.00

\$25.00

Clear Filter

This will sort the Goal column as well as all of the data tied to it. We can verify this by looking at the data contained in column A: the ID numbers are no longer in order.

Filter the Pledged column in descending order.

What value is recorded in the first cell of the column when this filter is applied?

- \$2,344,134.67 ✓
- \$1.00
- UthTopia
- 3Doodler: The World's First 3D Printing Pen

Response-specific feedback:

Correct. \$2,344,134.67 is the value that appears when the filter is applied.

[Finish ▶](#)

Now spend some time practicing how to format and sort different columns in your worksheet. Have fun with this! For example, try to find the highest goal, or combine filters by looking for the highest successful goal.

Remember, when you sort and filter multiple columns, it's easy to lose track of what's been sorted and filtered. Excel's subtle (but helpful) way of telling us which columns are sorted is to change the down arrow icon to a filter or arrow icon, depending on how the data was adjusted.

A screenshot of the Microsoft Excel application window titled "StarterBook". The Data ribbon tab is selected. In the center of the screen, there is a table with columns labeled D, E, F, G, and H. The first row contains the headers: "goal", "pledged", "comes", "country", and "currency". The second row shows data: "\$ 30,000.00" in column D, "\$2,344,134.67" in column E, "successful" in column F, "US" in column G, and "USD" in column H. The third row shows data: "\$ 5,000.00" in column D, "\$1,076,751.05" in column E, "canceled" in column F, "US" in column G, and "USD" in column H. A yellow arrow points to the "Clear" button in the Data ribbon's Sort & Filter group.

|   | D            | E              | F          | G       | H        |
|---|--------------|----------------|------------|---------|----------|
| 1 | goal         | pledged        | comes      | country | currency |
| 2 | \$ 30,000.00 | \$2,344,134.67 | successful | US      | USD      |
| 3 | \$ 5,000.00  | \$1,076,751.05 | canceled   | US      | USD      |

If you get lost in the data, that's okay! Clearing the filters and starting over can be done in two quick steps. To reset all filters, do the following:

1. Click the "Sort & Filter" button at the top right of the Excel window.
2. Click Clear.

A screenshot of the Microsoft Excel application window titled "StarterBook". The Data ribbon tab is selected. In the center of the screen, there is a table with columns labeled D, E, F, G, and H. The first row contains the headers: "goal", "pledged", "outcomes", "country", and "currency". The second row shows data: "\$ 30,000.00" in column D, "\$2,344,134.67" in column E, "successful" in column F, "US" in column G, and "USD" in column H. The third row shows data: "\$ 5,000.00" in column D, "\$1,076,751.05" in column E, "canceled" in column F, "US" in column G, and "USD" in column H. A yellow arrow points to the "Clear" button in the Data ribbon's Sort & Filter group.

|   | D            | E              | F          | G       | H        |
|---|--------------|----------------|------------|---------|----------|
| 1 | goal         | pledged        | outcomes   | country | currency |
| 2 | \$ 30,000.00 | \$2,344,134.67 | successful | US      | USD      |
| 3 | \$ 5,000.00  | \$1,076,751.05 | canceled   | US      | USD      |

# Format the Spreadsheet

Let's backtrack a bit. The Goal column was automatically sorted, but now some of the numbers in that column look odd. This is because the column width doesn't capture the entire number, so it has been shortened using **scientific notation**.

The screenshot shows a Microsoft Excel spreadsheet titled "StarterBook". The spreadsheet has three columns: "goal" (D), "pledge" (E), and "outcomes" (F). Row 1 contains the words "goal", "pledge", and "outcomes". Row 2 contains the value "8.50E+03" in cell D2 and "successful" in cell E2. Row 3 contains the value "1.03E+04" in cell D3 and "14653.00" in cell E3. A yellow arrow points from the text "14653.00" in cell E3 to the "Format" dropdown menu. The menu is open, showing various number formats: General, Number, Currency, Accounting, Short Date, Long Date, Time, Percentage, Fraction, Scientific (selected), Text, and More Number Formats... The "Scientific" option is highlighted with a blue background.

## IMPORTANT

**Scientific notation** is a method for displaying very large or very small numbers in a more compact way. Excel automatically adjusts those large or small numbers to fit within a column.

For example, the number **3E+07** can also be displayed as  **$3 \times 10^7$** . This is scientific notation for 30,000,000. This means that the number starts with a 3 and has an exponent of 7 (seven zeros).

You can automatically widen the column to hold all of the data by following these steps:

1. Place your cursor over the small line between columns D and E. The cursor should change in appearance to a vertical bar with an arrow pointing either left or right from the center.
2. Double-click to expand the column to fit the value with the most physical width.

You should now see the goal amounts arranged in order from highest to lowest. However, you might notice that when you scroll further down the sheet, it's difficult to determine what's in each column. Some of the data may not make much sense without a heading attached to it. The solution? Freeze the header.

---

## Freeze and Unfreeze Rows, Columns, and Panes

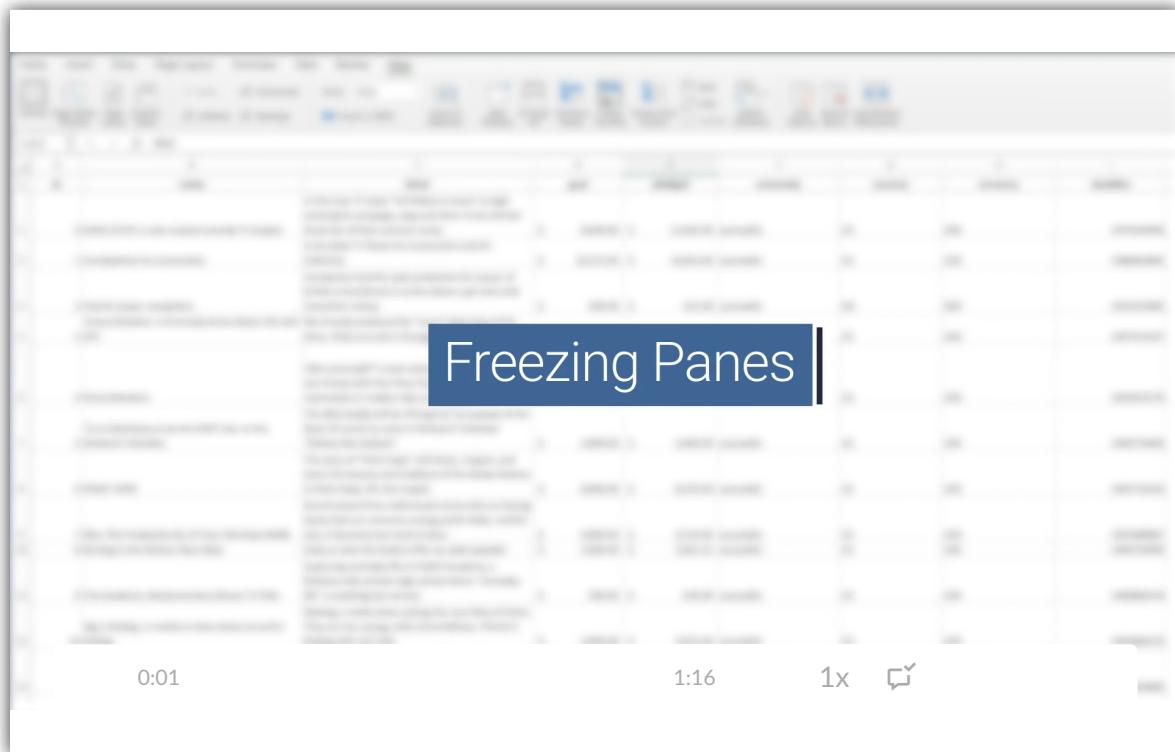
**Freezing** is a feature of Excel that allows portions of a spreadsheet, such as a row or column, to be locked in place so that it is always displayed, no matter what part of the sheet we're looking at. This means that you don't need to commit column headers to memory; simply freeze them so that they are always displayed as you scroll through the sheet.

When a column and row are frozen together, they become a **pane**. Think of this as a window: the window can be opened and moved up and down or left and right, but the pane stays in place. The same principle applies here, as shown in the following image.

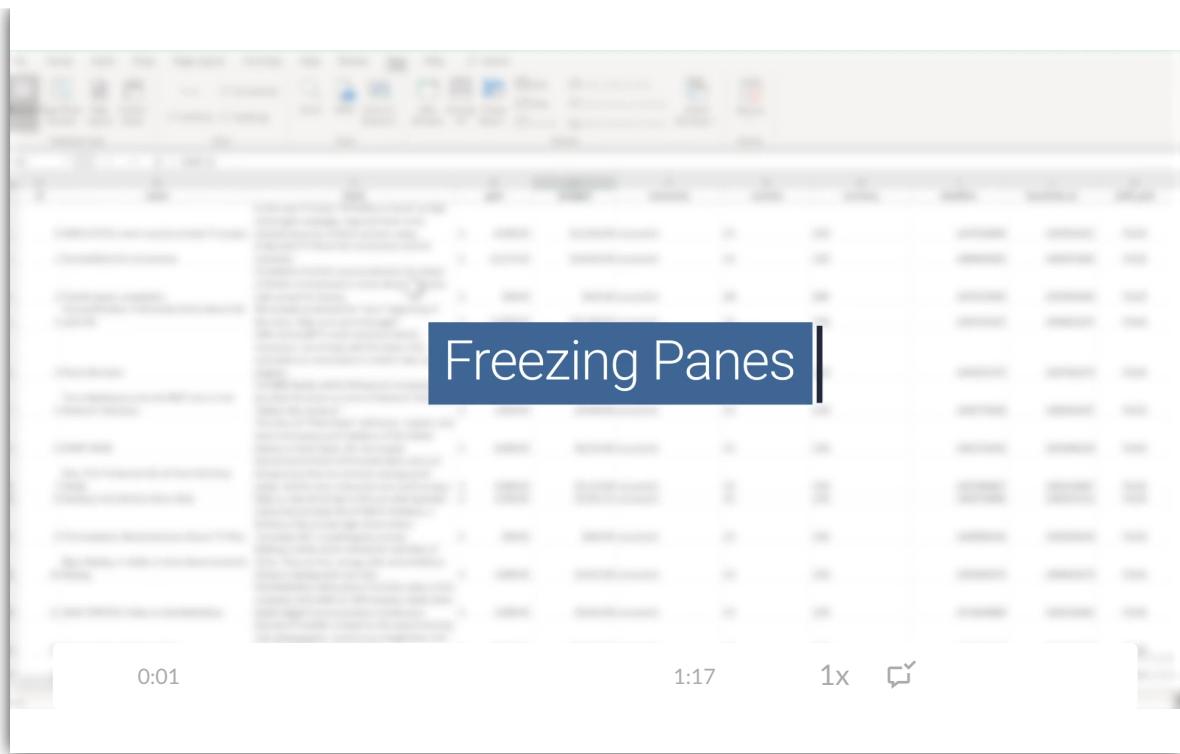
|   | D             | E                         | F        | G       | H       |
|---|---------------|---------------------------|----------|---------|---------|
| 1 | goal          | pledged                   | outcomes | country | current |
| 2 | \$ 400,000.00 | \$1,052,110.87 successful | US       | USD     |         |
| 3 | \$ 300,000.00 | \$471,567.00 successful   | NO       | NOK     |         |

Watch the following video to learn how to use the freezing function in Excel. Choose the video that corresponds to your operating system.

## macOS



## Windows



Now let's reset our view by unfreezing our panes. Click the Unfreeze Panes button below the View tab. Rows and columns are no longer static and everything is once again grouped together. Did you notice that when you clicked the Unfreeze button, the name changed to Freeze Panes? Let's investigate.

Say we want to set the name column and title row as a pane rather than freezing the first column and the top row. This is where the Freeze Panes button comes into play. This button freezes a column and row simultaneously.

Let's try it. In the worksheet, select cell C2 and then click Freeze Panes in the View tab. The new pane is created based on the location of the selected cell, in this case, C2.

C3 | Eyes as Big as Plates - The book! Featuring over 50 portraits, field notes and behind the scenes stories from seniors around the world.

|   | C   | D             | E              | F          |
|---|---|---------------|----------------|------------|
| 1 | blurb   | goal          | pledged        | outco      |
| 2 | castAR: bridging the physical world with the virtual worlds; 3D holographic like projections in AR, fully immersive environments in VR<br><br>Eyes as Big as Plates - The book! Featuring over 50 portraits, field notes and behind the scenes stories from seniors around the world. | \$ 400,000.00 | \$1,052,110.87 | successful |
| 3 | Open-source content-driven lighting system you can use with TV or PC, Mac, HTPC displays in movies,   | \$ 300,000.00 | \$471,567.00   | successful |

This tells Excel to freeze the column to the left of C2 and the row above C2. Now scroll through the data. See how both the **ID** and **name** columns stay in place? To complete the pane, the header row is also frozen. Pretty useful, right?

## SKILL DRILL

Create a pane with the first three columns and title row of the Excel worksheet. Remember that panes can be reset with the Freeze and Unfreeze buttons.

**IMPORTANT**

Remember to save your work periodically so that you don't lose it if the program crashes. Command+S (Mac) or CTRL+S (Windows) is a quick shortcut to save any work in progress.

© 2020 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.

## 1.2.4 Conditional Formatting

We're about to go a little deeper into data analysis with Excel and use conditional formatting to customize our worksheet. This will refine the visual presentation of the data in the worksheet to provide Louise with information at a glance.

By now, you should feel pretty familiar with the dataset. You've read it, resized it, filtered it, formatted it, and froze it. Now we're going to take our analysis a step further and customize the data to provide visual feedback.

What do we mean by "visual feedback"? Imagine if traffic lights were words rather than colors. What would that experience be like? Or think of a weather map, where the strongest part of a storm is shaded an angry red. In the cases of traffic lights and weather maps, colors provide a visual link to a specific situation. We can do the same in our data worksheet with conditional formatting.

---

### What's Conditional Formatting?

**Conditional formatting** in Excel refers to customizing the appearance of a cell based on its value. This is a great way to provide viewers with information at a glance. And, much like customizing an essay or report with Microsoft Word, there are several ways to format a spreadsheet: each cell can be filled with a different color, various fonts and text sizes can be used, and so on. For our project, we'll apply conditional formatting to some of the columns in the worksheet in order to more easily view the outcomes of Kickstarter campaigns.

### IMPORTANT

**Conditionals in Code** If you're an experienced Excel user who's looking forward to the coding portions of this course, that's great! Conditional formatting is a concept that you'll see in action often.

Conditionals make frequent appearances in all programming languages. When coding, they are referred to as conditional statements because they are paired with keywords such as *if*, *and*, and *or*.

When using conditionals in Excel, a cell can be highlighted a certain color based on its value. *If* the value of a cell is greater than 50, *then* color it green.

These same logical rules apply when conditional statements are included in coding script.

Let's put this into action by applying conditional formatting to our data.

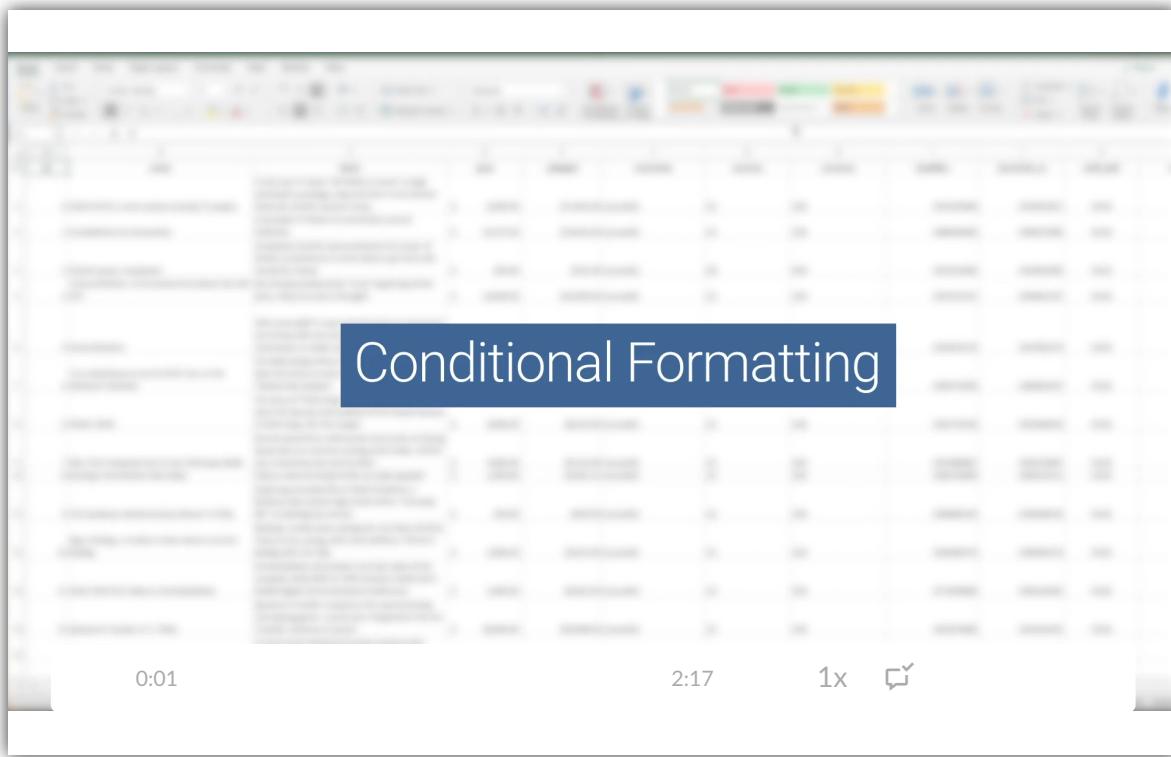
---

## Use Conditional Formatting

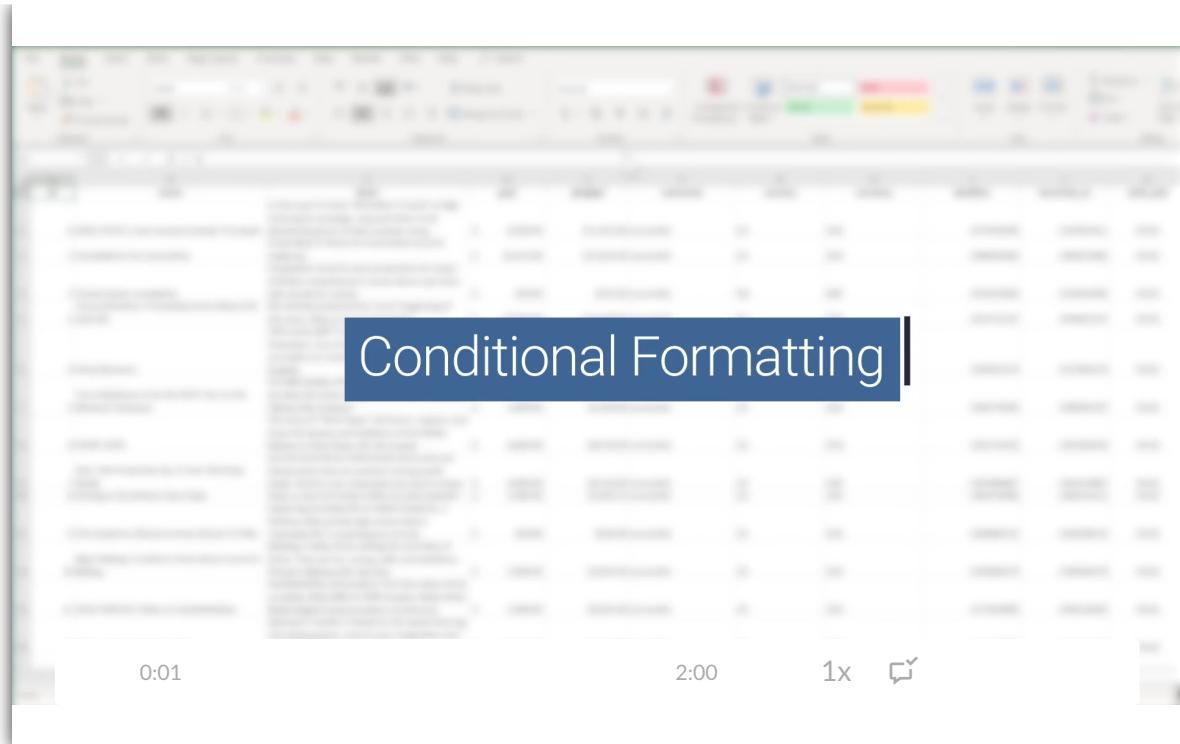
There are four main categories in the Outcome column—Successful, Failed, Canceled, and Live—so let's color-code by category. We can use conditional formatting to automatically apply a preselected color to the cell.

