## **Hands-On Activity: Cleaning with spreadsheets**

**TOTAL POINTS 2**

1.

Question 1



## Activity overview



You’ve learned about cleaning data and its importance in meeting good data science standards. In this activity, you’ll do some data cleaning with spreadsheets, then transpose the data.

By the time you complete this activity, you will be able to perform some basic cleaning methods in spreadsheets. This will enable you to clean and transpose data, which is important for making data more specific and accurate in your career as a data analyst.



### What you will need

To get started, first access the data spreadsheet.

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

Link to data spreadsheet: [Cleaning with spreadsheets](https://docs.google.com/spreadsheets/d/1PkAbgXC7C1g2dKzCCpaHBcAyPw-s1z7iUxIEJ0cCYWQ/template/preview)

OR

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

[Data Spreadsheet for Cleaning with Spreadsheets.xlsx](https://d3c33hcgiwev3.cloudfront.net/9Ss-5kXlRP6rPuZF5VT-qA_d5c83c4e13d643cd87cb22632fcaa9c6_Data-Spreadsheet-for-Cleaning-with-Spreadsheets.xlsx?Expires=1623024000&Signature=aqYu6tS40SGlaurliJ6cKH2EoSmMJvXJvuWMbKHHKm71Omb-Mrtzg2jRlbUNdIFkDVb4t~ulrMOo1ogBYXklZEgnt3d7LbKxWSzCge83gekhIervQ3tPLMxI5h3bjADpiOkus1~Rlvc1g8fcGqSRGC6LYuzOKsZxJxLnrd62nIE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

[](https://d3c33hcgiwev3.cloudfront.net/9Ss-5kXlRP6rPuZF5VT-qA_d5c83c4e13d643cd87cb22632fcaa9c6_Data-Spreadsheet-for-Cleaning-with-Spreadsheets.xlsx?Expires=1623024000&Signature=aqYu6tS40SGlaurliJ6cKH2EoSmMJvXJvuWMbKHHKm71Omb-Mrtzg2jRlbUNdIFkDVb4t~ulrMOo1ogBYXklZEgnt3d7LbKxWSzCge83gekhIervQ3tPLMxI5h3bjADpiOkus1~Rlvc1g8fcGqSRGC6LYuzOKsZxJxLnrd62nIE_&Key-Pair-Id=APKAJLTNE6QMUY6HBC5A)

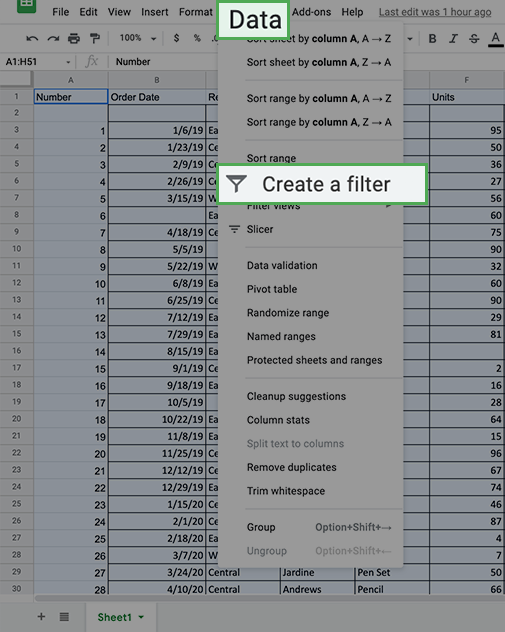
## Select and remove blank cells



The first technique we’ll use is to select and eliminate rows containing blank cells by using filters. To eliminate rows with blank cells:

