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.

Score to be Submitted   Based on last attempt

Score
Time spent: 1m 18s   Review

⟳ Retake

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.

Score to be Submitted   Based on last attempt

Score
Time spent: 1m 7s   Review

⟳ Retake

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, first click on the Home tab, then select cell 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.
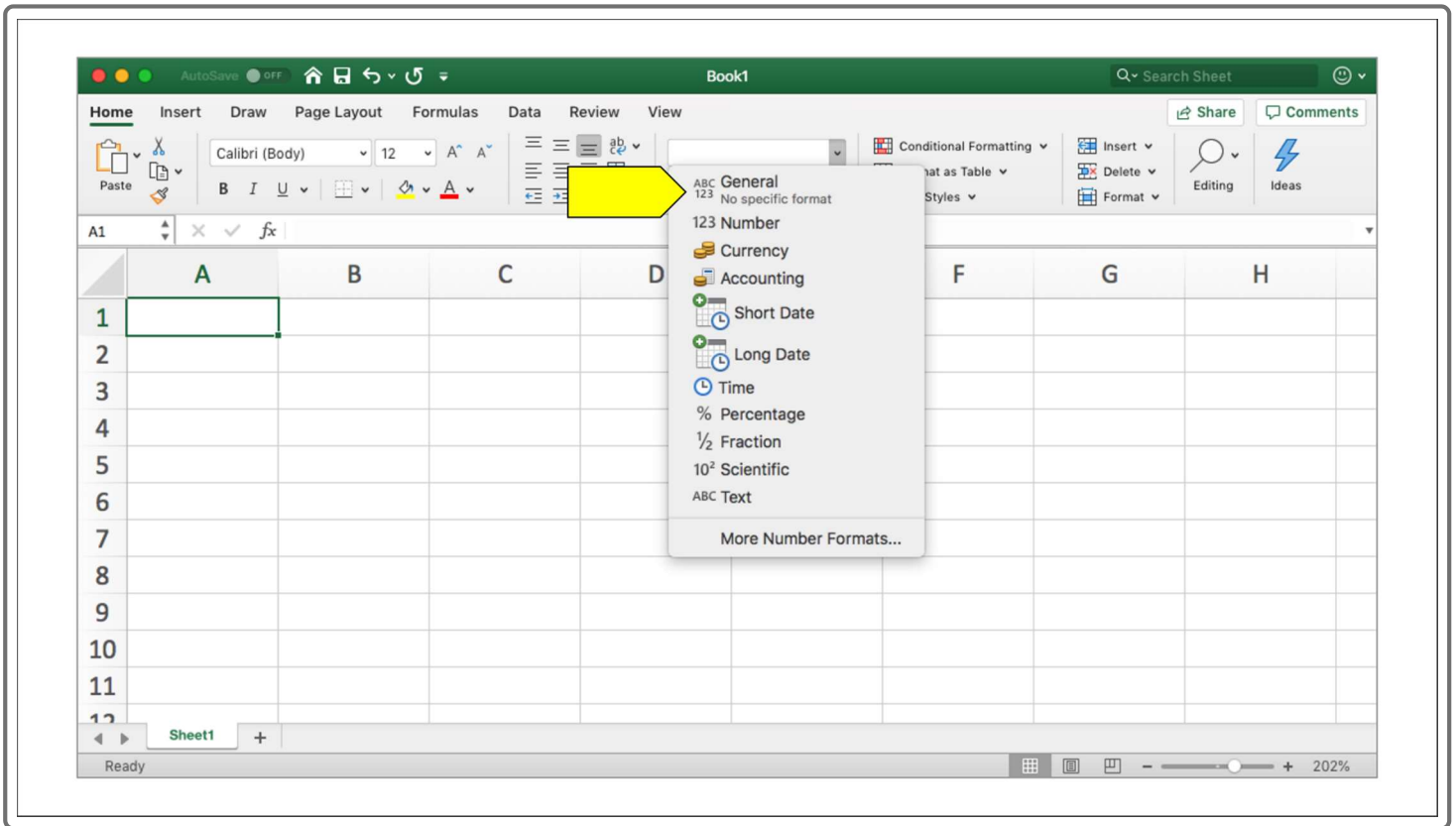
Now let's check the next column over, **D2**. The value has changed to Accounting because this cell contains monetary data.
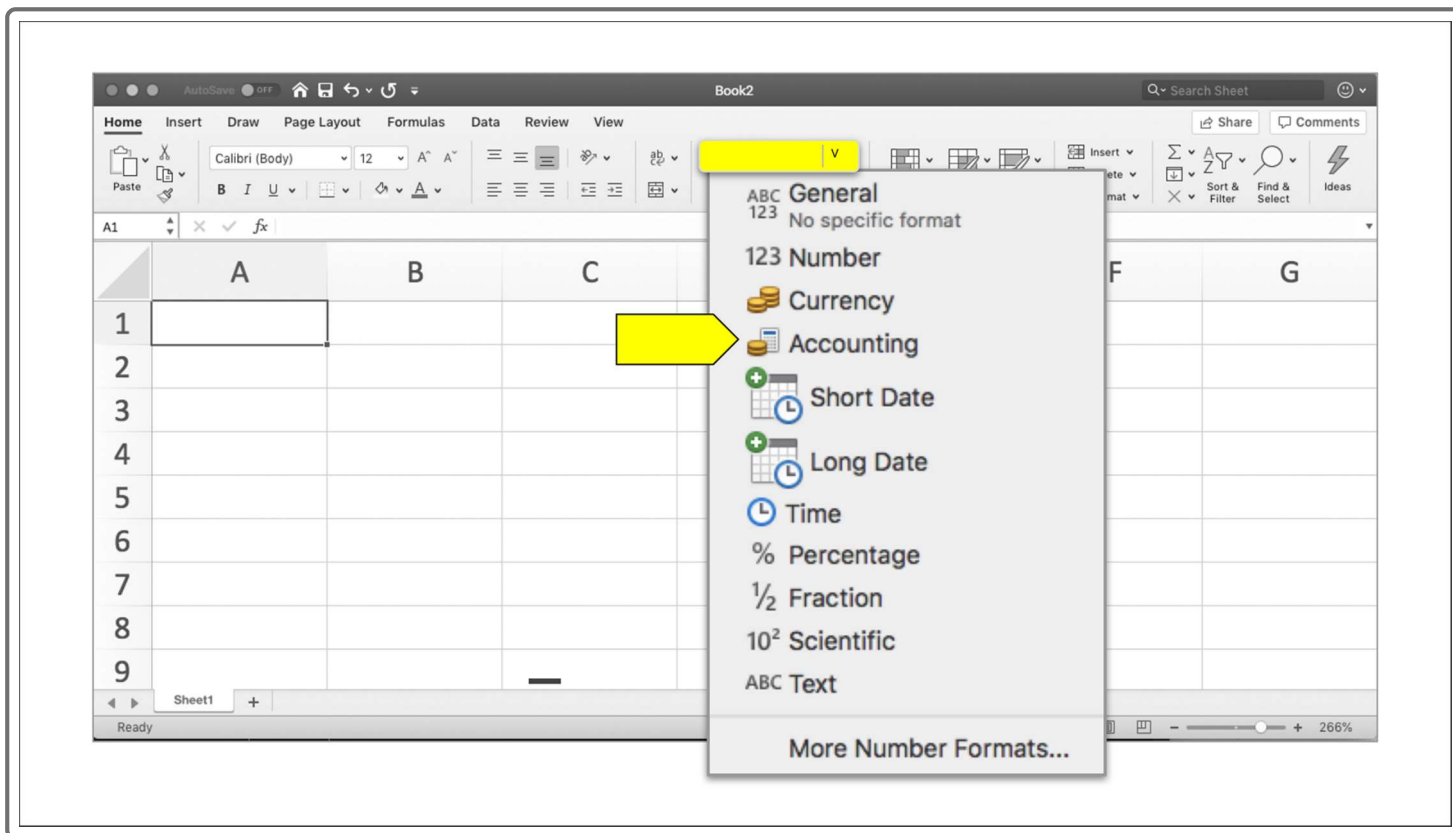
Be aware that sometimes the data isn't classified correctly, so it's a good idea to double-check them.

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

1:05

Windows

1:05

Let's format another column by changing a text column to a numeric column. Select all of column B 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 B back 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/)** tool.

> **NOTE**
>
> Learn more about the use of **Unix timestamps** ⤷ **(https://websiteseochecker.com/blog/what-is-timestamp/)** .

Select the text in one of the cells and copy it. Then, paste it into the online converter tool to see how it's interpreted. After clicking the "Timestamp to Human date" button, the tool will provide the date and time assigned to that particular string. This 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.