Watch the following video to learn about conditional formatting in Excel.  
Choose the video that corresponds to your operating system.

## macOS



## Windows



We now have our first two rules! These apply a color to the live and successful campaigns. There are two more rules to create. Give them a try in the Skill Drill.

### SKILL DRILL

Create two more rules for the remaining outcomes. Highlight failed outcomes in red and canceled outcomes in yellow.

## 1.2.5 Finding Averages

The ability to visually process outcomes at a glance will be very useful for Louise. Let's add a bit more customization to the sheet by creating another easy-to-interpret column.

Many of the campaigns missed their goal amount by a small margin. Instead of looking at both the Goal and Pledged columns to determine the deficit, let's create a new column that contains this information.

---

### Create a New Column

In your worksheet, name the next empty column Percentage Funded. In this column, we'll use Excel's **ROUND** formula to determine how much of the campaign's goal was met. Data from the Pledged and Goal columns will be used to find the percentage funded.

---

### Adding New Data to the Column

In the first row of our new Percentage Funded column, enter `=ROUND(E2/D2*100, 0)` into the cell. Let's break down this formula to see how it works.

1. `=ROUND` tells Excel that we'll be using the ROUND formula.
2. `(E2/D2)` specifies which data is being utilized in the formula: column E, row 2 (pledges) divided by column D, row 2 (goals). The result is the percentage of completed donations, shown in decimal form.
3. `*100,` multiplies that result by 100, giving us the percentage as a whole number.
4. `0)` specifies how many numbers after the decimal will be viewed. In our case, that's zero.

Because there are so many rows of data, it would be time-consuming to apply the formula to each individual cell. Thankfully, there's a shortcut we can use to apply the formula to the entire column.

After inputting the formula and running it, the result should be a whole number of the percentage. Now, to apply the formula to the rest of the column, click the cell containing the new data. See how it's outlined in green? This tells us that it's the active cell.

Next, place your cursor over the bottom right corner of the cell. The crosshairs of the cursor change from white to black; this is how you know you're in the right spot. When the crosshairs are black, double-click that spot; Excel will automatically apply that formula to the entire column.

Reset your sheet by clearing all filters and making sure the campaign id's are in order, from smallest to largest. What is the percentage funded for cell O522?

- 23
- 0
- 102
- 150.15



Response-specific feedback:

Correct. 102% of the campaign was funded.

[Finish ▶](#)

## Value Shading

To summarize what we just did: we created a new column of data using a new formula to find the percentage of a campaign's funding. Now we can quickly determine how close a campaign came to reaching—and in some cases, exceeding—their funding goal.

We'll now take it a step further and apply conditional formatting to the new column. This time, instead of assigning a color to a column based on a single word or phrase, we'll apply a range of colors, based on the percentage funded. Like we did for the Outcomes column, we'll create a formatting rule to let the colors do the talking for us.

Highlight the Percentage Funded column in the worksheet, and then click the Conditional Formatting button. Now we'll create a new rule by following these steps:

1. Select “Color Scales” followed by “More Rules.” At this point, we can select colors based on the lowest and highest values. The colors will automatically change based on the cell’s value.

2. Select the color red as the minimum color and blue for the maximum color.
3. Click OK.

Now we can more easily see which projects were fully funded.

## SKILL DRILL

Highlight the column again and click Conditional Formatting followed by Manage Rules. Try changing the maximum to the 90th percentile. How does it look?

So far, you've imported real-world data and sorted the pledges in descending order, created a new column of data, and applied conditional formatting that allows Louise to further analyze the data. Our data story is starting to come together!

## Average Donations

Kickstarter, like other crowdfunding platforms, allows project creators to add incentives for different pledge amounts. We can help Louise set up her incentives by first determining how much money people have pledged to campaigns historically. Is it \$3? \$10? Even \$50? Excel will help us perform this type of calculation.

The first step is to create a column that will provide a quick look at the average donation. For this step we'll follow a process like the one for creating and filling the Percentage Funded column. When we created the Percentage Funded column, we:

1. Selected an empty column and added a header
2. Input the round function into the cell below the header
3. Applied the formula to the rest of the column

Let's do this again with a new column that we'll name "Average Donation."

Next, we'll use the ROUND formula again, but with a few modifications, shown below.

=ROUND(E2/L2,2)

We last used this formula to create a percentage as a whole number. This time we're finding an average, so we want to view the next two digits after the decimal. By modifying the formula to use a 2 instead of a 0, we're telling Excel to include those extra two digits.

To find the average, we'll use data from the Pledged column again. But instead of dividing by the goals (as we did last time), we will divide by the backers.

Use this ROUND formula on the first cell of the Average Donation column. Then, in the next empty cell, use the same formula with a 4 instead of a 2 to specify how many digits will follow the decimal: =ROUND(E2/L2,4). See the difference? Now let's revert the formula back to using a 2, as we don't actually need four digits in the column: =ROUND(E2/L2,2).

Now apply the ROUND formula for the first cell to the entire column. Make sure there are only two digits after the decimal point.

Final score: 100%

 [Retake](#)

## 1.2.6 Errors and Debugging

One of the pain points in data analysis is errors. The new column (Average Donation) you created has errors, so let's fix them in order to present a clean analysis to Louise.

When working with data, you're bound to encounter errors. Some errors occur due to a mistake, while others pop up because Excel doesn't recognize a formula's output format, even if it's valid.

Techniques used to fix errors vary. Sometimes errors are obscure and require research, while other errors have simpler fixes, such as correcting a typo. The process of researching an error and incorporating its solution is called **debugging**.

Recall the `#DIV/0!` error we encountered earlier. Why did it occur? Let's investigate.

Kickstarter requires every campaign to have a fundraising goal. However, not every campaign has backers, which means, in some cases, there is no number to divide by in the formula. Our formula, `=ROUND(E2/L2,2)` uses data from the Pledged and Backers columns. Let's look at row 124 and

plug in the numbers ourselves. Now, our formula becomes `=ROUND(0/0,2)`. If we were to take out a calculator and try to divide 0 by 0, we'd get an error there, too. No wonder it's not working correctly. The `#DIV/0!` error occurs because numbers are not divisible by zero.

While this error doesn't hinder our research, we can and should clean it up.

---

## Using IFERROR()

To improve the look of our formula output, we'll need to integrate the `=IFERROR(value,value_if_error)` formula. This formula catches errors and replaces them with a user-defined input. In addition, we'll add a bit of a twist by nesting this formula and the ROUND formula.

### IMPORTANT

Nesting formulas occurs when one formula is nested, or lies inside, another formula. This is powerful because it allows both formulas to run simultaneously.

The **IFERROR** formula is designed specifically to hold another formula within it, while the **ROUND** formula is designed only to perform mathematical calculations. If the order was reversed, neither formula would be correct, resulting in additional errors. Therefore, the order of nesting is important.

To address the division error, we'll use this formula:

`=IFERROR(ROUND(E2/L2,2),0)`

Let's break down what's happening here.

- **IFERROR** is now the beginning of the entire formula, so the equals sign is in front of it instead of **ROUND**.

- The value is our intact `ROUND` formula; nothing has been changed.
- `,0)` tells Excel that we want a zero-value input when the formula attempts to divide by 0. The formula is completed by closing the parenthesis.

We specify a zero-value input instead of text (such as “no backers”) because we want a numerical data type throughout the column. This way, if we perform analysis on that column’s data, we won’t encounter additional errors.

After applying the updated formula to the entire column, scroll down to row 124. You should see that the error has been replaced by a zero.

Because Louise is interested in starting a campaign for theater, let’s filter the worksheet to show only theater campaigns in the Category/Subcategory column. This will allow a more focused view of our category of choice by trimming down the data and eliminating what we don’t need.

## SKILL DRILL

In the first cell of the Percentage Funded column, replace the 0 with text, such as “no backers.” Then, apply the updated formula to the column and view the result.

Repeat these steps, but this time, update the string to “N/A.”

There are many different ways to address errors with the IFERROR formula. Because the column we’re working with is a numerical data type, change the formula back to `=IFERROR(ROUND(E2/L2,2),0)`.

### 1.3.1 Pivoting Toward Success

**Louise** is looking forward to your analysis, but she would benefit from additional visualization of the data so that she can see the outcomes of all the categories. Let's help Louise by creating summary tables, charts, and graphs.

Now we've organized our data and viewed data in the Theater category. Why not break it down even further?

There's a column in the worksheet named "Category and Subcategory," which groups a main category (such as "film and video") with all of its subcategories (like science fiction, drama, animation, and so on). We can make our data more detailed by splitting the Category and Subcategory column into two distinct columns: "Parent category" and "Subcategory." This gives us additional data to use in our analysis.

While looking at the theater data as a whole has been helpful, including more data will make our analysis even more detailed and thus generate additional insight.

---

## Create Subcategories

Start by putting the subcategories into their own column. In the worksheet, clear all the active filters (if there are any), and then follow these steps:

1. Select the "Category and Subcategory" column.
2. Copy the column using the keyboard shortcut Command+C (Mac) or CTRL+C (Windows).
3. Paste the data into the next empty column using the keyboard shortcut Command+V (Mac) or CTRL+V (Windows).
4. Click the Data tab.
5. Click the “Text to Columns” button.
6. The "Convert Text to Columns Wizard" appears.
7. Inside the "Convert Text to Columns Wizard":
  - Select “Delimiter” and click “Next.”
  - Uncheck the “Tab” box and check “Other.”
  - Place a backslash ( / ) in the box, then click “Next.”
  - Select “Text” from the “Column data format.”
  - Click "Finish."

Good work—you've broken down the data into more categories. Remember to name the new columns “Parent category” and “Subcategory” respectively. Let's use this new data to create a pivot table.

---

## Pivot Tables

**Pivot tables** are a powerful Excel tool that condenses data into a summary that delivers information based on our questions.

This is a lot to take in, so let's build a pivot table together, step by step.

## **IMPORTANT**

Pivot tables and pivot charts (which we'll learn about next) are extremely versatile and powerful tools. They allow us to pick and choose the data we want to analyze and then tweak it with visual customizations. Pivot tables also let us continue to tweak the view by filtering our chosen data after it's been set to a graph.

Before we can create a pivot table, we need to know what questions we want the table to answer. Ask yourself the following:

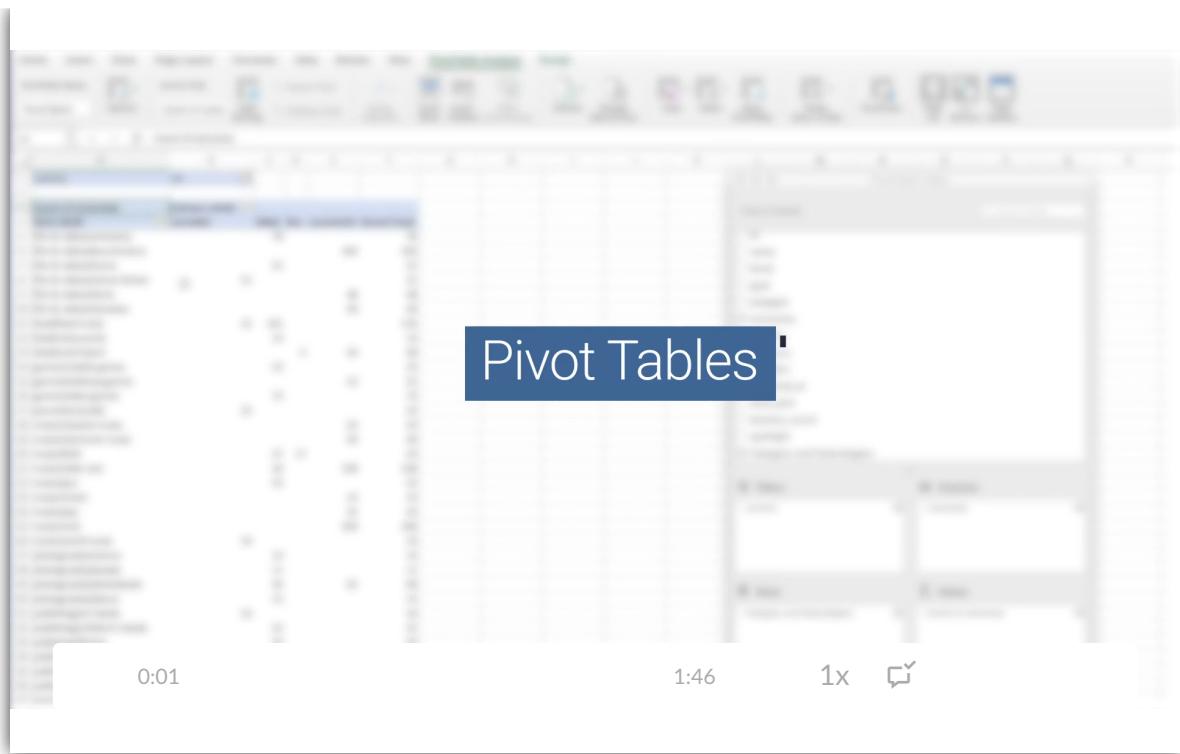
- Which data do we want to see summarized?
- How do we want the data to be presented?

Now we have a starting point. Let's dive in! Watch the following video to learn more about pivot tables. Choose the video that corresponds to your operating system.

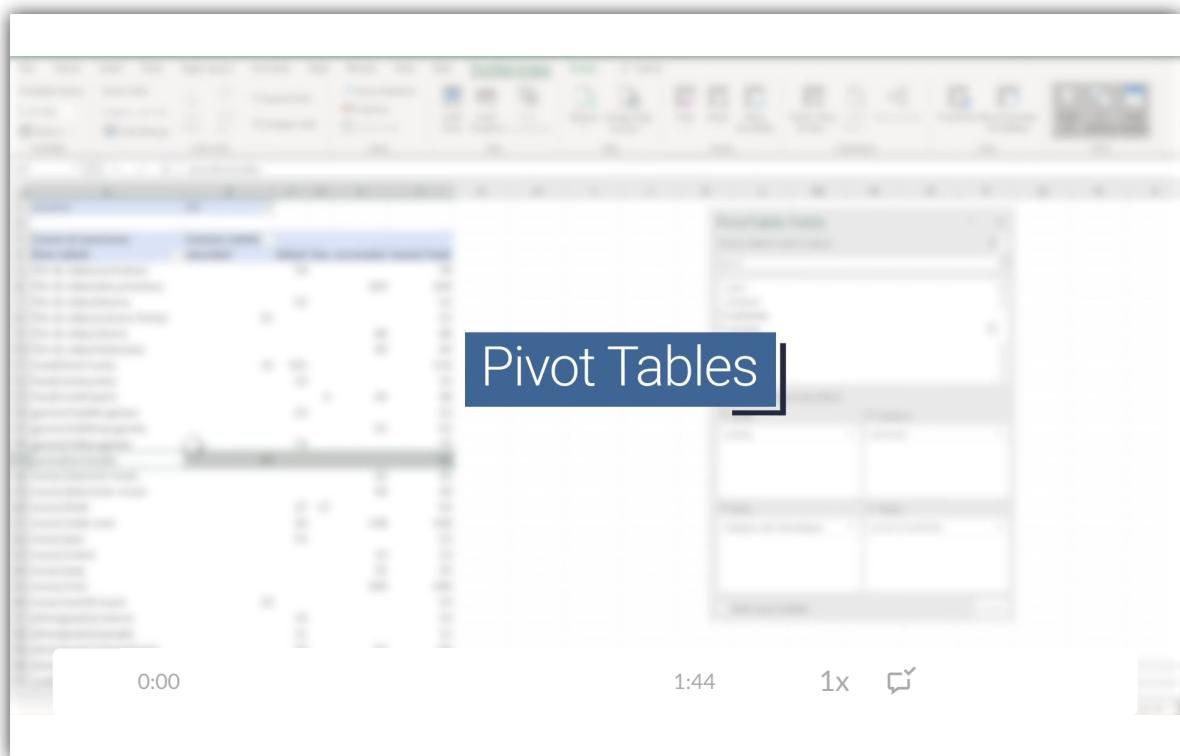
## **macOS**

## **IMPORTANT**

There is an issue with some versions of Excel 2016 for Mac. Make sure you update your Excel to the latest available version.



## Windows



Experiment with the pivot table you just created by switching the outcomes to rows and the parent category to columns. Note that the information is still being delivered, but it's displayed differently. However, the first version of the pivot table was more compact and easier to read, so let's switch back.

## FINDING

We have found that there were 525 successful theater campaigns in the United States.

## SKILL DRILL

Sometimes creating pivot tables requires some experimentation. Feel free to experiment by creating a new pivot table and switching the row and column fields, adding multiple items to multiple fields, or even using a different value.

### 1.3.2 Charting the Parent Category

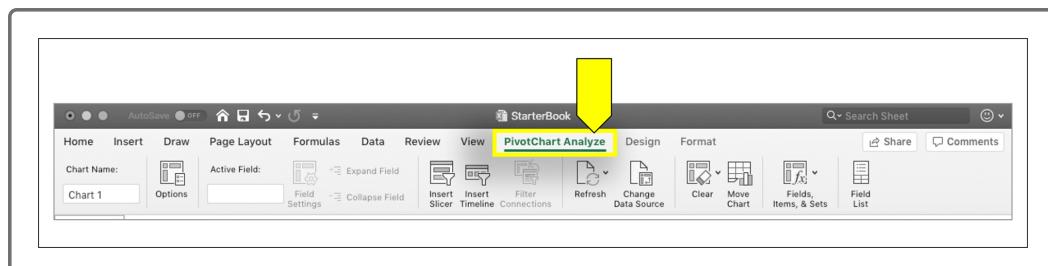
To provide Louise with a visual summary of the data in our pivot tables, we need to add visualizations using the same data.

## Pivot Charts

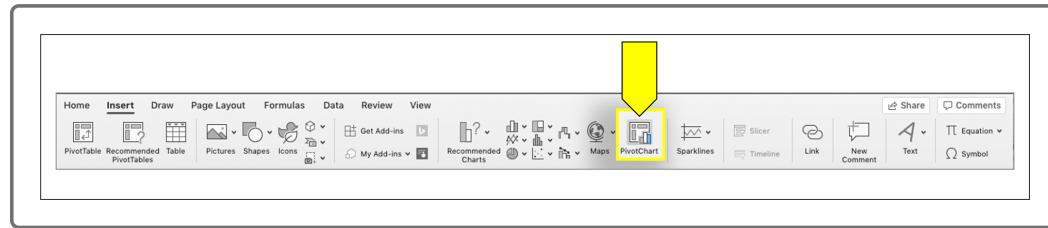
**Pivot charts** complement pivot tables by using table data to create visualizations such as bar charts and line graphs. Visualizations aid in uncovering a link or trend, especially when a table isn't able to.