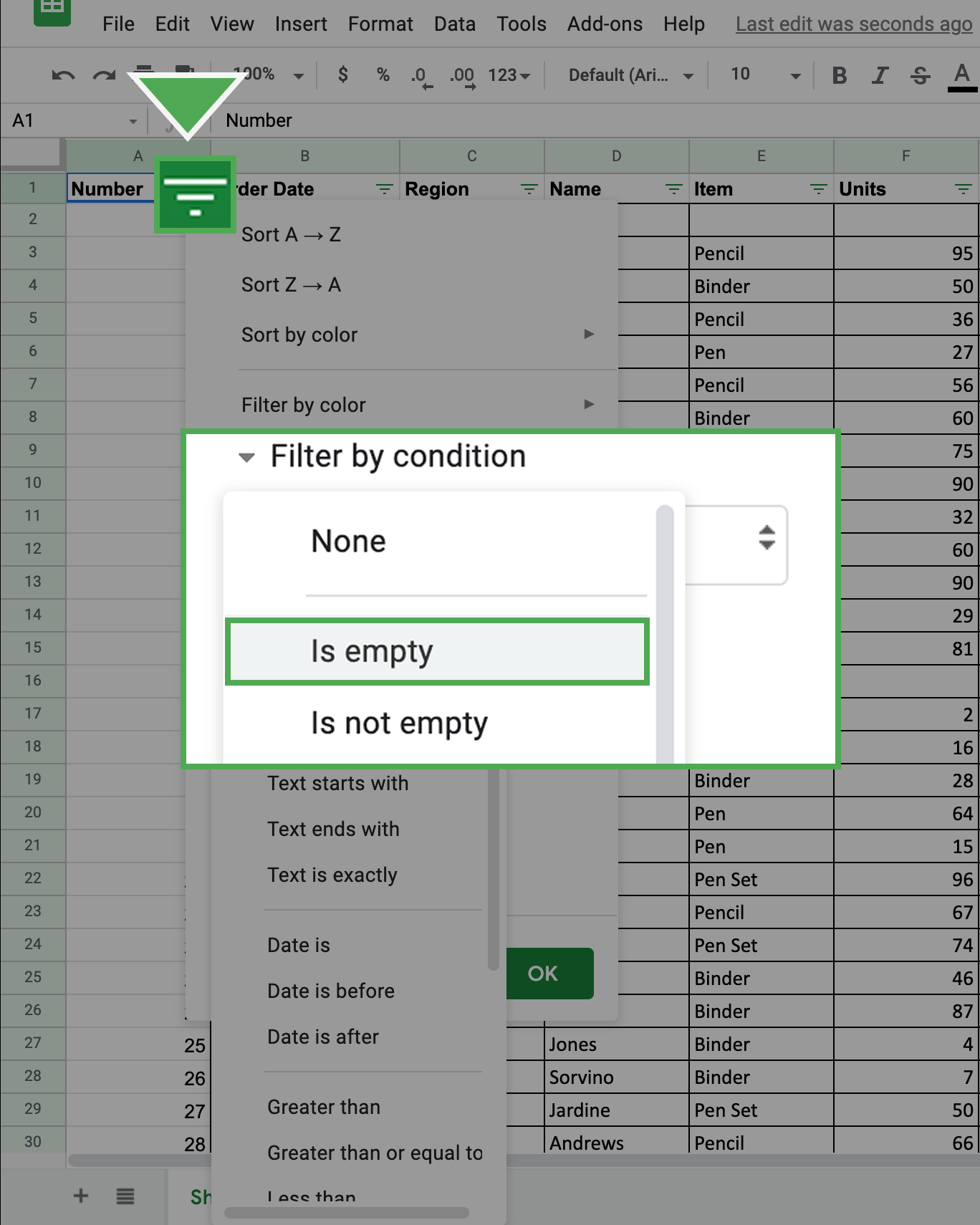
1. Highlight all cells in the spreadsheet. You can highlight Columns A-H by clicking on the first cell in Column A, holding Shift, and clicking on Column H.

2. Click on the Data tab and pick the Create a filter option.



3. Every column now shows a green triangle in the first row next to the column title. Click the green triangle in Column 1 to access a new menu.

4. On that new menu, click on Filter by condition and open the dropdown menu to select Is empty. Click OK.



You can then review a list of all the rows with blank cells in that column.

5. Select all these cells and delete the rows except the row of column headers.

6. Return to the Filter by condition and return it to None.

* Note: You will now notice that any row that had an empty cell in Column A will be removed (including the extra empty rows after the data).

7. Repeat this for Columns B-H.

All the rows that had blank cells are now removed from the spreadsheet.