Return to your pivot table. To add a visual representation of the data, follow these steps:

1. Click anywhere on the pivot table to bring up the PivotTable Analyze tab, and then click it.

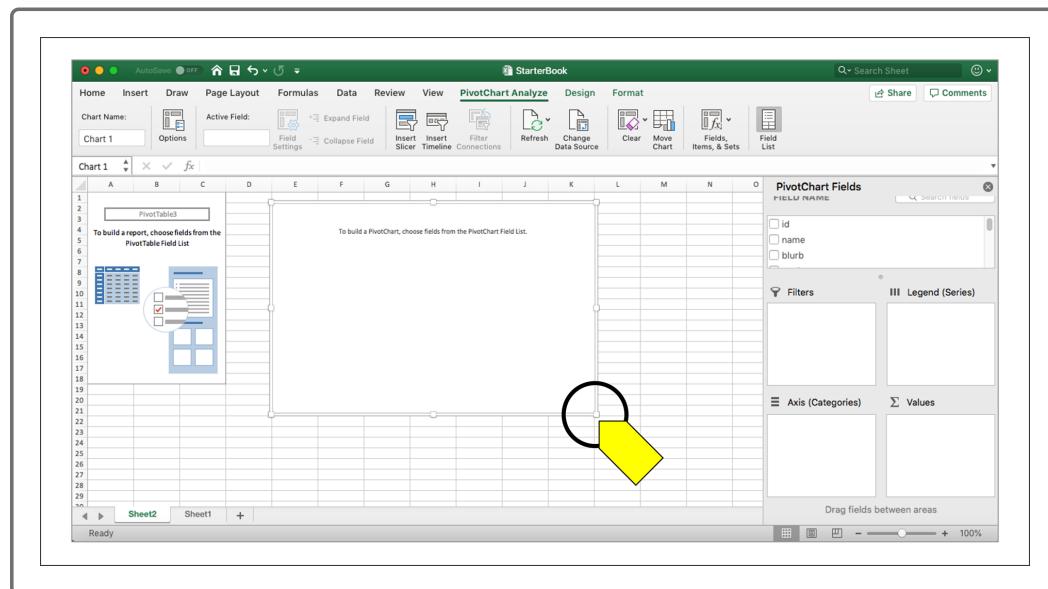


2. Click the PivotChart button on the toolbar.



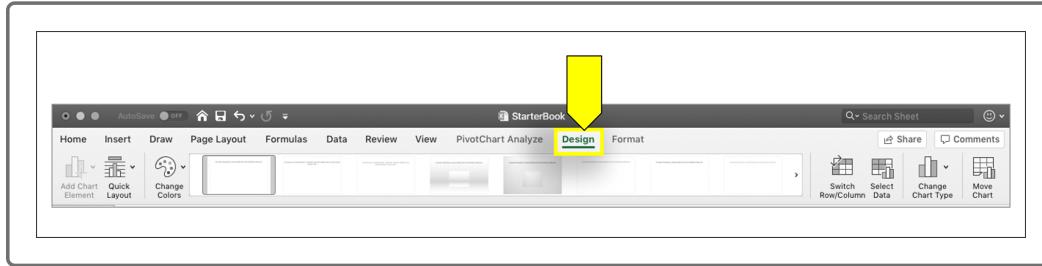
When this button is clicked, Excel will automatically choose a chart based on the table data, which in our case is a bar chart.

Resize the chart as needed by dragging the corner out from the center of the image.

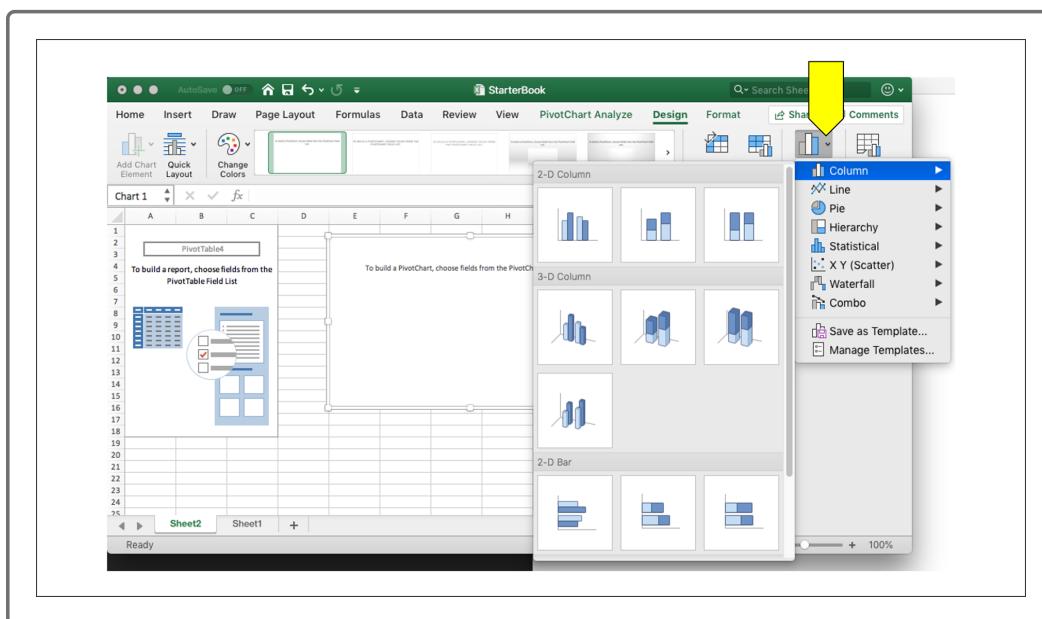


This bar chart is great, but it can be improved even further by changing it to a stacked bar chart.

3. Click on the chart, and then click the Design tab on the toolbar.



4. Click the Change Chart Type button, and then select Stacked Bar Graph from the pop-up menu.



This is a great chart. Each category has a single bar that tells us the same information as the table. Even better, this chart is just as interactive as the table it's linked to. If we change a filter, the table will automatically update itself to reflect the new data. Additionally, information is displayed in a small box when we hover our mouse over a column.

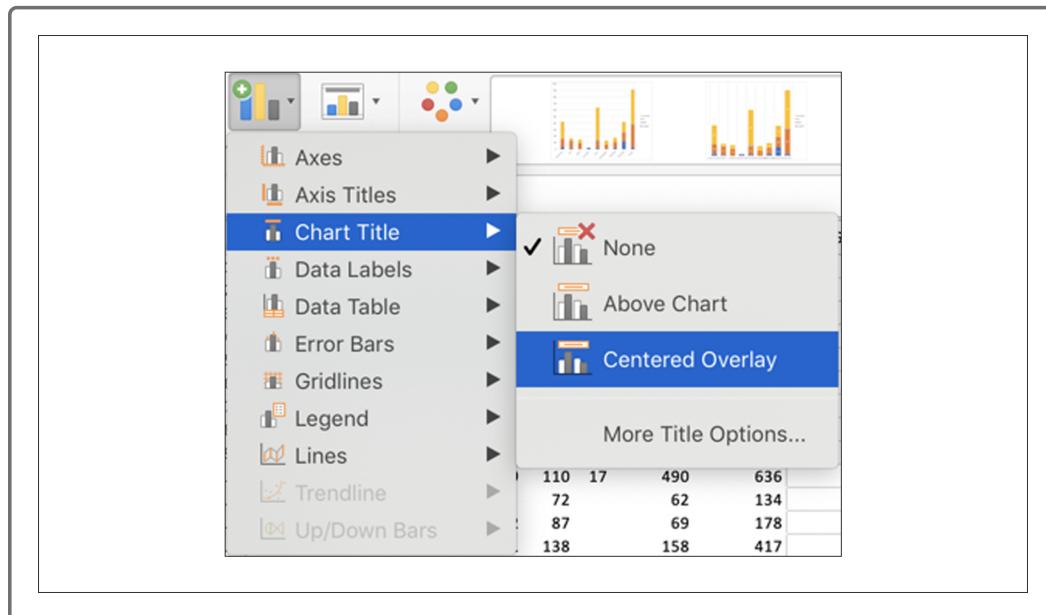
Final score: 100%

[Retake](#)

We're not quite done yet, though. It's a good practice to give visualizations a title to link the chart to specific data. This way, if we revisit the image after not working directly with the data for any length of time, the title will remind us of what the data represents.

Add a title to your bar chart by completing the following steps:

1. Return to the Design tab and click Add Chart Element on the left side of the toolbar.
2. From the pop-up menu, select Chart Title followed by Centered Overlay.

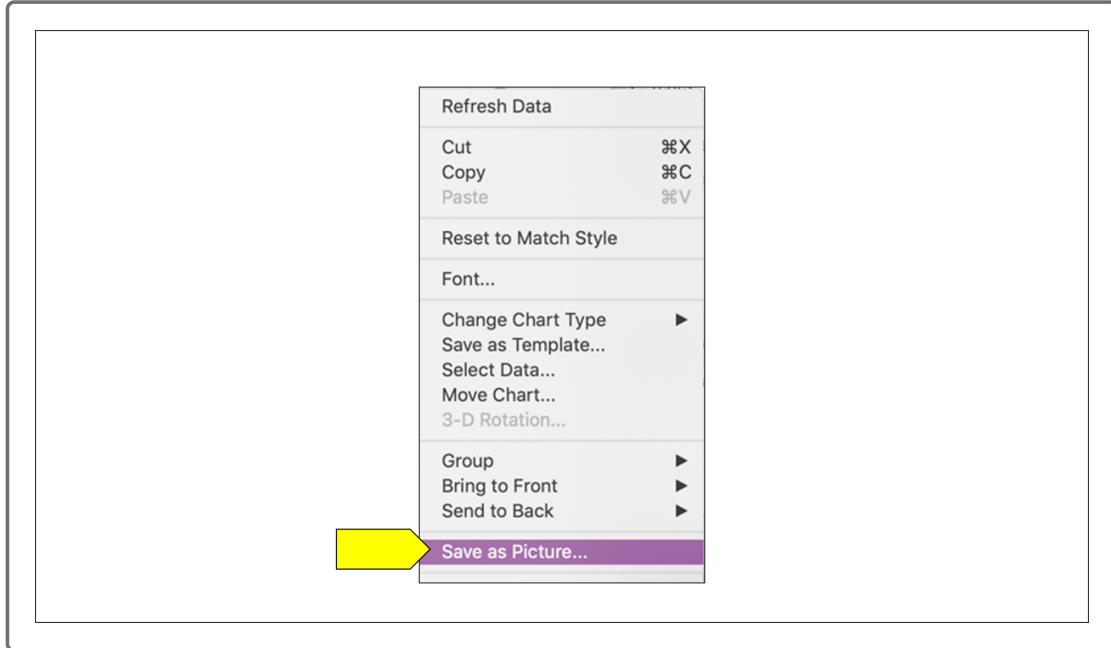


A text box with Chart Title will appear at the top of your chart. Enter a new name by double-clicking the placeholder text and typing in your own title, such as Parent Category Outcomes.

Now that we have a chart with a title, let's save it as an image to use in our report later. We'll be giving Louise a report of our findings when we're done, complete with images of our analysis. It's good to save images as we go; this way, we won't need to retrace our steps to save them later.

**macOS**

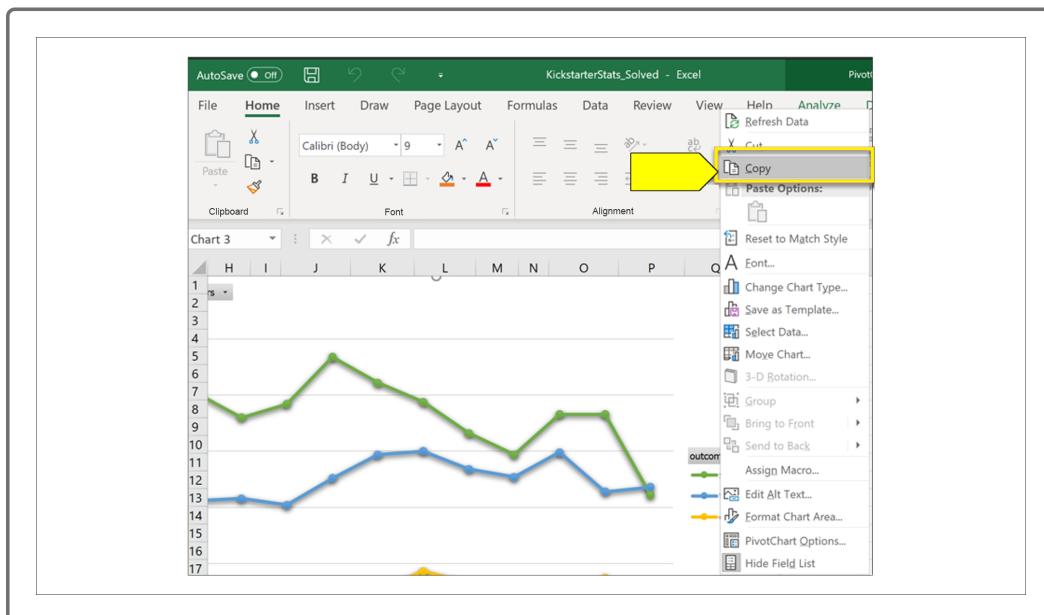
To save an image of your chart, select the chart by clicking anywhere on it. Then right-click and select "Save as Picture."



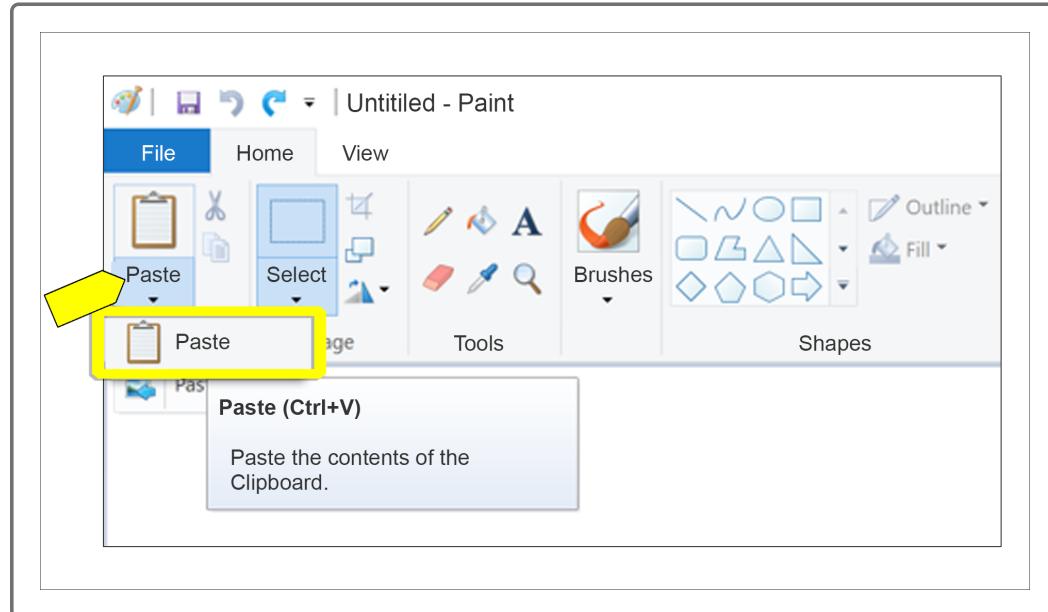
## Windows

To save an image of your chart, follow these steps:

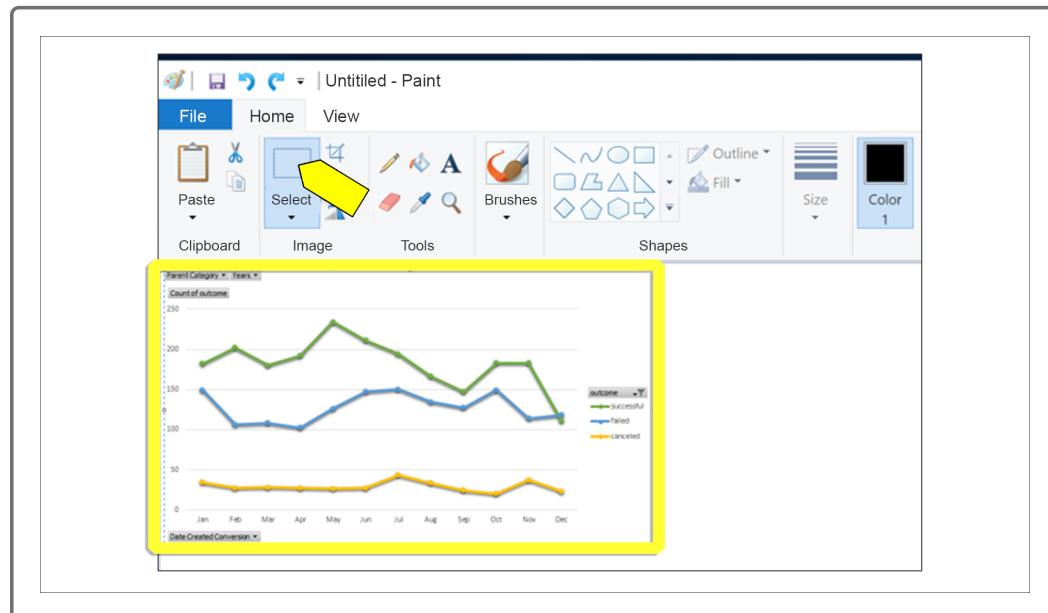
1. Right-click the image and select “Copy” (or use the CTRL+C keyboard shortcut).



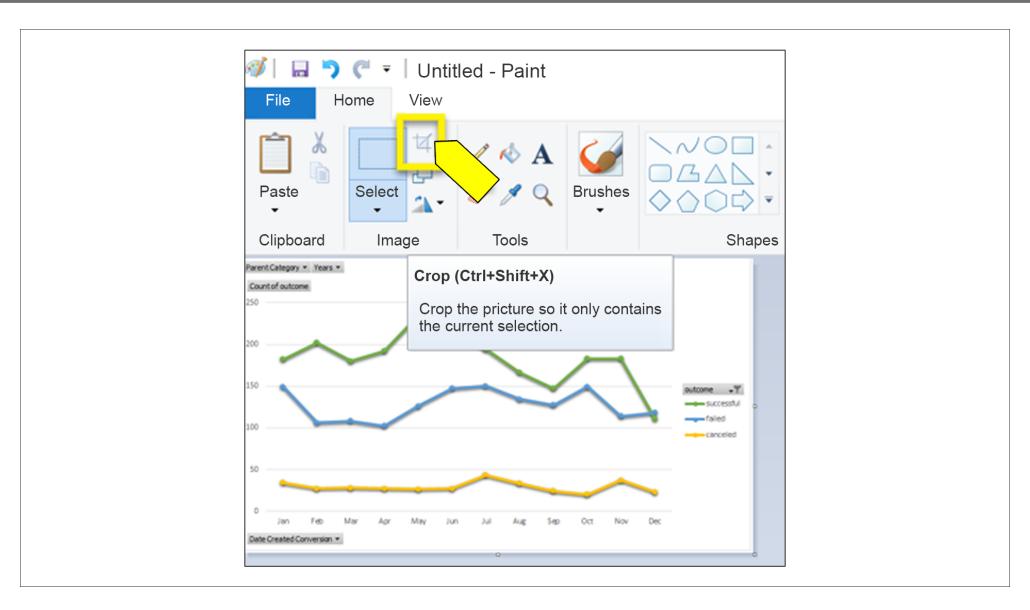
2. Open a graphics editor, such as Microsoft Paint, and paste the image using the Paste function from the ribbon or by pressing CTRL+V on your keyboard.



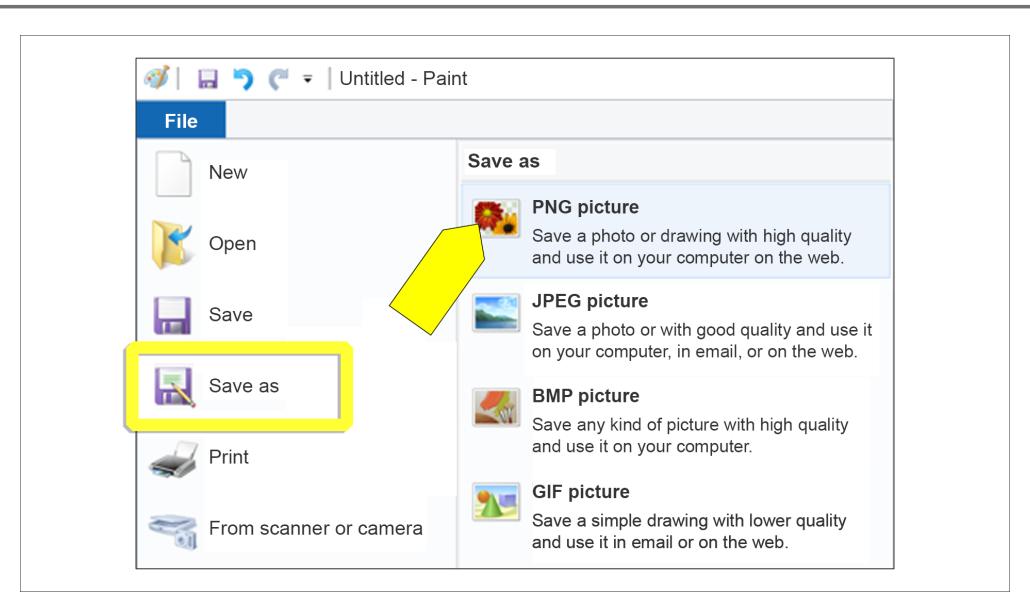
3. There may be excess white space around your chart. To trim it, use the Select tool to select only the chart.



4. Click the “Crop” button.



5. Save your image by clicking “File” followed by “Save as,” and then select “PNG picture.” Save this new image in the same directory as the data file. We’ll need to access it later to add it to our report.



This new chart is really informative. We can easily see which parent categories performed well and which ones did not. But what about subcategories? Louise is going to create a play, not build a theater. Let's

create a new pivot table and pivot chart, but this time we'll filter by subcategory.

---

## Chart the Subcategories

Let's make another chart, this time using subcategories. We're using subcategories in order to focus our analysis on an area that is more relevant for Louise: theatrical productions. We'll follow a process similar to the one used to create the original chart using the parent categories, but this time we'll create a chart and pivot table at the same time. Return to the original spreadsheet to get started, and then follow these steps:

1. Click the Insert tab.
2. Click the PivotChart button.
3. Keep the default range selection and place the PivotChart in a new worksheet. Click OK.

**Tip:** The default range includes the entire worksheet, unless there is already a highlighted section. Make sure there are no sections highlighted before completing this step.

4. Name the new sheet Subcategory Statistics and relocate it to the right-most sheet tab (if it isn't there already).
5. Drag and drop the following:
  - Country and Parent Category to Filters
  - Outcome to Columns
  - Subcategory to Rows
  - Outcome to Values
6. Click the PivotTable Analyze tab and then click PivotChart.
7. Click the Design tab at the top of the screen and select Change Chart Type. Choose Column followed by Stacked Column.

## 8. (Optional) Customize the chart colors and grid.

The grid is made up of horizontal (and sometimes vertical) lines that help us quickly measure the data being displayed. This can be customized in Excel to change the line style or even remove them completely.

What do you notice? This new pivot chart looks almost exactly like the first one we created and has the same functionality. Let's now filter this chart by country to see if the origin of the campaign has an impact. First, select the U.S. and take a look at the statistics.

Now we know that the total number of Kickstarter campaigns is just over 4,000, and 3,038 of them are from the U.S. Let's look at Great Britain's statistics.

### FINDING

We have found that while there is only a total of 604 Kickstarter campaigns for plays in Great Britain, the "plays" category is the most successful.

### 1.3.3 Timing Success

**Another** valuable piece of data is the length of fundraising campaigns. Is the length of a campaign correlated with its success? To help Louise plan her campaign timeline, let's take a closer look at how campaign length might be tied to its outcome. In order to do this, we'll need to convert the Unix timestamps to a more readable format.

Now we've looked at the different categories and compared countries of origin, which has provided a lot of insight into Kickstarter campaigns. Now let's consider time, or more specifically, whether the length of a campaign makes a difference in determining its success. For example, is a shorter or longer campaign more effective? Is there a certain time of year when campaigns tend to be more successful? We have the data, so let's find the answers.

---

## Convert Unix Timestamps to Readable Format

Look at the Deadline and Launched\_at columns in your worksheet. The data in these columns is not exactly easy to read. They contain Unix

timestamps, which measure time as the number of seconds since midnight of January 1, 1970. While interesting, these timestamps are not exactly something we can use easily.

Thankfully, Excel helps us convert these timestamps into a day-month-year format that we can interpret. For this conversion, we'll need to use another formula.

In a new column, add the heading "Date Created Conversion." Then, enter the following formula, making sure that J2 is replaced with the first data cell of the Launched\_at column:

```
=((J2/60)/60)/24)+DATE(1970,1,1)
```

Let's break this down.

- `=` tells Excel that we're using a formula.
- `((J2/60)/60)/24` iterates the following: cell J2 (the first cell in the Launched\_at column) divided by 60 (seconds), then divide that by 60 (minutes), then divide that by 24 (hours).
- `+DATE` tells Excel that we're using the DATE formula as well.
- `(1970,1,1)` is the date that the Unix timestamps began counting from, also known as the **epoch**.

Essentially, we're using the formula to figure out how many days, minutes, and seconds the timestamp translates to, and then we're adding it to the January 1, 1970 date. It's a little confusing, but it adds up to the actual time of the campaign launch.

## SHOW HINT

Follow these same steps to create a Date Ended Conversion column, but use the data in Deadline instead of Launched\_at.

Input the same formula, but change J2 to the first cell of the Deadline column (I2). Test the formula on the first cell to ensure the conversion goes smoothly, and then apply it to the rest of the column.

---

## Create a New Pivot Chart

Now that we have a date range for each project, let's create another pivot chart. This time, instead of a stacked bar graph, we'll use a line chart to view the data. Filtering by parent category and years, we will want the Columns value to be "outcome," Rows value to be "Date Created Conversion," and Values to be "outcome." Note that when "outcome" is placed in the Values box, it will be renamed "Count of outcome."

The steps for setting up a pivot line chart are similar to those followed to create stacked bar charts:

1. In the Kickstarter sheet, create a new PivotChart.
2. Rename the new sheet "Outcomes Based on Launch Date."
3. Select Line, and then select the "Line with Markers" style.

### NOTE

Some versions of Excel automatically parse the dates when pivot tables are created. This means that additional columns may be created without additional input from the user.

Line charts are helpful when trying to determine trends. We can see by looking at our new chart that the months of May and June both have a greater success rate. A bar chart wouldn't be able to convey this information in the same manner, so it's important to keep in mind the type of data story we want to tell.

Save this image in your project folder so we can add it to our report later.

## FINDING

The most successful Kickstarter campaigns were started in May. On the other hand, December doesn't seem like a great time to launch a campaign.

© 2020 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.

### 1.4.1 Extra Filters

To continue to help Louise plan her campaign, we need to be more specific with our searches. Using filters, we can update our charts to view data for only theater and plays. We can also search for unique campaigns by using Excel's search function. By searching for unique campaigns, we'll be able to pinpoint the ones most similar to Louise's vision. This helps us research projects similar in scope and type—maybe they have done something right (or wrong), and their specific data will help Louise's project find success.

Now that our worksheet is organized and we're familiar with a few key filtering and sorting tools, we can filter specific data to deepen our analysis. We'll start by looking more closely at subcategories.

## Filter Charts

Go back to the Category Statistics worksheet and filter the chart to display only theater. Follow these steps:

1. In the Row Labels section on the right side of the page, click the arrow to activate the dropdown menu.
2. Click the "Select All" button to deselect everything. Scroll through the list to select only "theater."

We'll repeat these steps for each chart, but with a few minor differences:

- In the "Subcategory Statistics" sheet, we'll select Plays.
- In the "Outcomes Based on Launch Date" sheet, we'll select "theater" for the parent category.

Just by glancing at the data, we can determine that theater is a popular and successful type of campaign overall. By using filters, we can see that theater follows the overall trend: there is a spike of successful campaigns that began in June, but that tapers off by the end of the year. By comparison, the data around technology campaigns reveals a different story. Instead of one large spike, their trend lines are a bit all over the place and less predictable.

---

## Use the Search Function

So far, we've visualized all of the campaigns in our Category Statistics sheet. We've discovered trends in the theater category with our Outcomes sheets. (June seems to be a good month to launch a campaign!) All of this has provided us—and Louise—with information to help plan her campaign. Now let's return to the Kickstarter sheet and focus on a single play: *Foresight*. This is a play from Great Britain that Louise enjoyed. She's curious about the market in Great Britain and would like to learn more about this particular play. We'll filter the dataset to view this campaign as well as other pertinent information.

Filter the Subcategory column to show only plays, and then use the Find function to search for "Foresight." The keyboard shortcut to initiate the Find function is (Mac) Command+F or (Windows) CTRL+F. Alternatively, type "Foresight" into Excel's search bar in the upper-right corner.

Looking at the color of the outcome allows us to quickly determine that the Foresight campaign was successful. It reached 100% of its goal – it even went over by four dollars!

What else can we learn about this campaign from this data? The average donation is surprisingly high, considering there are only 17 backers. Scrolling further, we can also see that the campaign wasn't active for very long—just under a month.

Final score: 100%

 [Retake](#)

## SKILL DRILL

Search for the play *Walken on Sunshine* and note the status of the campaign, average donation, duration, and backers.

## 1.4.2 VLOOKUP

**Louise** was inspired by five plays she saw at the Edinburgh Festival Fringe and wants to know how they were funded. We'll help Louise learn more about these plays using a new Excel function: VLOOKUP.

Louise enjoyed five plays at the Edinburgh Festival Fringe: *Be Prepared*, *Checkpoint 22*, *Cutting Off Kate Bush*, *Jestia and Raedon*, and *The Hitchhiker's Guide to the Family*. She wants to find out more about how these plays got started and whether they were funded by Kickstarter.

We could use Command+F or CTRL+F to search for these titles in the worksheet, but there's a more efficient method that will allow us to search for these titles and gather information about the plays' campaigns simultaneously: VLOOKUP.

**VLOOKUP** lets us pull specific columns from our main dataset into a new sheet without having to search for each column and then copy and paste the data. This way, we can pull only the data points we're interested in. For example, if we only want to see the blurb of the play, we can tell Excel to pull only the data in the blurb cell for that play.

|      | A      | B  | C  | D            |
|------|--------|--|--|--------------|
| 3389 |        | LIBERTY! EQUALITY! AND FIREWORKS!... A Civil Rights Play | Pollyanna just completed an extremely successful run of this new educational play and wants to tour to more under-served communities.  | \$ 3,000.00  |
| 3390 |        | 3388 ICONS   | ICONS is a unique new play about the Amazon warrior women from Greek myth and re-imagines them from a contemporary female perspective. | \$ 1,500.00  |
| 3391 |        | 3389 Chimera Ensemble Productions Fund                   | Chimera Ensemble is launching 2 inaugural theater productions, and we need support to do high quality work!                            | \$ 10,000.00 |
|      | Sheet1 |  |  |              |

To use VLOOKUP, first get all of the titles in a new sheet, which we'll name "Edinburgh Research." In our new sheet, let's add headers: column A will be named "Name" and column B will be named "blurb". Next, write the list of plays in column A. Be sure to spell the titles correctly and use proper capitalization. Here's our list:

- Be Prepared
- Checkpoint 22
- Cutting Off Kate Bush
- Jestia and Raedon
- The Hitchhiker's Guide to the Family

We'll start using the VLOOKUP formula in column B. In cell B2, type the following formula, then press Enter to run it:

```
=VLOOKUP(A2, Kickstarter!B:C, 2, FALSE)
```

You should now see the blurb for *Be Prepared* in cell B1. But what is VLOOKUP actually doing?

1. We're starting the formula with =VLOOKUP.

2. Excel uses the value in A2 ("Be Prepared") to look for data for that campaign.
3. We tell Excel to look for this data in the Kickstarter worksheet.
4. It begins looking in column B; then, it searches the next column, column C, for the data in the same row.
5. Excel will then reference the data from the second column, "blurb" (B is the starting column, so it's counted as the first). Then, Excel will retrieve that data and display it in the new table where we executed VLOOKUP.
6. FALSE tells Excel that we're interested in only exact matches.

Copy the formula into B3 through B6 to get the blurbs for all five plays. Let's also add what each play's goal was and the amount pledged. In the new table in the Edinburgh Research worksheet, first update our headers by adding "Goal" and "Pledged" to columns C and D. To add additional values to the table, paste the same formula into cell C2, and then update its range to B:C. Originally, we told Excel to search for matching data in columns B through C (B:C), but now we want to search further down the same row, through column E instead.

It can be a little confusing, so let's work through this one together. In C2, use VLOOKUP to search for the play we typed into cell A2. The VLOOKUP formula starts in the left-most column of B:E in the Kickstarter worksheet, and then returns the value from the third column. The updated formula will look like this:

```
=VLOOKUP(A2, Kickstarter!B:E, 3, FALSE)
```

Don't forget to specify exact matches.

Now, copy the formula into cells C2 through C5. We can use the same formula to get the amount pledged into column D by adjusting the VLOOKUP formula to take the value from the fourth column.

Which of the following statements is correct?

- Be Prepared* had a goal of \$2,000. ✓
- Checkpoint 22* had a goal of \$1,500.
- Cutting Off Kate Bush* had a goal of \$1,000.
- Jestia and Raedon* had a goal of \$1,500.
- The Hitchhiker's Guide to the Family* had a goal of \$5,000.

Response-specific feedback:

Correct. This is a true statement.

[Finish ▶](#)

## SKILL DRILL

In column E, use VLOOKUP to search for the average donation. Then, in column F, use VLOOKUP to search for the number of backers.

1.5.1

## Measures of Central Tendency

**Adding** statistical components will deepen our analysis for Louise. Statistics provide an unbiased view of the data and make conclusions based on calculations rather than gut feelings. With the addition of statistics, Louise will be able to make decisions about her campaign with confidence. We'll start with the measures of central tendency: mean, median, and mode.

So far, we've compiled some solid information to present to Louise to help inform her campaign strategy. We've organized and sorted the data, as well as created visualizations that lend strength to our analysis. Now we'll use statistics to beef up our report even more.

When we talk about the statistics of a data set, we're generally concerned with how the data is **distributed**. Are data points clustered around one value, or is the data more spread out? Statistical measures distill information about the distribution of data into a single number. The first measures we're going to look at are **measures of central tendency**. **Central tendency** refers to the tendency of data to be toward the middle of the dataset. The three key measures of central tendency are the mean, median, and mode.

## Mean, Median, and Mode

The **mean** is the sum of the data divided by the number of data points. You can think of the mean as answering the question “If every data point contributed the same amount, what would that amount be?” For example, if you and two friends all chipped in to buy a pizza, and you put in \$12, one friend put in \$7, and the other friend put in \$5, the mean cost would be calculated this way:

$$(12 + 7 + 5) / 3 = 24 / 3 = \$8.$$

The **median** answers the question “Where is the midpoint of the data?” Also known as the 50th percentile, the median is the value that splits the data into two equal halves: 50% of the data is lower than the median, and 50% of the data is higher. To calculate the median, sort the data points in order, and then locate the point in the middle. For example, if the grades on a quiz are 82, 79, 79, 77, 70, 90, 71, 86, 83, first we would put them in order:

70, 71, 77, 79, 79, 82, 83, 86, 90

Since there are 9 scores, the midpoint will be the 5th score after sorting. There are 4 scores above it and 4 scores below it:

70, 71, 77, 79, **79**, 82, 83, 86, 90

The median score on the quizzes is **79**.

### SKILL DRILL

Make a small dataset (fewer than 10 data points) with a median of 50 and a mean of 55.

The **mode** answers the question “What value shows up the most?” This question can be trickier than it seems. Take our quiz scores example. In

this example, 79 shows up more than any other score, so 79 is the mode. Let's say there is a student who was absent that day and then takes a makeup quiz, scoring 82. Then there would be two students who scored 82, and two students who scored 79. In this case, the data has two modes: 79 and 82. A dataset can have any number of modes—or even no mode!

If one large number is added to a dataset of relatively small numbers, which measure of central tendency would change the most?

- Mean
- Median
- Mode
- The mean, median, and mode would all be affected equally.



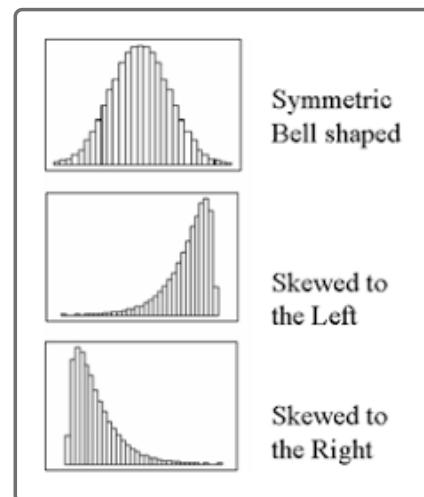
Response-specific feedback:

Correct. The mode won't change at all, the median might shift slightly, and the mean will increase significantly.

Finish ►

### IMPORTANT

When the mean and median are close to each other, the data is roughly **symmetric**: half the data is above the mean, half the data is below. If the mean is significantly different than the median, the data is **skewed**, meaning that some number of extreme values are pulling the mean higher or lower. If the mean is much higher than the median, the data is **skewed to the right**. If the mean is much lower than the median, the data is **skewed to the left**.



**Skewness** is a statistic that quantifies how skewed, or asymmetrical, a distribution is.

This is an advanced concept in statistics that you won't need to worry about in this course.

---

## Use Measures of Central Tendency with Crowdfunding Data

Let's see how measures of central tendency work in practice. We'll consider Kickstarter campaigns for plays in the U.S. and compare the statistics for the campaigns that succeeded versus those that failed.

Follow these steps:

1. Clear any filters on the dataset.
2. Apply the following filters:
  - Filter on subcategory for "plays."
  - Filter on country for "U.S."
  - Filter on outcome for "successful."
3. Copy the filtered dataset and paste it into a new worksheet named "Successful U.S. Kickstarters."

### SHOW PRO TIP

4. Return to the Kickstarter worksheet and change the filter on outcome to "failed."
5. Copy and paste this dataset into a new worksheet and name it "Failed U.S. Kickstarters."
6. Create another worksheet and name it "Descriptive Statistics."

We'll be pulling data from each of these new worksheets and performing a few measures of central tendency on them—which is just a fancy way of saying that we'll be finding the mean and median for each dataset's (the failed and successful U.S. campaigns) goal and pledged columns.

In the new worksheet, create a table like the one below to hold our results. This way, we'll be able to easily compare the goals and pledges for failed and successful campaigns alike. By comparing the two, we'll be able to determine whether there are any trends between the goals and pledges in successful or failed campaigns.

| A1 | B              | C      |
|----|----------------|--------|
| 1  | Successful     | Failed |
| 2  | Mean Goal      |        |
| 3  | Median Goal    |        |
| 4  |                |        |
| 5  | Mean Pledged   |        |
| 6  | Median Pledged |        |

## Failed Campaigns

In B2, enter the formula used to find the mean of a dataset:

```
=AVERAGE( 'Successful US Kickstarters' !D:D)
```

### IMPORTANT

Excel doesn't have a MEAN function; it uses the less precise AVERAGE function to calculate the mean. Statisticians use the term "average" in many contexts, but prefer to be precise in their calculations. You can

tell when a statistician is using Excel by the grumbling noise they make when they have to type AVERAGE instead of MEAN.