## Transpose the data

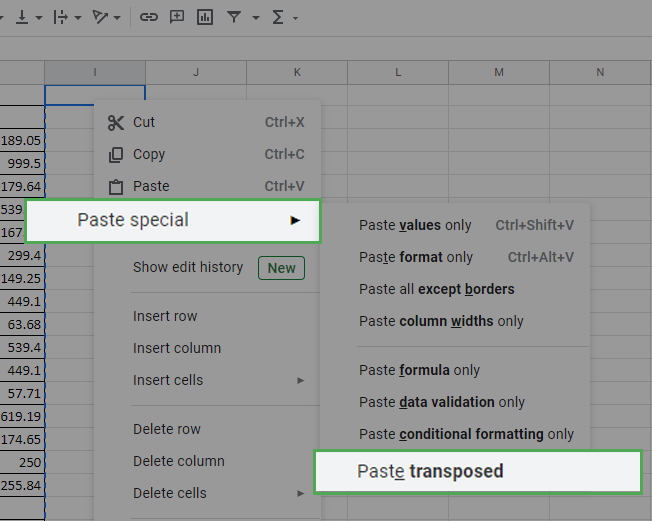


The second technique you will practice will help you convert the data from the current long format (more rows than columns) to the wide format (more columns than rows). This action is called transposing. To transpose your data:

1. Highlight and copy the data that you want to transpose including the column labels. You can do this by highlighting Columns A-H, right-clicking the highlighted rows, and selecting Copy.

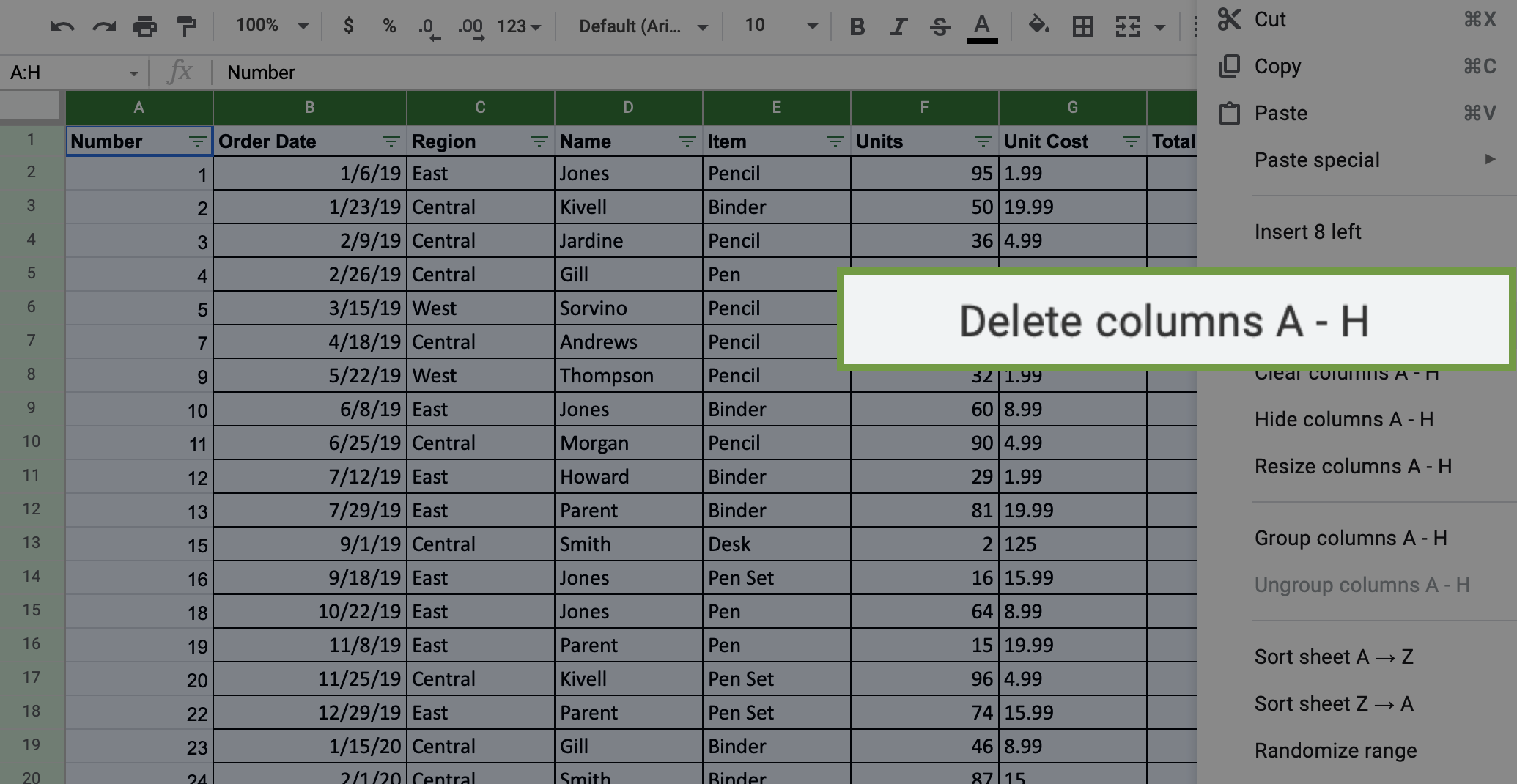
2. Right-click on cell I1. This is where you want the transposed data to start.