This formula tells Excel that we're looking for the average number in a dataset, but we're only looking for the average amount of "Successful U.S. Kickstarters." By adding (D:D), we're also pinpointing which column we're applying the formula to the Goals column. The colon indicates a range of data, so by adding D:D to the formula, we're specifying the entire column.

Let's add a few more to our new worksheet.

- In C2, enter the formula `=AVERAGE('Failed US Kickstarters'!D:D)`.
- In B3, enter the formula `=MEDIAN('Successful US Kickstarters'!D:D)`.
- In C3, enter the formula `=MEDIAN('Failed US Kickstarters'!D:D)`.

## SKILL DRILL

In cells B5 and C5, enter the AVERAGE formula.

Remember to use the correct columns for each: B5 is finding the average pledged for successful campaigns, and C5 is finding the average pledged for failed campaigns.

In cells B6 and C6, enter the MEDIAN formula. Be sure to pull data from the Pledged column for these as well and filter them as either successful or failed.

Format the cells B2:C6 to Currency with no decimals.

You should now see a table that looks like this:

|   | A              | B          | C        |
|---|----------------|------------|----------|
| 1 |                | Successful | Failed   |
| 2 | Mean Goal      | \$5,049    | \$10,554 |
| 3 | Median Goal    | \$3,000    | \$5,000  |
| 4 |                |            |          |
| 5 | Mean Pledged   | \$5,602    | \$559    |
| 6 | Median Pledged | \$3,168    | \$103    |

This simple table allows us to determine a few things. For one, failed Kickstarter campaigns have much higher fundraising goals than successful Kickstarter campaigns. Louise is asking for more than twice the average successful Kickstarter goal, so this isn't great news for her campaign. In addition, the mean and median pledged amounts are much lower than the successful pledges, which indicates that failed Kickstarter campaigns are unsuccessful for reasons other than asking for too much money. In other words, if the failed projects were also getting a median pledge amount of around \$3,000, it's possible that those that failed just asked for too high of a price. Since the median is much lower, there must be another factor keeping people from pledging to those unsuccessful projects.

## SKILL DRILL

Make the same table for two other subcategories. (They don't need to be in the Theater category.) How do the distributions differ, if at all?

## 1.5.2 Measures of Spread

We've now learned about measures of central tendency and found the mean and median of our dataset. Now let's take our analysis a step further and help Louise measure the spread of the dataset. We'll be looking at the range of the dataset as well as adding more statistics to our analysis: standard deviation and variance.

Measures of central tendency distill a lot of information about the distribution of a dataset down to one number. However, two datasets can have the same mean or median but still look very different—that is, the spread of data between the two datasets can vary quite a bit. When considering the distribution of a dataset, we also want to have measures of its spread. **Measures of spread** include range, variance, standard deviation, and quartiles.

### Range

The simplest measure of spread is the range of a dataset. The **range** is the difference between the maximum value of the dataset and the minimum value of the dataset. For our purposes, the range does not capture as

much information as we'd like. What we would really like to know is roughly how far each data point is from the center, or mean, or how much of the data is near the center.

---

## Variance

**Variance** is a measure of how far data points are from the center, or mean. To calculate the variance, do the following:

1. Subtract the mean from each data point.
2. Square the difference so that it's positive.
3. Take the average of those squared differences.

Because we've taken the average of the *squared* differences, the unit of variance doesn't quite match our dataset. To get the unit to match, we take the square root of the variance to standardize it, or get the **standard deviation**. Standard deviation is often represented with a lowercase sigma ( $\sigma$ ). You'll also see variance represented as the standard deviation squared ( $\sigma^2$ ). Let's look at an example.

Imagine we have a dataset of five backers. We'll signify that this is a dataset by placing the data within brackets: [1, 3, 6, 7, 8], which then makes it into a "set" of numbers. How do we find the standard deviation? Let's begin working through the standard deviation equation:

$$\sigma = \sqrt{\frac{\sum (X - \bar{X})^2}{n-1}}$$

- $\sigma$ , lowercase sigma, is the symbol for standard deviation.
- $\Sigma$ , uppercase sigma, is the symbol for summation or the sum.
- $X$ , represents each point of data in the dataset.
- $\bar{X}$ , represents the mean of the dataset and is pronounced "x-bar."

- $n$ , represents the total number of points in the dataset.

This equation looks fairly complicated, so let's talk through what's happening.

1. Find the mean.
2. For each number in the dataset, subtract the mean and square the result.
3. Find the mean of these new numbers.
4. Take the square root of the mean.

Let's apply the equation to our set:

1. Find the mean:  $(1 + 3 + 6 + 7 + 8) / 5$  (the sum the values divided by the number of values), or  $25/5 = 5$ .
2. Next, we find the deviations. The deviations from the *sample* mean are
  - $(1 - 5) = -4$
  - $(3 - 5) = -2$
  - $(6 - 5) = 1$
  - $(7 - 5) = 2$
  - $(8 - 5) = 3$

Ideally, we'd like to know the deviations from the actual population mean, but because we don't know the actual population mean, these deviations have a subtle and slight bias to them. We'll correct that bias in the next step.

3. Find the variance. First, square all of those deviations; this way, we will always be working with positive numbers. This results in 16, 4, 1, 4, and 9. If we take the average of these values, we'll get a slightly smaller variance than the actual population. To correct for this bias, we instead divide by the number of samples minus 1 (if you're

curious, this is known as Bessel's correction). Thus, the unbiased average of these values is  $(16 + 4 + 1 + 4 + 9) / (5-1) = 34 / 4 = 8.5$ .

4. The square root of 8.5 is about 2.92, which is the standard deviation.

Another method for measuring the spread is to calculate the interquartile range. After organizing the dataset from lowest value to highest value, we can break it into four separate parts known as quartiles.

---

## Quartiles

Like medians, **quartiles** are percentiles. The lower quartile is the 25th percentile, that is, 25% of the data is less than the lower quartile. Similarly, the upper quartile is the 75th percentile, so 75% of the data is less than the upper quartile. You may also see these referred to as the 1st and 3rd quartiles. (The 2nd quartile is the median, so that one already has its own fancy name).

The difference between the upper and lower quartiles is known as the **interquartile range** (IQR). The IQR gives us a sense of how far out you can go from the mean to get 50% of the data.

### IMPORTANT

Determining the interquartile range is done with the following formula:

$$\text{IQR} = \text{Q3} - \text{Q1}$$

That is, the IQR is equal to the third quartile minus the first quartile. When used to describe a list of data, the IQR tells us how the data is spread around the median.

To learn more about IQR, read [this article that provides additional examples and explanations for the uses of IQR](#) (<https://www.thoughtco.com/what-is-the-interquartile-range-rule-3126244>).

To see these concepts in action, let's add measures of spread to our "Descriptive Statistics" worksheet. Add new rows for the standard deviation, upper and lower quartiles, and IQR.

Here's how your updated table should look:

|                               | Successful | Failed   |
|-------------------------------|------------|----------|
| Mean Goal                     | \$5,049    | \$10,554 |
| Median Goal                   | \$3,000    | \$5,000  |
| Standard Deviation of Goal    |            |          |
| Upper Quartile of Goal        |            |          |
| Lower Quartile of Goal        |            |          |
| IQR of Goal                   |            |          |
|                               |            |          |
| Mean Pledged                  | \$5,602    | \$559    |
| Median Pledged                | \$3,168    | \$103    |
| Standard Deviation of Pledged |            |          |
| Upper Quartile of Pledged     |            |          |
| Lower Quartile of Pledged     |            |          |
| IQR of Pledged                |            |          |

The function to calculate the standard deviation of a population in Excel is **STDEV.P**. (The other option is **STDEV.S**, which calculates the standard deviation based on a sample of the whole population. There's a subtle difference between these formulas (one is for the entire population of a dataset while the other is for a sample of the whole) that statisticians care about, but we're going to ignore it. Don't tell any of your statistician friends.

We'll be using the same range and worksheet data as we did with the AVERAGE formula, so the STDEV.P formula we enter into B4 is

=STDEV.P('Successful US Kickstarters'!D:D).

To calculate the upper and lower quartiles, use the **QUARTILE.EXC** function. QUARTILE.EXC takes two arguments: the first argument is the data array, and the second is the quartile to be calculated. For the upper quartile, put 3:

```
=QUARTILE.EXC('Successful US Kickstarters'!D:D, 3)
```

For the lower quartile, put 1:

```
=QUARTILE.EXC('Successful US Kickstarters'!D:D, 1)
```

The IQR cell will be the difference between the upper and lower quartiles, so the two cells would be subtracted. In B7, type `=B5-B6`.

## SKILL DRILL

Using the same formulas, update them to complete the standard deviation and quartile calculations in column C.

Your table should now look like this:

|    |                               | Successful | Failed   |
|----|-------------------------------|------------|----------|
| 1  |                               |            |          |
| 2  | Mean Goal                     | \$5,049    | \$10,554 |
| 3  | Median Goal                   | \$3,000    | \$5,000  |
| 4  | Standard Deviation of Goal    | \$7,749    | \$21,968 |
| 5  | Upper Quartile of Goal        | \$5,000    | \$10,000 |
| 6  | Lower Quartile of Goal        | \$1,500    | \$2,000  |
| 7  | IQR of Goal                   | \$3,500    | \$8,000  |
| 8  |                               |            |          |
| 9  | Mean Pledged                  | \$5,602    | \$559    |
| 10 | Median Pledged                | \$3,168    | \$103    |
| 11 | Standard Deviation of Pledged | \$8,335    | \$1,331  |
| 12 | Upper Quartile of Pledged     | \$5,699    | \$501    |
| 13 | Lower Quartile of Pledged     | \$1,717    | \$9      |
| 14 | IQR of Pledged                | \$3,982    | \$492    |

## FINDING

Based on these statistics, we can determine the following:

- The mean of each distribution is around the 3rd quartile, so the data follows similar distributions in each subset.
- The standard deviations are larger than the mean, which means everything below the mean is considered “close” to the center.
- Some large values are driving all of these distributions. The standard deviations are all roughly twice the IQR in each distribution, except in the failed Kickstarters, where the standard deviation is closer to three times the IQR. There must be some failed Kickstarters with really high goals!

© 2020 Trilogy Education Services, a 2U, Inc. brand. All Rights Reserved.

### 1.5.3

## Identifying and Addressing Outliers

Now that we've worked on measures of central tendency, variance, and standard deviation, let's find the outliers of the dataset. The concept of outliers is something Louise could use to better plan her campaign by eliminating extreme data points that are not representative of the data we want. For example, there may be large funding goals in the dataset that are for theater-building proposals. Louise just wants funding for her play, not a theater, so she'll be able to eliminate those data points. Let's take a closer look at outliers and how to handle them when working with data.

In datasets, **outliers** are extreme points of data; they can be much larger than the rest of the data or much smaller. But how do we define "extreme"? We can use the tools we've just learned along with some guidelines generally accepted by statisticians. There are two main techniques for determining outliers, and each technique uses a measure of central tendency and a measure of spread. We can use either the mean and standard deviation together, or the median and interquartile range (IQR) together.

## NOTE

Why don't we use variance to determine outliers? We use standard deviation because taking the square root of the variance standardizes the "units" of the variance to match the "units" of the dataset. (This is also why it's called "standard" deviation.)

If we decide to use the first method—mean and standard deviation—the guideline is that any value that is more than 3 standard deviations higher or lower than the mean is considered an outlier. If we decide to use the second method—median and IQR—two guidelines need to be followed:

- Any value greater than the upper quartile plus  $1.5 \times \text{IQR}$  is considered an outlier.
- Any value less than the lower quartile minus  $1.5 \times \text{IQR}$  is considered an outlier.

Considering the IQR rule, which of the following statements is correct?

- Any value equal to the upper quartile plus  $1.5 \times \text{IQR}$  is considered an outlier.
- Any value greater than the upper quartile is considered an outlier.
- Any value greater than the upper quartile plus  $1.5 \times \text{IQR}$  is considered an outlier. ✓
- Any value less than the upper quartile plus  $1.5 \times \text{IQR}$  is considered an outlier.

Response-specific feedback:

Correct. For a data point to be considered an outlier, it must be greater than the upper quartile plus  $1.5 \times \text{IQR}$ .

[Finish ►](#)

Which method do we use? In almost all cases, the IQR rule is preferred. Medians and quartiles are **robust statistics**, which means that they are less sensitive to outliers.

Consider a county with a small population of people making a modest living. Now, imagine a billionaire moves into the county. The median income would barely change, if at all, but the mean would catapult to a much higher value. In fact, if the county is small enough, everyone but the billionaire could end up being “below average” based on the mean.

So if the IQR rule is preferred, why is there a method that uses mean and standard deviation to determine outliers? For one thing, mean and standard deviation can be calculated more quickly. Finding percentiles requires sorting the data, which can be time-consuming with large datasets. The mean and standard deviation can be calculated without sorting data, which means that our computers won’t need to work as hard to perform the calculations.

Now that we can identify outliers, what do we do with them? This is a tricky question. Changing or removing data points changes the story you’re trying to tell with your data. If the identified outliers are a mistake (say, the data was entered with a typo), ideally, we would just want to correct the mistake and leave the data point in the dataset; if that’s not possible, we would have to throw out the data point. However, if an outlier is a legitimate member of the dataset, it’s better to leave it in and tell the full story of the data.

Final score: 100%

 [Retake](#)

#### 1.5.4

## Visualizing Distributions with Box Plots

Louise mentioned earlier that she's also interested in Great Britain's theater market, especially musicals. While she's committed to creating a play in the U.S., she's also interested in researching musicals in Great Britain for a future project with an estimated budget of £4,000. To present Louise with the big picture, we will create box plots using statistical computations.

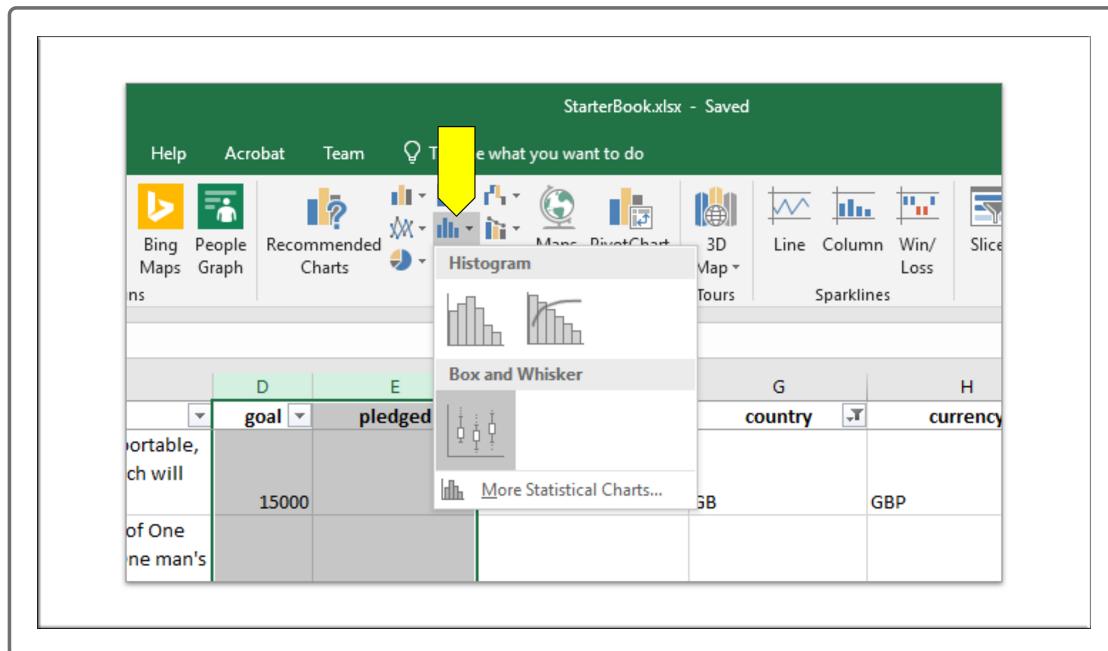
Taken together, measures of central tendency, measures of spread, and identification of outliers tell us about the distribution of our data, but *showing* the distribution makes the data story easier to understand—which has an even greater impact. **Box plots** are an effective way to show large amounts of information about a distribution in a small amount of space.

## Box Plots

We'll use **box plots**, also called **box and whisker plots**, to compare the distribution of campaign goals and the distribution of total amounts pledged for plays in Great Britain. Remember that Louise estimates she'll need to raise £4,000 for her future project.

First, clear all the filters from your data. Set a filter on "country" to only show rows with GB as the country. Then, set a filter on "Subcategory" to only show rows with "musical" as the subcategory.

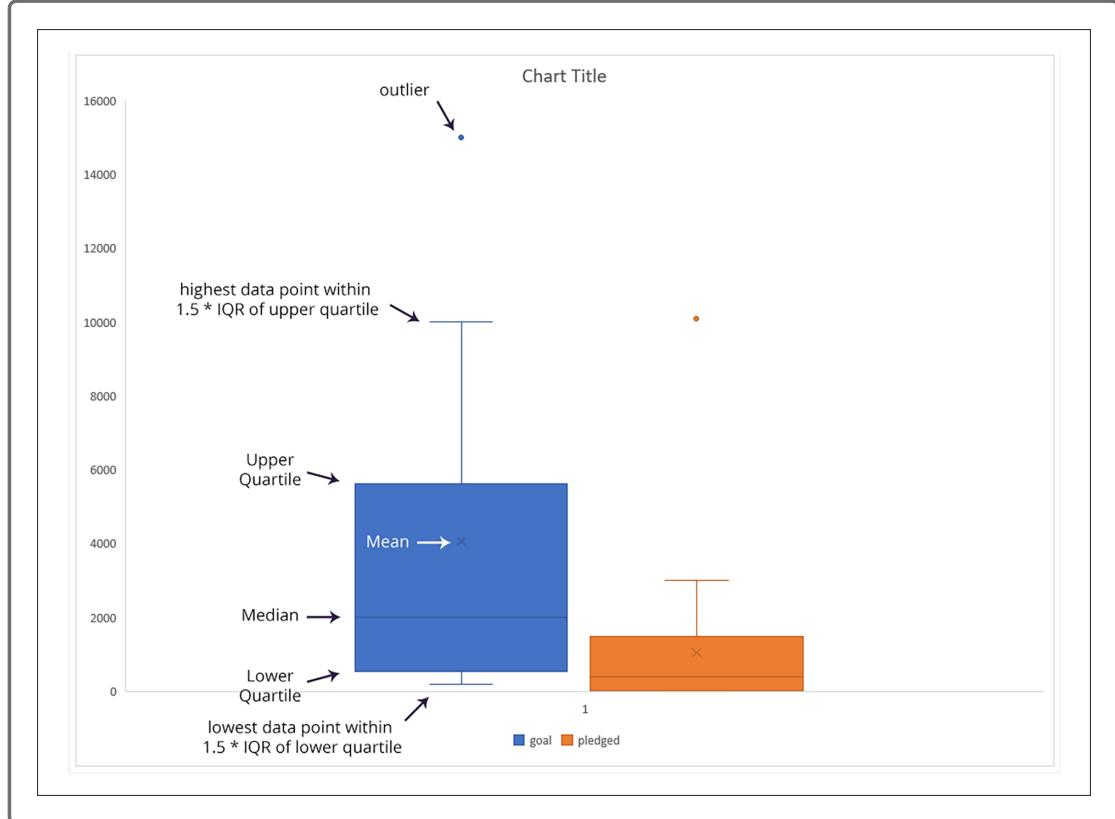
Select the Goal and Pledged columns and then go to the Insert tab. Click the Statistical Charts button and select "Box and Whisker."



Now move the chart to its own sheet by going to the Design tab and selecting the Move Chart button.



How do we read a box and whisker plot? The box shows the interquartile range with a line for the median and an "X" to indicate the mean. The whiskers show the extreme values within 1.5 times the interquartile range. Outliers are represented by labeled dots.



From these plots, we can see that the mean campaign goal is around £4,000. This is outside of the range of outliers for amount pledged, so Louise should probably try to get her play produced for less than £4,000. Half of the campaign goals are less than £2,000, which is just over the 3rd quartile for amounts pledged.

When the mean is considerably higher than the median of a dataset, what does it signify?

- Nothing; there is no statistical relevance when this occurs.
- It means the data has a symmetrical distribution.
- It means the distribution is heavily skewed. ✓
- It means the distribution is only slightly skewed.

Response-specific feedback:

Correct. When the mean is higher than the median, the data is considered to be skewed to the right. When the median is higher than the mean, the data is skewed left.

[Finish ►](#)

What do the data indicate when the lower quartile of pledges starts on the x-axis?

- All of the Kickstarter campaigns for musicals in Great Britain got little to full funding at the completion of the campaign.
- 25% of the Kickstarter campaigns for musicals in Great Britain got no funding at all. ✓
- 25 Kickstarter campaigns in the dataset did not get funding at all.
- 25% of the Kickstarter campaigns in Great Britain got a negative amount of funding for their projects.

Response-specific feedback:

Correct. Because each quartile is a quarter of the data, when the lower quartile sits exactly on the x-axis, it means that a quarter of the data does not have a value.

[Finish ►](#)

### 1.6.1 Using GitHub

Now that we've completed our analysis, we need to create a report to showcase our findings. We also need a place to host our report and images, so we'll create an account on GitHub and set it to share our report.

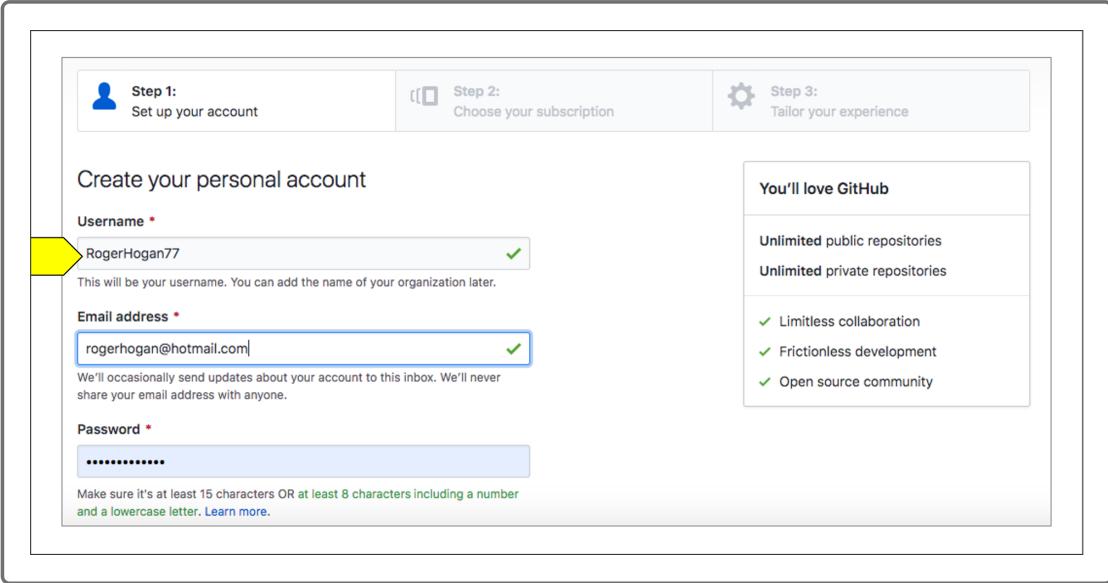
We've gleaned a lot of helpful information and insights from this data. We've organized, sorted, filtered to narrow down our scope, and then we created visualizations and performed statistical computations. Now it's time to put it all together and share our findings with Louise. For this step we'll be using GitHub.

## Create a GitHub Account

GitHub is an online hosting service for storing files, images, and other content online. GitHub is quick, convenient, and widely used by developers of all types.

To get started, go to the [GitHub website](https://github.com/) (<https://github.com/>) and create an account, or log in if you already have one. Note that your username is

public and will be tied to each repository you create.



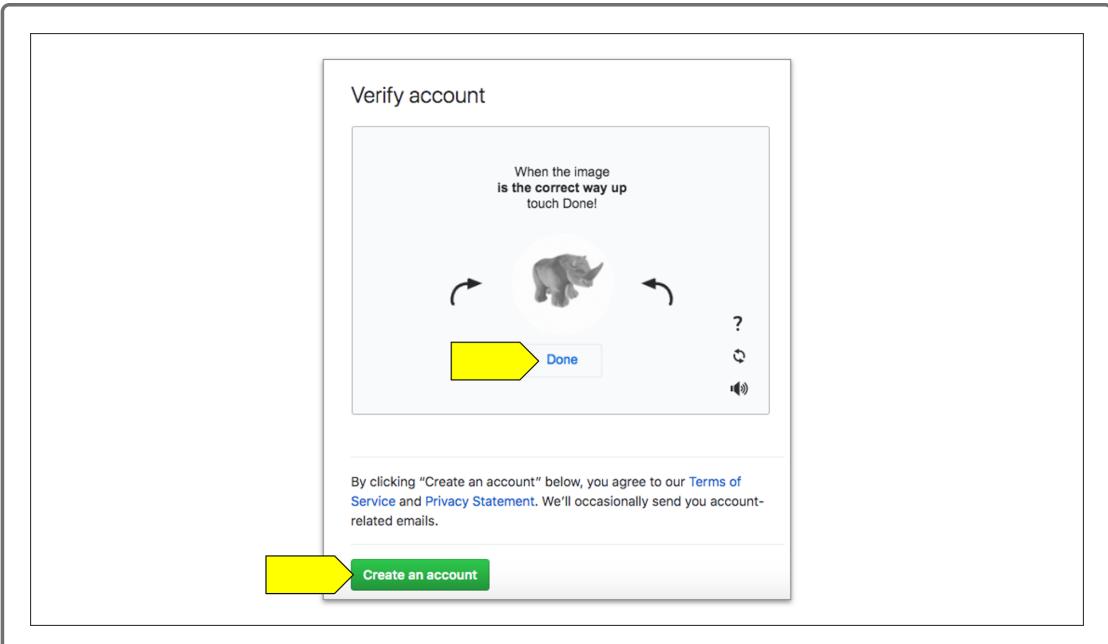
The screenshot shows the first step of creating a GitHub account. It's titled "Create your personal account". There are three tabs at the top: "Step 1: Set up your account", "Step 2: Choose your subscription", and "Step 3: Tailor your experience". The "Step 1" tab is active. The form fields are:

- Username \***: RogerHogan77 (highlighted with a yellow arrow)
- Email address \***: rogerhogan@hotmail.com
- Password \***: (redacted)

Below the form, there's a note: "We'll occasionally send updates about your account to this inbox. We'll never share your email address with anyone." To the right, a sidebar titled "You'll love GitHub" lists benefits:

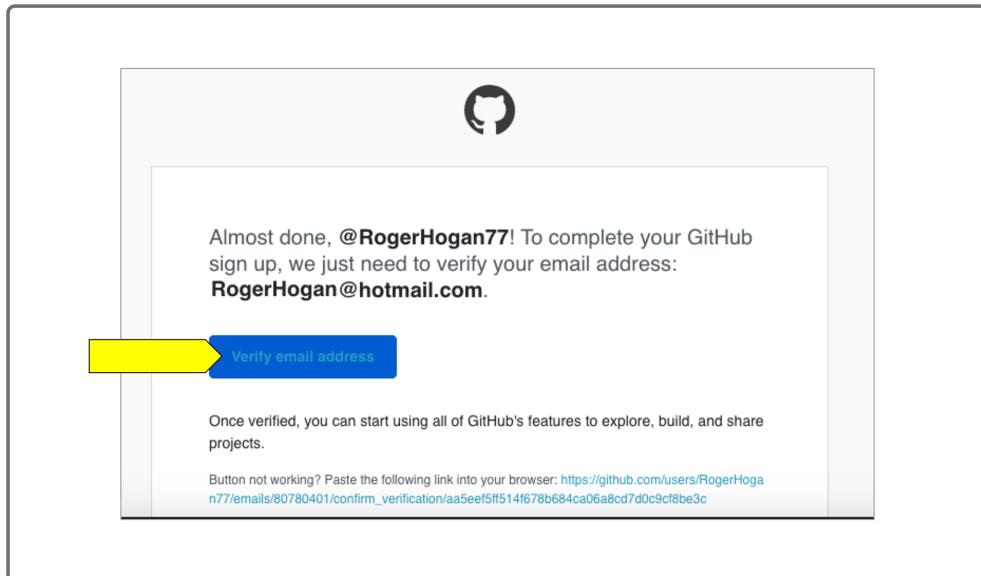
- Unlimited public repositories
- Unlimited private repositories
- ✓ Limitless collaboration
- ✓ Frictionless development
- ✓ Open source community

After you enter your username, email, and password, you will be sent to a page to verify your account. This is where GitHub will confirm that you aren't a bot (an automated program). This verification step will likely entail some kind of short puzzle, such as using arrows to rotate an image.

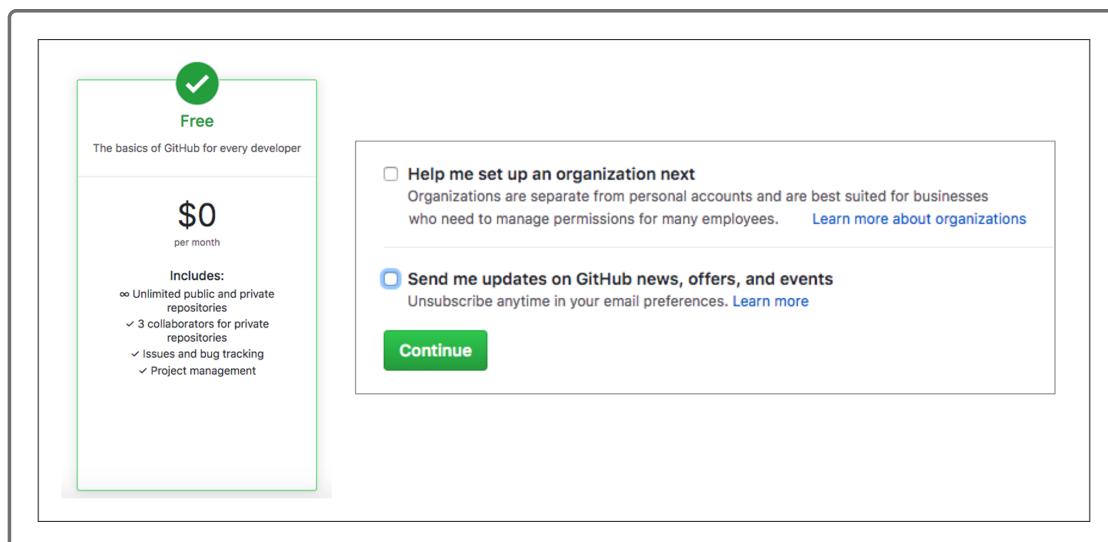


The screenshot shows the second step of verifying the account. It's titled "Verify account". The main content area contains a puzzle: a rhinoceros image with two curved arrows around it, one pointing clockwise and one counter-clockwise, and a "Done" button. A yellow arrow points to the "Done" button. Below the puzzle, there's a note: "When the image is the correct way up touch Done!". At the bottom, there's a note: "By clicking "Create an account" below, you agree to our [Terms of Service](#) and [Privacy Statement](#). We'll occasionally send you account-related emails." A yellow arrow points to the "Create an account" button.

You'll also receive an email from GitHub to verify your email address. Don't skip this step!



After completing the verification, you'll select a subscription type. Be sure to select the free subscription option that is outlined in green. Scroll down to uncheck the box allowing your email address to be added to GitHub's mailing list (unless you want to receive emails from GitHub). Then click Continue.



The next page you're directed to is a survey to help tailor your GitHub experience. It isn't necessary to complete this form, and you can click

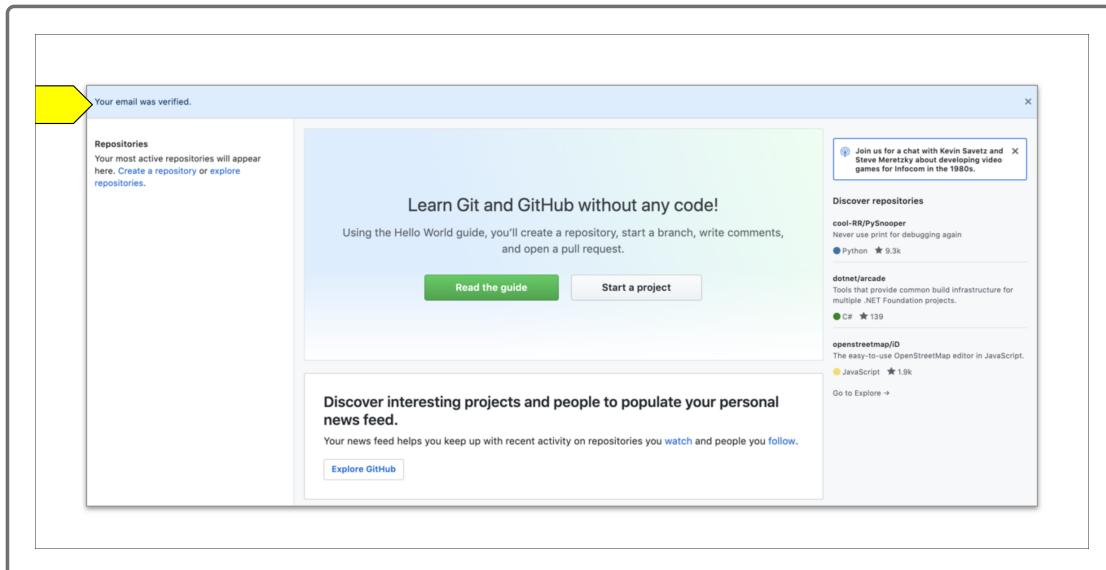
“Skip This Step” to finish creating your account.

What are you interested in?

What languages, frameworks, industries, or disciplines are you interested in?  
e.g. webapp, python, js13kgames

**Submit** skip this step

If you haven’t already verified your email address, take the time to complete this step now. Navigate to your email inbox, open the message from GitHub, and click the provided link to verify your account. This will open a new window that will automatically open in your web browser:



Now you’re ready to create your first repository!

### IMPORTANT

A **repository**, or repo for short, is the project-specific holding space for the files used in a project. Another great thing about GitHub is that

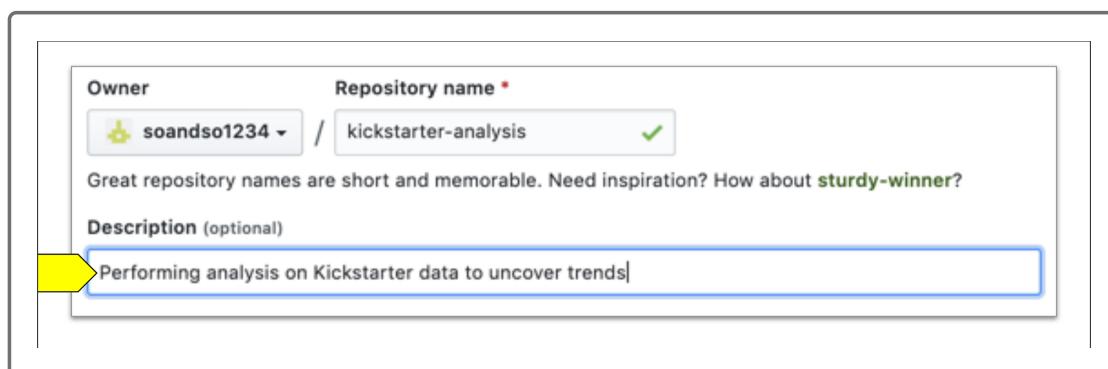
there are *unlimited* repositories per user! Many companies use GitHub because it is a perfect collaboration hub for development teams.

## Create a Repository

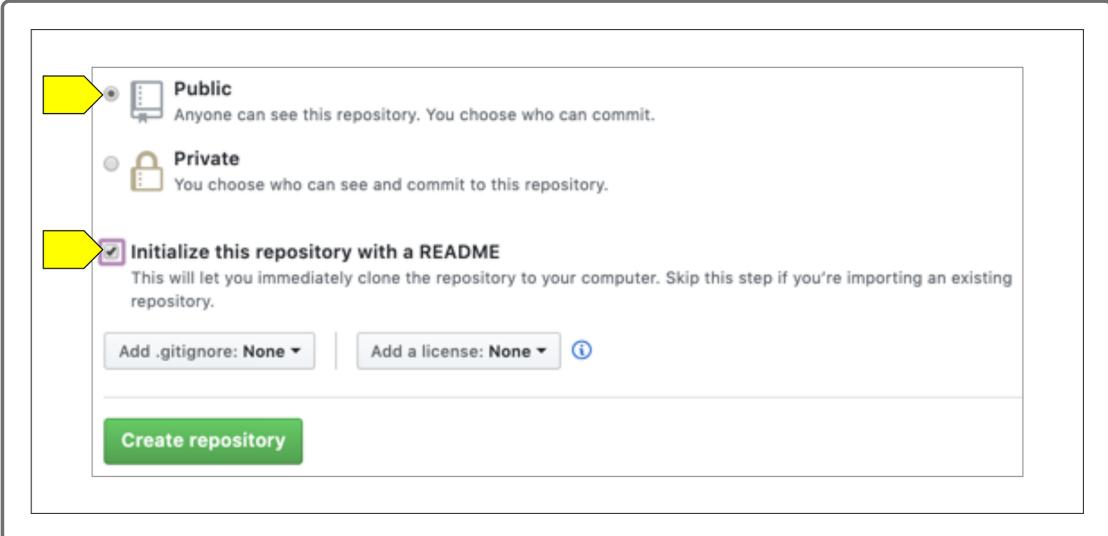
Now that we're logged in, we'll create a new repository. Click the "Start a Project" button to go to the "Create a Repository" page, which is where we'll begin to set up our project.

Each repo should have a unique name. You can use GitHub's random repo name generator, but it's better to use a name that describes the project. Let's name our repo `kickstarter-analysis` to get started.

Enter a brief description of the project. Keep it short and succinct; the point of the description is to help others understand the purpose of the repo. A project description will also help jog your memory when you revisit old repositories weeks or months down the line. See the following example:



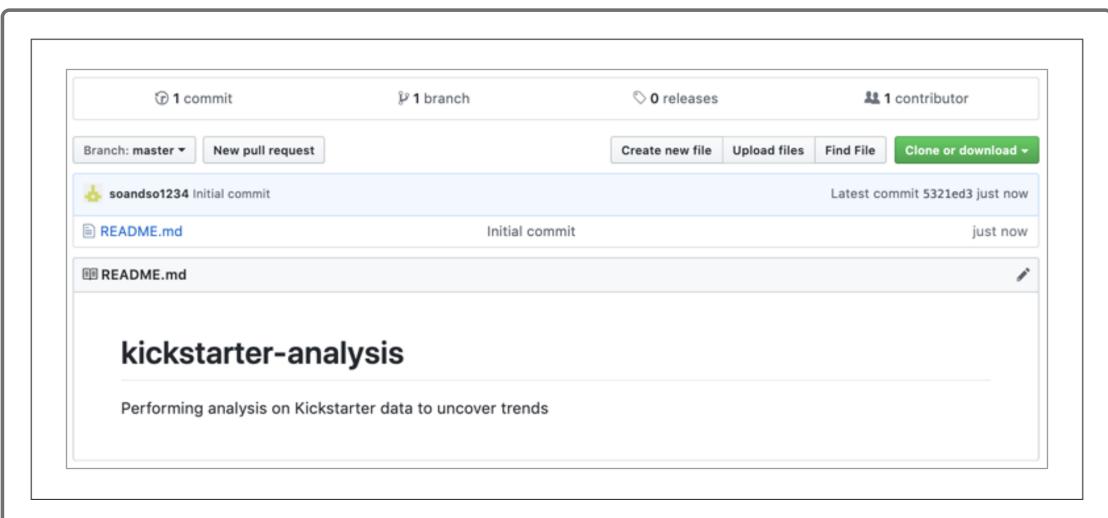
Once you've given your repo a name and added a description, check the box that says "Initialize this repository with a README." Keep the default setting to make the repo public.