3. Hover over Paste Special from the right-click menu. Select the Paste transposed option.

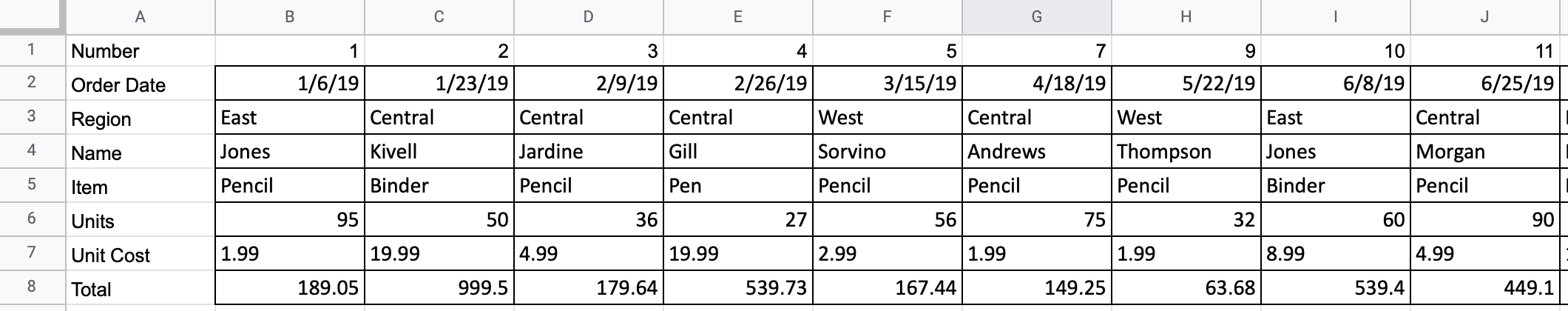


You should now find the data transformed into the new wide format. At this point, you should remove the original long data from the spreadsheet.

4. Delete the previous long data. The easiest way to do this is to click on Column A, so the entire column is highlighted. Then, hold down the Shift key and click on Column H. You should find these columns highlighted. Right-click on the highlighted area and select Delete Columns A - H.



You should now have something that appears like this:



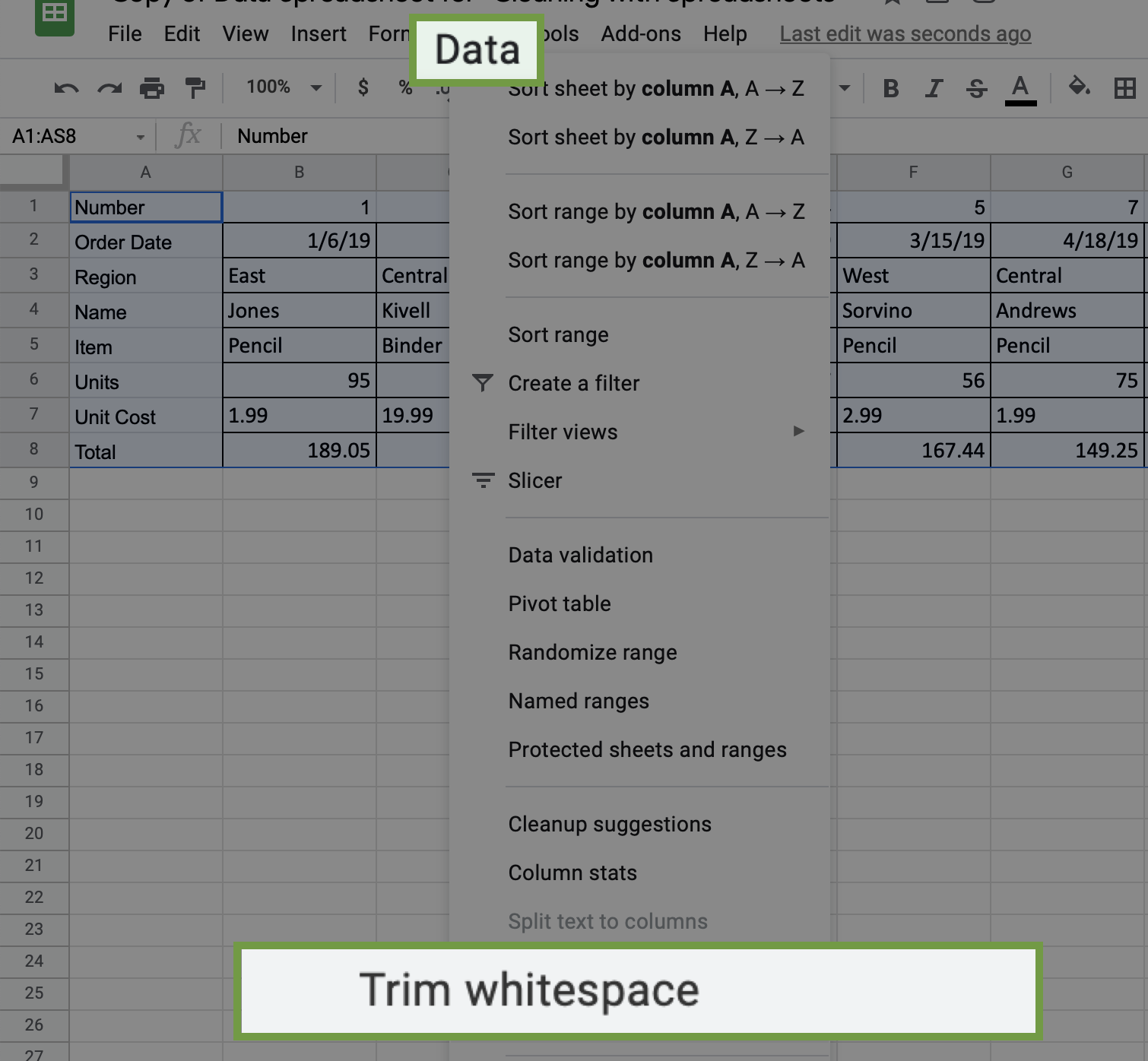
## Get rid of extra spaces in cells with string data



Now that you have transposed the data, eliminate the extra spaces in the values of the cells. To do that, use the TRIM command.

1. Highlight the data in the spreadsheet.

2. Click on the Data tab and select the Trim whitespace option.



Now all the extra spaces in the cells have been removed.

## Change Text Lower/Uppercase/Proper Case



Next, you’ll process string data. The easiest way to clean up string data will depend on the spreadsheet program you are using. If you are using Excel, you’ll use a simple formula. If you are using Google Sheets, you can use an Add-On to do this with a few clicks. Follow the steps in the relevant section below.

### Microsoft Excel



If you are using Microsoft Excel, [this documentation](https://support.microsoft.com/en-us/office/change-the-case-of-text-in-excel-adc65f5b-958f-46a2-4d23-ab4d5faf48a8) explains how to use a formula to change the case of a text string. Follow these instructions to clean the string text and then move on to the confirmation and reflection section of this activity.