## NOTE

A README file contains an explanation or description of the project so that others know how to use the information provided. These files are typically written in markdown format and will be automatically displayed when someone visits your repository. We'll cover markdown in greater detail later.

Click the green Create Repository button to make the new repo official. You will then be automatically redirected you to a new (and somewhat empty) page.



# Upload Files

Now we'll upload our Excel spreadsheet with the completed analysis. Click the Upload Files button in the upper-right corner of the screen. You can either drag and drop the appropriate .xls file or click "Choose Your Files" to manually navigate through your directory to locate the correct file.

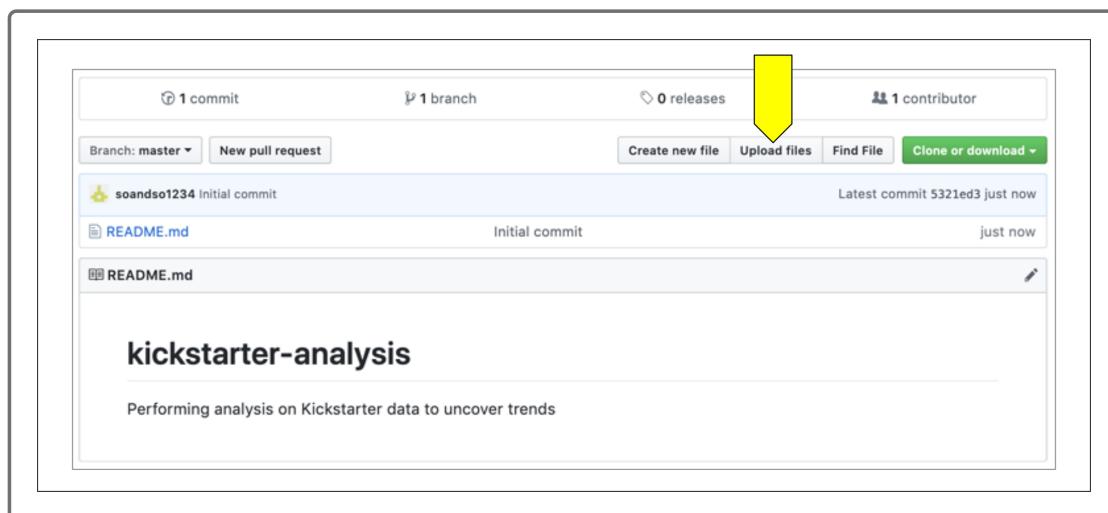
If your file size is more than 25 MB, then you'll run into an error when you try to upload your work. You'll need to compress your file first (also known as zipping it), and then upload the zipped file.

For instructions on how to compress your Excel workbook file, visit one of the following links:

macOS (<https://www.imore.com/how-compress-file-your-mac>)

Windows (<https://support.microsoft.com/en-us/help/14200/windows-compress-uncompress-zip-files>)

Later, you'll learn more about how to upload large files without zipping them.



Note that multiple files can be uploaded at once, which will save some time as we continue to prepare our analysis report.

Next, we'll "commit" our changes. A **commit** is part of the GitHub process that captures why a change is made. Any action—from adding a file, updating it, or deleting a part of it—is reflected through commit messages. This feature is similar to saving our work in Excel, and it's easy to revert back to or reference a previous version if needed.

A commit message should be descriptive but to the point. For example, when uploading our data, the message could read, "Uploading completed Kickstarter analysis." This way, we aren't inundated with text if we check our commits; a quick glance will give us the pertinent information.

Make sure the "Commit directly to master branch" button is selected, and then click "Commit changes."

### IMPORTANT

**Branching** is a common component of working in GitHub. The **master** branch is the main branch that showcases your completed project. For a large-scale project, different aspects of the project will each have its own branch. This is especially useful when working in teams.

For example, Branch A was created to add another component to a project. This branch contains a copy of the master branch. When a developer adds items to this branch, the new changes are isolated within Branch A. This allows developers to iron out any kinks that may come up when making file changes while keeping the master branch intact.

Now that there are two branches, master and Branch A, how do new additions get integrated? After the changes in Branch A have been checked for errors and quality, they are merged into the master branch.

This process, called **merging**, pulls in only the newly added or adjusted files.

We can view our files and commit messages on the main page of the repo. The README, which is automatically displayed, doesn't have any content—yet! Let's fix that.

---

## Edit the README

A **README** file is the face of a repository: it provides an explanation of the project. It can be customized using **markdown** to provide a pleasing aesthetic.

**Markdown** is a plain text editing language used to make plain text prettier. It can insert images, create hyperlinks, create bulleted lists, and many other customizations. Through GitHub, markdown is even rendered as a webpage, though it doesn't possess as much customization as a page powered with HTML. Commonly found in forums and README files, markdown is a language that is fairly simple to pick up and put into practice.

We can edit the README markdown by clicking the button with a pencil icon on the main repository page. The very first line in our new editor should read `# kickstarter-analysis`. The pound sign, or hashtag, signifies a first-level header, or the title of the document. “Kickstarter-analysis” is the repository name, which is integrated into the document automatically by GitHub.

This file is where we'll compose our analysis report to be shared with others—or, in this case, Louise. Here's a breakdown of basic markdown syntax:

- **Headers** are signified with a #. The number of hashtags indicates the level of the header. For example, ### is a third-level header.
- **Bullets** are added in three ways: a typical bullet with an asterisk, a numbered bullet with numbers, or a hyphenated bullet with a hyphen.
- **Images** can be added with the following syntax: `![image_name](path/to/image_name.png)`.

- **Hyperlinks** to relevant files are added in a similar manner: [filename] (path/to/filename.xlsx).
- **Line breaks** are added by using three consecutive hyphens: ---.

Refer to the [GitHub Help page](https://help.github.com/en/articles/basic-writing-and-formatting-syntax) (<https://help.github.com/en/articles/basic-writing-and-formatting-syntax>) for additional information about creating documentation using markdown format.

Now we're ready to flesh out our analysis!

Begin by updating the header to a more appropriate title like "An Analysis of Kickstarter Campaigns." Below the header, add a brief description of the project.

In addition to your findings, you should also include charts and graphs (visualizations) in your report. Visualizations help communicate the story the data is telling by making it more understandable and coherent.

At the end of the report, add any recommendations you have for Louise based on your findings. Then add a descriptive commit message and click "Commit changes." This descriptive README file adds depth to the repository as well as your analysis. It's ready to be shared!

Return to the main page of your GitHub repository, highlight the address, and copy it. This link, which should read

<https://github.com/yourusername/kickstarter-analysis>,  
(<https://github.com/yourusername/kickstarter-analysis>), can now be emailed to our hopeful playwright, Louise, so she can begin her Kickstarter preparations.

# Module 1 Challenge

[Re-submit Assignment](#)

---

**Due** Jan 19 by 11:59pm    **Points** 100    **Submitting** a text entry box or a website url

---

Welcome to the first weekly challenge in this course. Your weekly challenges provide an opportunity for you to demonstrate the skills you have learned in the module. These challenges are required and will be graded.

In this challenge, you will continue to exercise your Excel prowess! This time, you will create two new analyses: outcomes based on goals and outcomes based on launch date.

---

## Background

Louise's play *Fever* came close to its fundraising goal in a short amount of time. How many other Kickstarter campaigns were able to do this as well? In this challenge, you'll conduct a data analysis to answer this question and determine whether the length of a campaign contributes to its ultimate success or failure.

---

## Objectives

The goals of this challenge are for you to:

- Use filters and Excel formula to create new datasets
- Create visualizations using Excel charts and pivot tables
- Interpret the summary data provided from the visualizations

# Instructions

Open Excel and create two sheets. Name them "Outcomes Based on Launch Date" and "Outcomes Based on Goals."

## In the "Outcomes Based on Goals" sheet, do the following:

1. Create 8 new columns with the following headings:

- Goal
- Number Successful
- Number Failed
- Number Canceled
- Total Projects
- Percentage Successful
- Percentage Failed
- Percentage Canceled

2. In the Goal column, create 12 new rows to sort each project. This will hold each goal range we're charting:

| Goal               |
|--------------------|
| Less Than 1000     |
| 1000 to 4999       |
| 5000 to 9999       |
| 10000 to 14999     |
| 15000 to 19999     |
| 20000 to 24999     |
| 25000 to 29999     |
| 30000 to 34999     |
| 35000 to 39999     |
| 40000 to 44999     |
| 45000 to 49999     |
| Greater than 50000 |

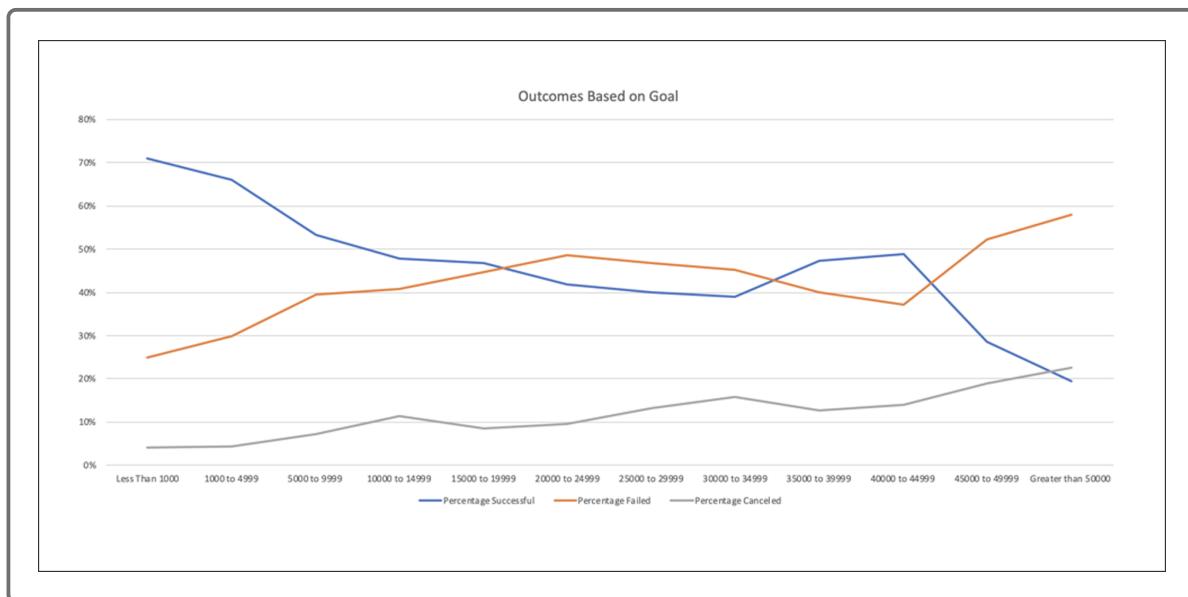
3. Use `COUNTIFS()` to count the number of successful, failed, and canceled projects within the newly defined ranges. (Filter the Kickstarter sheet to show only the Plays subcategory, otherwise, our scope includes all campaigns instead of the ones we are interested in.) Populate the appropriate columns.

4. Add each of the values in the Number Successful, Number Failed, and Number Canceled columns to populate the Total Projects column.

5. Use division to find the percentage of projects that fall into each category (percentage successful, percentage failed, and percentage canceled).

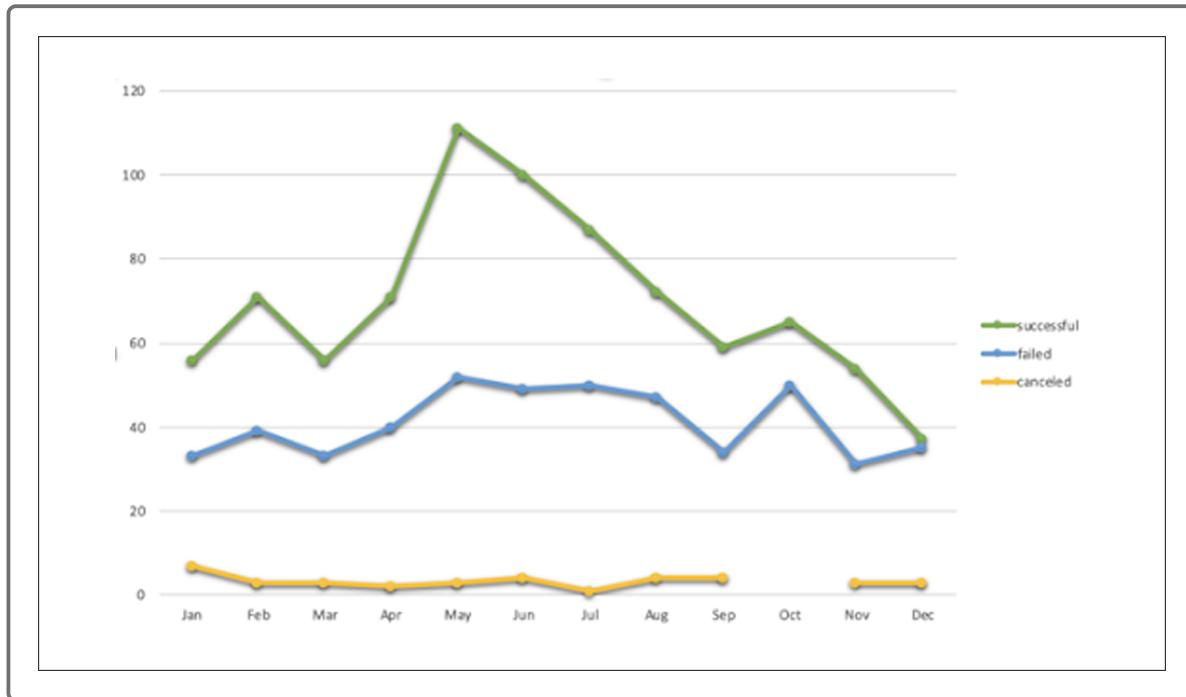
6. Create a line chart to visualize the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.

7. Save the line chart as a .png image, like the one below.



**In the "Outcomes Based on Launch Date" sheet, do the following:**

1. Create a pivot table that charts the outcome of each Kickstarter project based on the launch date.
2. Set the “Parent Category” filter to “theater” and save the chart as a .png image, like the one below.



## Rubric

View the following rubric for detailed evaluation criteria for this Challenge:

[Module 1 - Kickstarting with Excel Detailed Rubric .pdf](#)

## Submission

Once your final analysis is complete, make sure the two new analyses are in the **Outcomes Based on Goals** and **Outcomes Based on Launch Date** sheets as described above. Save your worksheet and reupload it to your GitHub repository.

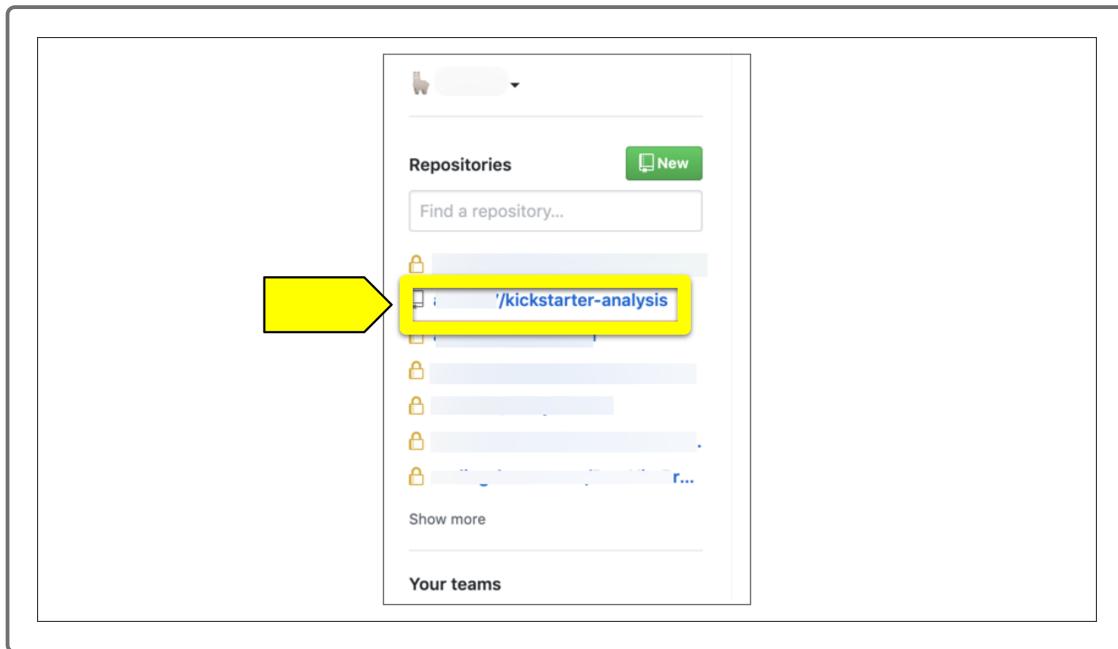
Next, update your README.md file as follows:

- At the bottom of your sheet, add a third-level header for the challenge i.e., **### Challenge**
- Under this section, add a short description of the new analysis and each .png file you saved earlier.

Save your README file. Make sure the most recent copy is in your GitHub repository.

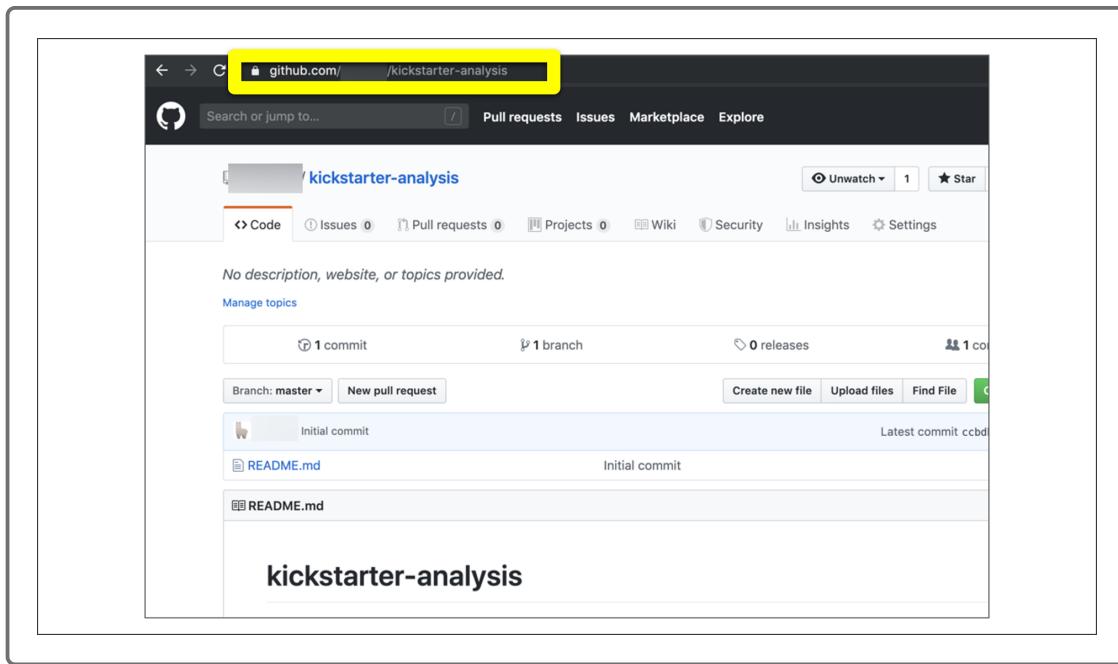
Your homework will be submitted by providing a link to your new repository. To get this link, follow these steps.