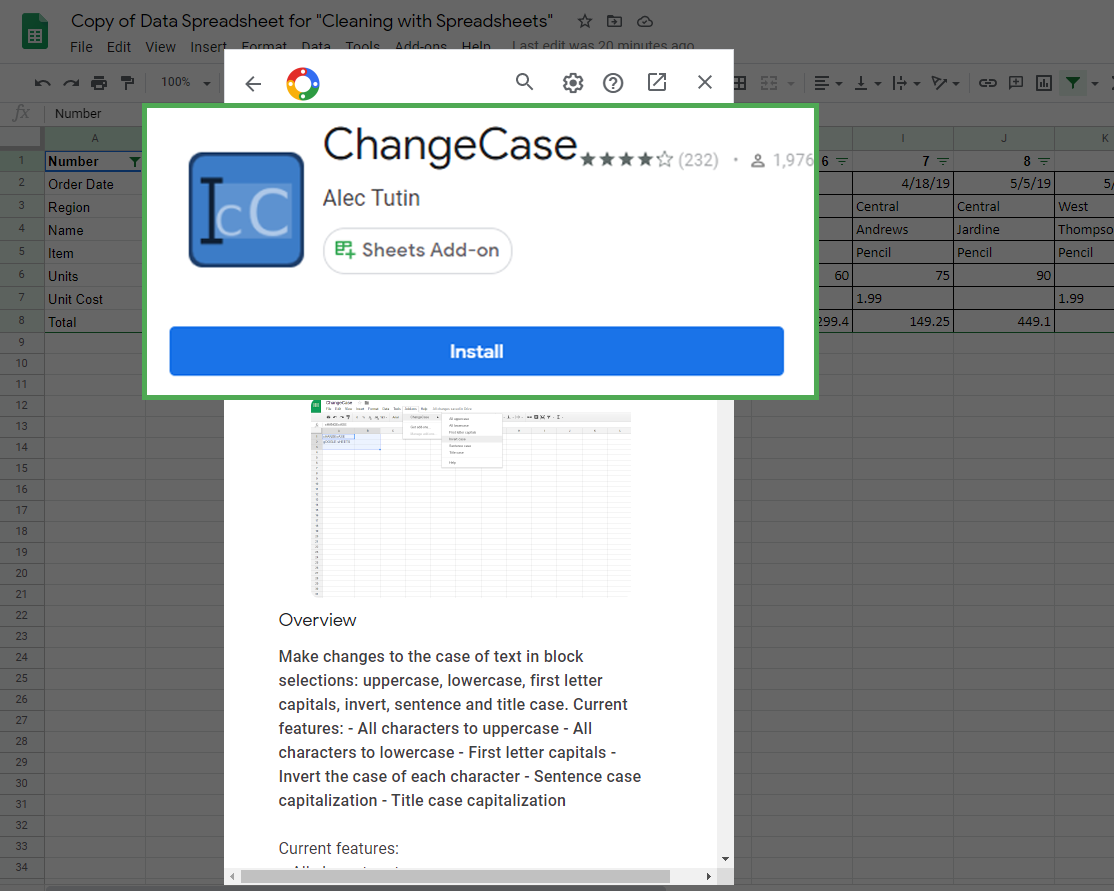
### Google sheets



If you’re completing this exercise using Google Sheets, you’ll need to install an add-in that will give you the functionality needed to easily clean string data and change cases.

*Google Sheets Add-on Instructions:*

1. Click on the Add-Ons option at the top of Google Sheets.
2. Click on Get add-ons.
3. Search for ChangeCase. It should appear like this:



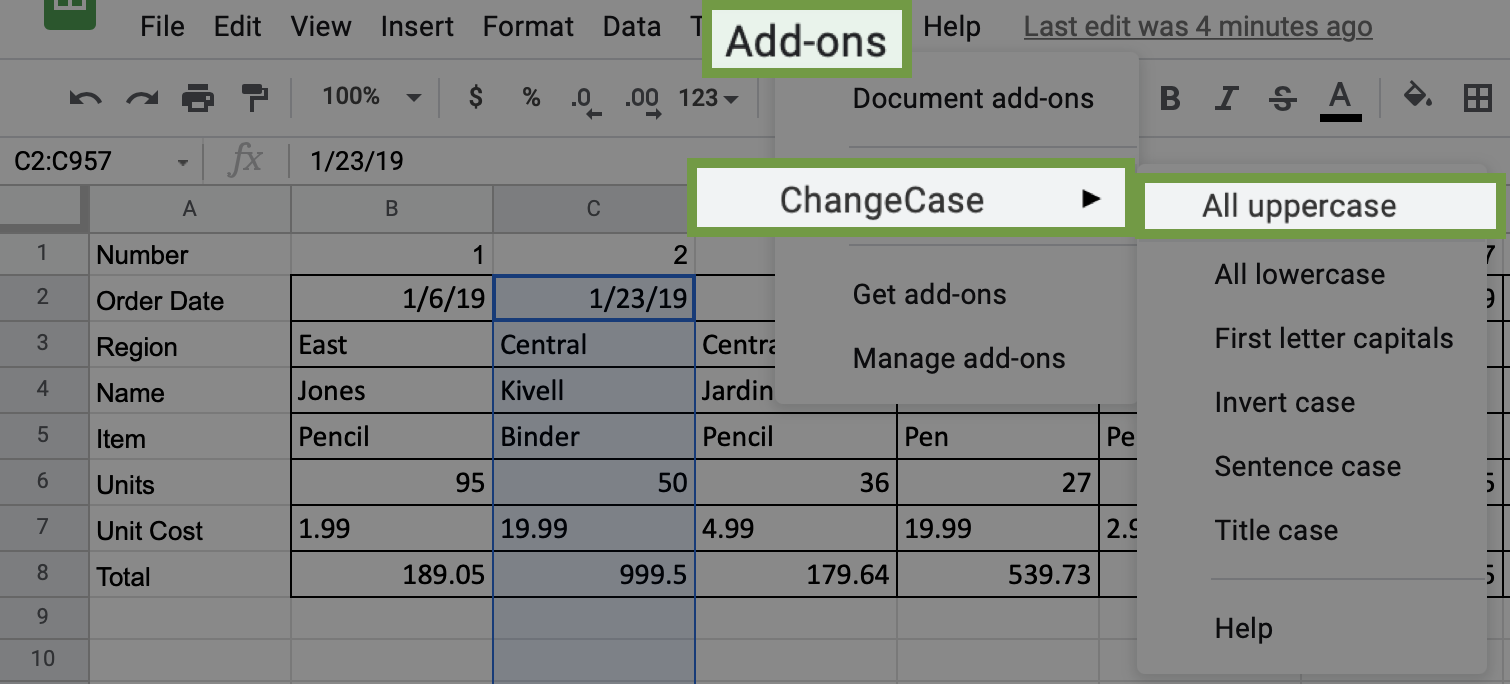
4. Click on Install to install the add-on. It may ask you to login or verify the installation permissions.

Once you have installed the add-on successfully, you can access it by clicking on the Add-ons menu again.

Now, you can change the case of text data that shows up. To change the text in Column C to all uppercase:

1. Click on Column C. Be sure to deselect the column header, unless you want to change the case of that as well (which you don't).

2. Click on the Add-Ons tab and select ChangeCase. Select the option All uppercase. Notice the other options that you could have chosen if needed.



## Delete all formatting



If you want to clear the formatting for any or all cells, you can find the command in the Format tab. To clear formatting:

1. Select the data for which you want to delete the formatting. In this case, highlight all the data in the spreadsheet by clicking and dragging over Rows 1-8.

2. Click the Format tab and select the Clear Formatting option.

You will notice that all the cells have had their formatting removed.

## Confirmation and reflection



Review the final product of the spreadsheet you cleaned during this activity. Row 1 of this spreadsheet is the Number row. In which column of Row 1 is the value 49?

1 point

Column AA

Column AS

Column Z

Column AZ

2.

Question 2

In this activity, you practiced cleaning and transposing data. In the text box below, write 2-3 sentences (40-60 words) in response to each of the following questions:

* What was the most challenging part of cleaning the data?
* Why is cleaning and transposing data important for data analysis?
* If you had to clean this data again, would you do differently? Why?

that cleaning is a fundamental step in data science as it greatly increases the integrity of the data.

Good data science results rely heavily on the reliability of the data. Data analysts clean data to make it more accurate and reliable. This is important for making sure that the projects you will work on as a data analyst are completed properly.