1. Navigate to [www.github.com](http://www.github.com) (<http://www.github.com>) in your browser and log in if you need to.
2. From the homepage, select your repository from the list on the left. Click the link to go to the homepage for that repository.



3. Select the address in your search bar and copy it by highlighting the entire address and then right-clicking to select Copy. You can also

copy by using the keyboard shortcut. This link can be shared with others so they can view the files in your repository or clone the repo if necessary.



#### 4. Submit the link to your repository through Canvas.

**Note:** You are allowed to miss up to two Challenge assignments and still earn your certificate. If you complete all Challenge assignments, your lowest two grades will be dropped. If you wish to skip this assignment, click Submit and then type "I choose to skip this assignment" in the text box.

#### Module 1 Challenge Rubric

| Criteria  | Ratings   |   |   |   |   | Pts                             |
|---|---|---|---|---|---|---------------------------------|
| <p>Row and Column Creation<br/>Please find detailed rubric in assignment description</p> <p>Please find detailed rubric in assignment description</p> | <p><b>10.0 pts</b><br/><b>Mastery</b><br/>8 new columns and 12 new rows were correctly created for categories listed in detailed rubric.</p> <p><b>15.0 pts</b><br/><b>Mastery</b><br/>COUNTIFS() function correctly applied to the three columns specified in detailed rubric.</p> | <p><b>7.5 pts</b><br/><b>Approaching Mastery</b><br/>6 or 7 new columns and 10 or 11 new rows were correctly created for categories listed in detailed rubric.</p> <p><b>7.5 pts</b><br/><b>Approaching Mastery</b><br/>COUNTIFS() function correctly applied to two of the columns specified in detailed rubric.</p> | <p><b>5.0 pts</b><br/><b>Progressing</b><br/>4 or 5 new columns and 8 or 9 new rows were correctly created for categories listed in detailed rubric.</p> <p><b>5.0 pts</b><br/><b>Progressing</b><br/>COUNTIFS() function correctly applied to one of the columns specified in detailed rubric.</p> | <p><b>2.5 pts</b><br/><b>Emerging</b><br/>Fewer than 4 new columns and fewer than 8 new rows were correctly created for categories listed in detailed rubric.</p> <p><b>3.75 pts</b><br/><b>Emerging</b><br/>COUNTIFS() function applied incorrectly to the three columns specified in the detailed rubric.</p> | <p><b>0.0 pts</b><br/><b>Incomplete</b><br/>No submission was received -OR- Submission was empty or blank -OR- Submission contains evidence of academic dishonesty</p> <p><b>0.0 pts</b><br/><b>Incomplete</b><br/>No submission was received -OR- Submission was empty or blank -OR- Submission contains evidence of academic dishonesty</p> | <p>10.0 pts</p> <p>15.0 pts</p> |
| <p>Simple formulas (Addition and division)</p>  | <p><b>15.0 pts</b><br/><b>Mastery</b><br/>Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for all twelve rows.</p>  | <p><b>11.25 pts</b><br/><b>Approaching Mastery</b><br/>Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for ten or eleven rows.</p>  | <p><b>7.5 pts</b><br/><b>Progressing</b><br/>Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for eight or nine rows.</p>  | <p><b>3.75 pts</b><br/><b>Emerging</b><br/>Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for fewer than eight rows.</p>   | <p><b>0.0 pts</b><br/><b>Incomplete</b><br/>No submission was received -OR- Submission was empty or blank -OR- Submission contains evidence of academic dishonesty</p>  | <p>15.0 pts</p>                 |

| Criteria                     | Ratings  |   |  |   |   | Pts      |
|------------------------------|--|---|--|---|---|----------|
| Line Chart 1                 | <b>20.0 pts</b><br><b>Mastery</b><br>Correctly creates a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.                          | <b>15.0 pts</b><br><b>Approaching Mastery</b><br>Creates, with one or two minor errors, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.                            | <b>10.0 pts</b><br><b>Progressing</b><br>Creates, with more than two minor errors or a significant error, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation. | <b>5.0 pts</b><br><b>Emerging</b><br>Creates, with more than three minor errors and/or multiple significant errors, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.                      | <b>0.0 pts</b><br><b>Incomplete</b><br>No submission was received - OR- Submission was empty or blank - OR- Submission contains evidence of academic dishonesty | 20.0 pts |
| Pivot Table and Line Chart 2 | <b>20.0 pts</b><br><b>Mastery</b><br>Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent   | <b>15.0 pts</b><br><b>Approaching Mastery</b><br>Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with one or two minor errors.                             | <b>10.0 pts</b><br><b>Progressing</b><br>Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with more than two minor errors or a significant error.  | <b>5.0 pts</b><br><b>Emerging</b><br>Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with more than three minor errors and/or multiple significant errors.                       | <b>0.0 pts</b><br><b>Incomplete</b><br>No submission was received - OR- Submission was empty or blank - OR- Submission contains evidence of academic dishonesty | 20.0 pts |
| Written Report               | <b>20.0 pts</b><br><b>Mastery</b><br>Presents a cohesive written analysis that draws three reasonable conclusions from the data and states limitations of the dataset and suggestions for additional tables of graphs. | <b>15.0 pts</b><br><b>Approaching Mastery</b><br>Presents a developing cohesive written analysis that draws at least one reasonable conclusion from the data and either states limitations of the dataset or suggestions for additional tables of graphs. | <b>10.0 pts</b><br><b>Progressing</b><br>Presents a developing written analysis that draws at least one reasonable conclusion from the data and either states limitations of the dataset or suggestions for additional tables of graphs.         | <b>5.0 pts</b><br><b>Emerging</b><br>Presents a limited written analysis or no written analysis that draws at least one or fewer reasonable conclusions from the data and does not state the limitations of the dataset or include suggestions for additional tables of graphs. | <b>0.0 pts</b><br><b>Incomplete</b><br>No submission was received - OR- Submission was empty or blank - OR- Submission contains evidence of academic dishonesty | 20.0 pts |
|                              | Total Points: 100.0  |   |  |   |   |          |

**Rubric for Kickstarting with Excel:**

|  | <b>Mastery</b>   | <b>Approaching Mastery</b>   | <b>Progressing</b>   | <b>Emerging</b>  | <b>Incomplete</b>  |
|--|--|--|--|--|--|
| <b>Row and Column Creation (10 points)</b>                 | <p>Eight new columns were correctly created for:<br/>Goal<br/>Number Successful<br/>Number Failed<br/>Number Canceled<br/>Total Projects<br/>Percentage Successful<br/>Percentage Failed<br/>Percentage Canceled<br/><b>And</b> twelve new rows were correctly created for:<br/>Less Than 1000<br/>1000 to 4999<br/>5000 to 9999<br/>10000 to 14999<br/>15000 to 19999<br/>20000 to 24000<br/>25000 to 29999<br/>30000 to 34999<br/>35000 to 39999<br/>40000 to 44999<br/>45000 to 49999</p> | <p>Six or Seven new columns were correctly created for:<br/>Goal<br/>Number Successful<br/>Number Failed<br/>Number Canceled<br/>Total Projects<br/>Percentage Successful<br/>Percentage Failed<br/>Percentage Canceled<br/><b>And</b> Ten or Eleven new rows were correctly created for:<br/>Less Than 1000<br/>1000 to 4999<br/>5000 to 9999<br/>10000 to 14999<br/>15000 to 19999<br/>20000 to 24000<br/>25000 to 29999<br/>30000 to 34999<br/>35000 to 39999<br/>40000 to 44999<br/>45000 to 49999</p> | <p>Four or Five new columns were correctly created for:<br/>Goal<br/>Number Successful<br/>Number Failed<br/>Number Canceled<br/>Total Projects<br/>Percentage Successful<br/>Percentage Failed<br/>Percentage Canceled<br/><b>And</b> Eight or Nine new rows were correctly created for:<br/>Less Than 1000<br/>1000 to 4999<br/>5000 to 9999<br/>10000 to 14999<br/>15000 to 19999<br/>20000 to 24000<br/>25000 to 29999<br/>30000 to 34999<br/>35000 to 39999<br/>40000 to 44999<br/>45000 to 49999</p> | <p>Fewer than four new columns were correctly created for:<br/>Goal<br/>Number Successful<br/>Number Failed<br/>Number Canceled<br/>Total Projects<br/>Percentage Successful<br/>Percentage Failed<br/>Percentage Canceled<br/><b>And</b> fewer than eight new rows were correctly created for:<br/>Less Than 1000<br/>1000 to 4999<br/>5000 to 9999<br/>10000 to 14999<br/>15000 to 19999<br/>20000 to 24000<br/>25000 to 29999<br/>30000 to 34999<br/>35000 to 39999<br/>40000 to 44999<br/>45000 to 49999</p> | No submission was received<br>-OR-<br>Submission was empty or blank<br>-OR-<br>Submission contains evidence of academic dishonesty |
| <b>COUNTIFS() function (10 points)</b>                     | COUNTIFS() function correctly applied to the three columns listed below.<br>Number Successful<br>Number Failed<br>Number Canceled  | COUNTIFS() function correctly applied to two of the columns listed below.<br>Number Successful<br>Number Failed<br>Number Canceled   | COUNTIFS() function correctly applied to one of the columns listed below.<br>Number Successful<br>Number Failed<br>Number Canceled   | COUNTIFS() function applied incorrectly to the three columns listed below.<br>Number Successful<br>Number Failed<br>Number Canceled  |  |
| <b>Simple formulas (Addition and division) (10 points)</b> | Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for all twelve rows.   | Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for all ten or eleven rows.  | Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for all eight or nine rows.  | Addition function correctly applied to the Number Successful, Number Failed, and Number Canceled columns to calculate Total Projects column for fewer than eight rows.   |  |
| <b>Line Chart 1 (15 points)</b>                            | Correctly creates a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.   | Creates, with one or two minor errors, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.  | Creates, with more than two minor errors or a significant error, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.  | Creates, with more than three minor errors and/or multiple significant errors, a line graph that visualizes the relationship between the goal amount and the campaign's chances of success, failure, or cancellation.  |  |

|  |   |  |  |  |  |
|--|---|--|--|--|--|
| <b>Pivot Table and Line Chart 2</b><br>(25 points) | Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category.   | Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with one or two minor errors.   | Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with more than two minor errors or a significant error.   | Create a pivot table that correctly charts the outcome of each kickstarter project based on the launch date and is filtered by parent category with more than three minor errors and/or multiple significant errors.   |  |
| <b>Written Report</b><br>(30 points)               | <p>Presents a cohesive written analysis that:</p> <ul style="list-style-type: none"> <li>✓ Draws three reasonable conclusions from the data</li> <li>✓ States limitations of the dataset and suggestions for additional tables of graphs</li> </ul> | <p>Presents a cohesive written analysis that:</p> <ul style="list-style-type: none"> <li>✓ Draws at least two reasonable conclusions from the data</li> <li>✓ States either limitations of the dataset or suggestions for additional tables of graphs</li> </ul> | <p>Presents a developing written analysis that:</p> <ul style="list-style-type: none"> <li>✓ Draws at least one conclusion from the data</li> <li>✓ States either limitations of the dataset or suggestions for additional tables of graphs</li> </ul> | <p>Presents a limited written analysis or no written analysis that:</p> <ul style="list-style-type: none"> <li>✓ Draws one or fewer conclusions from the data</li> <li>✓ Does not include limitations of the dataset or suggestions for additional tables of graphs</li> </ul> |  |

# Module 1

## Career Connection

Welcome to your first Career Connection! This lesson at the end of each module will review real-world applications for the material covered in the past week. We'll prepare you to be tested on your new knowledge in technical interviews and suggest strategies to become employer-competitive.



CAREER  
SERVICES

We've consulted many employers to find out what will help you stand out from the crowd. Armed with that knowledge, we'll guide you through steps you can take every week to get there.

### Becoming Employer-Competitive

Here at Career Services, we've come up with two job-preparedness levels: employer-ready and employer-competitive. Let's look at what these terms mean:

- **Employer-ready** means that you meet the minimum requirements to apply for a developer job. You have strong, complete job search materials (e.g., resume, portfolio, LinkedIn profile). With some luck, an employer might notice you among a stack of applications and decide you fit their needs.
- **Employer-competitive** means that you've given yourself the best chance to secure your desired job. You've put together excellent job search materials, you know how to prepare for interviews, and you use

multiple networking strategies to proactively seek and pursue the jobs you want.

## DEEP DIVE ▼

Most of the students who work with Career Services to reach the employer-competitive level find employment within six months after graduation!

Being employer-competitive means starting now to position yourself as a developer, think like a developer, and engage with the online development community. Ready to get started? Complete the following three steps, and you'll be that much closer to acing the job interview when the time comes.

### **Step #1: Just see what's out there.**

Peruse open source projects to get a feel for what's out there. Pay attention to what other developers write in their commit messages. You can start by reviewing the following projects on GitHub:

- [Project Open Data](https://github.com/project-open-data/) (<https://github.com/project-open-data/>)
- [Sherlock](https://github.com/sherlock-project/sherlock) (<https://github.com/sherlock-project/sherlock>)
- [United States Congress/Legislators](https://github.com/unitedstates/congress-legislators)  
(<https://github.com/unitedstates/congress-legislators>)

### **Step #2: Learn how to market yourself.**

Browse professionally designed websites and apps for inspiration as you prepare to build a portfolio. The following links provide numerous examples of style and layout:

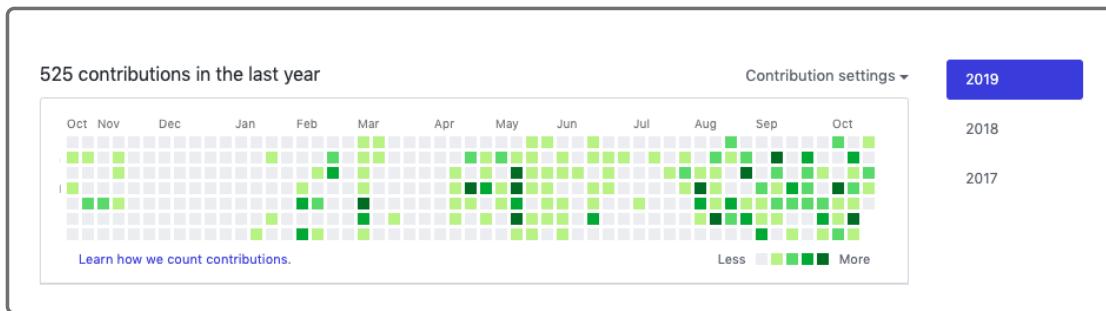
- [10 Fantastic Portfolio Websites from Silicon Valley Design Leads](https://medium.com/@bestfolios/10-fantastic-portfolio-websites-from-silicon-valley-design-leads-2d84b384dba6)  
(<https://medium.com/@bestfolios/10-fantastic-portfolio-websites-from-silicon-valley-design-leads-2d84b384dba6>)

- [One Page Love](https://onepagelove.com/inspiration/portfolio) (<https://onepagelove.com/inspiration/portfolio>)

### Step #3: Code a lot—and put it on GitHub.

Many companies will ask you to “pair program” with one of their employees during an interview, so practice this skill with a friend or classmate. Also, because GitHub makes it easy for employers to check your commit history, make sure both you and your partner commit to GitHub regularly. Don’t forget to keep your commit messages clean and professional!

Review your own code commit chart periodically to see how you appear to employers. The following image shows an example of a commit chart:



## Look to the Future

This week, you learned Excel, which is fundamental to learning any data analysis and visualization technologies you'll develop in the future. You're already on your way to becoming a coder!

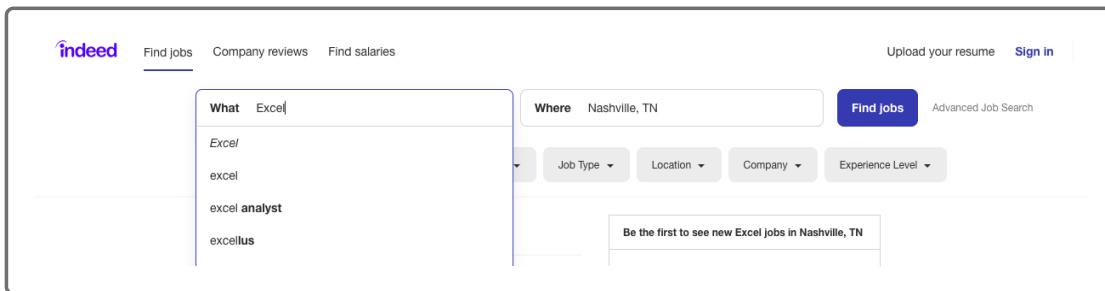
Senior analysts and/or engineers work on complex projects with a variety of technologies and tools. So when you're learning to code, use plenty of tools and methodologies from current real-world projects.

But it all begins with Excel. Though it's unlikely that you'll ever interview directly for an “Excel job,” you'll almost definitely face questions about Excel and spreadsheet data at some point in the interview process. Start practicing these skills now.

## Step #1: Investigate job listings.

[Indeed.com](http://Indeed.com) (<http://Indeed.com>) is a great resource for finding a variety of job postings. Visit the website and search for Excel.

The following screenshot shows what that search may look like:



Look through the search results for each category. Then write your answers to the following questions:

- **Where does Excel typically appear in the job description?**

You may find it listed with other technologies under a heading like “Skills Required,” “Tech Stack,” or something similar. These lists can be intimidating, but think of them as nothing more than wish lists—most companies know they won’t get everything they want. Don’t be afraid to apply for jobs where you only know 60% or 70% of the required technologies. Employers are also looking for that all-important cultural fit, meaning they want someone who will work well with the team. Technical skills can be taught but cultural fit can’t.

- **What other skills are listed alongside HTML, CSS, and Git?**

Besides Excel, you might see Python, SQL, NoSQL, and Tableau, as well as a variety of other technologies. By the end of this course, you’ll possess these skills too. You might see other technologies listed that aren’t covered in this course. Yet the skills you’ll learn transfer directly to these other technologies. They aren’t all that different, anyway!

- What level of experience, in years, are employers typically looking for?

You'll notice that almost all job postings specify at least two to three years of experience. Those looking for senior analysts will call for five to eight years minimum. Again, treat these requests like a wish list. If you see a job posting along the lines of "data analyst, two years experience," you can and should apply for that job.

### Step #2: Add technical skills to your resume.

Before you move on to the next module, make sure to add Excel to the "Technical Skills" section on your resume. This will help you pass those pesky automated resume scanners and will indicate to potential employers that you have the skills they need.

## DEEP DIVE ▼

---

## Technical Interview

Many companies use algorithm questions during the technical interview. We'll start diving into those when we learn JavaScript. For Excel, consider the following common technical interview questions:

- What is the difference among `COUNT`, `COUNTA`, `COUNTIF` and `COUNTBLANK` in Excel?
  - `COUNT` is used to count cells containing numbers, dates, etc.
  - `COUNTA` is used to count any cell value containing numbers, text, logical values, etc. any type of value excluding blanks.
  - `COUNTBLANK` will count blank cells or cells containing an empty string.

- `COUNTIF` and `COUNTIFS` count cells matching a certain criteria in the IF statement.
- Is it possible to make a pivot table using multiple sources of data?
  - Yes, but the multiple sources must at least be within the same workbook
- What filter should we use if you want more than two conditions?
  - The "Advanced Criteria Filter" will work for us to analyze the list or test more than two conditions.

When you've answered these questions on your own, search the internet to check your answers.

---

## Continue to Hone Your Skills



If you're interested in learning more about the technical interviewing process and practicing algorithms in a mock interview setting, check out our [upcoming workshops](#).  
[\(https://careerservicesonlineevents.splashthat.com/\)](https://careerservicesonlineevents.splashthat.com/)