Hands-On Activity: Cleaning data with spreadsheets

TOTAL POINTS 2

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

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

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 header of **Column A**, holding **Shift**, and clicking on the header of **Column H**.

2. Click on the **Data** tab and pick the **Create a filter** option. In Microsoft Excel, this is called **Filter**.
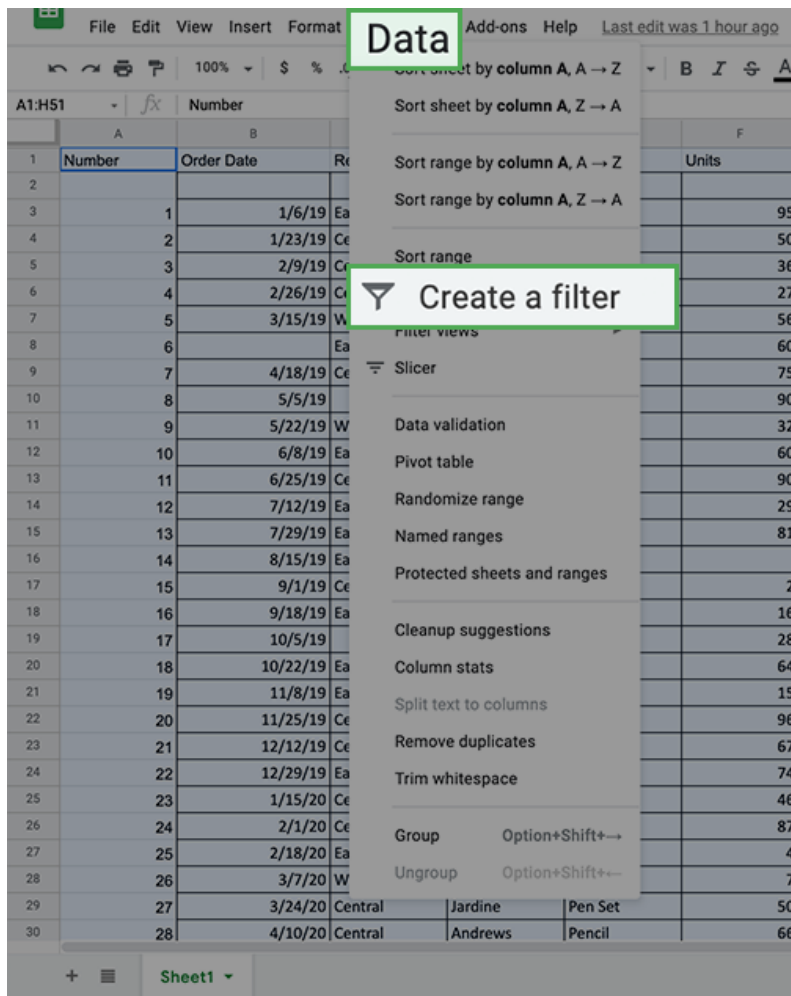
Excel:
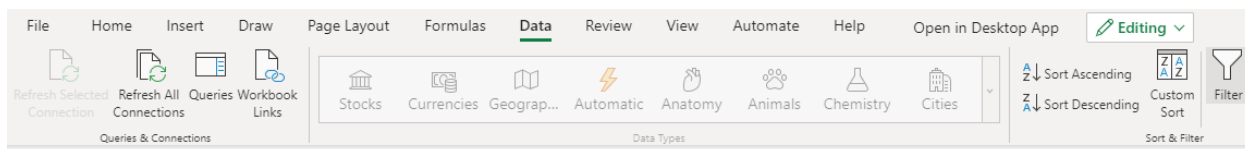
3. Every column now shows a green triangle in the first row next to the column title. Click the green triangle in **Column B** to access a new menu.

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

In Excel, click the dropdown, then **Filter...** then make sure only **(Blanks)** is checked. Click **OK**.

Excel:

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**. In Excel, click **Clear Filter from 'Column'**.

- **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**. In Excel, highlight only the relevant cells (**A1-H45**) instead of the headers.

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 **Transposed** option. In Excel, select the **Transpose** icon under the paste options.

	H	I	J	K	L	M	N
Total							
	189.05						
	999.5						
	179.64						
	539.73						
	167.44						
	149.25						
	63.68						
	539.4						
	449.1						
	57.71						
	1,619.19						
	250						
	255.84						
	575.36						
	299.85						
	479.04						
	1,183.26						
	413.54						
	1,305.00						
	19.96						
	139.93						
	249.5						
	131.34						
	479.04						
	68.37						
	719.2						
	625						
	309.38						
	686.95						

Context menu:

Cut	Ctrl+X
Copy	Ctrl+C
Paste	Ctrl+V
Paste special ►	
Insert 1 row above	
Insert 1 column left	
Insert cells ►	
Delete row	
Delete column	
Delete cells ►	
Show edit history	
Insert link	
Comment	Ctrl+Alt+M
Insert note	
Convert to people chip	
View more cell actions ►	

Paste special submenu:

Values only	Ctrl+Shift+V
Format only	Ctrl+Alt+V
Formula only	
Conditional formatting only	
Data validation only	
As text	
CSV as columns	
Transposed	
Column width only	
All except borders	

Excel:

Search

Cut

Copy

Paste Options

Insert ＞

Delete ＞

Clear Contents

Sort ＞

Transpose

New Comment

Number Format...

Hyperlink...

Show Changes

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**.

Your screen should now appear like this:

	A	B	C	D	E	F	G	H	I	J
1	Number	1	2	3	4	5	7	9	10	11
2	Order Date	1/6/19	1/23/19	2/9/19	2/26/19	3/15/19	4/18/19	5/22/19	6/8/19	6/25/19
3	Region	East	Central	Central	Central	West	Central	West	East	Central	
4	Name	Jones	Kivell	Jardine	Gill	Sorvino	Andrews	Thompson	Jones	Morgan	
5	Item	Pencil	Binder	Pencil	Pen	Pencil	Pencil	Pencil	Binder	Pencil	
6	Units	95	50	36	27	56	75	32	60	90
7	Unit Cost	1.99	19.99	4.99	19.99	2.99	1.99	1.99	8.99	4.99	
8	Total	189.05	999.5	179.64	539.73	167.44	149.25	63.68	539.4	449.1	

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.

1. Highlight the data in the spreadsheet.

2. Click on the **Data** tab, then hover over **Data cleanup** and select **Trim whitespace**.

Sort sheet ▸

Sort range ▸

Create a filter

Filter views ▸

Add a slicer

Protect sheets and ranges

Named ranges

Randomize range

Column stats

Data validation

Data cleanup ▸

Split text to columns

Data connectors ▸

	F	G	H
	5	7	9
	3/15/19	4/18/19	5/22/19
		Central	West
no		Andrews	Thompson
il		Pencil	Pencil
	56	75	32
		1.99	1.99
	167.44	149.25	63.68

Cleanup suggestions

Remove duplicates

Trim whitespace

In Excel, you can use the TRIM command to get rid of white spaces. In any space beneath your data (such as cell **A10**), type =TRIM(A1). Then, drag the bottom right corner of the cell to the bottom right to call the data without the white spaces.

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 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.

1. Click on the **Add-Ons** option at the top of Google Sheets.

2. Click on **Get add-ons**.

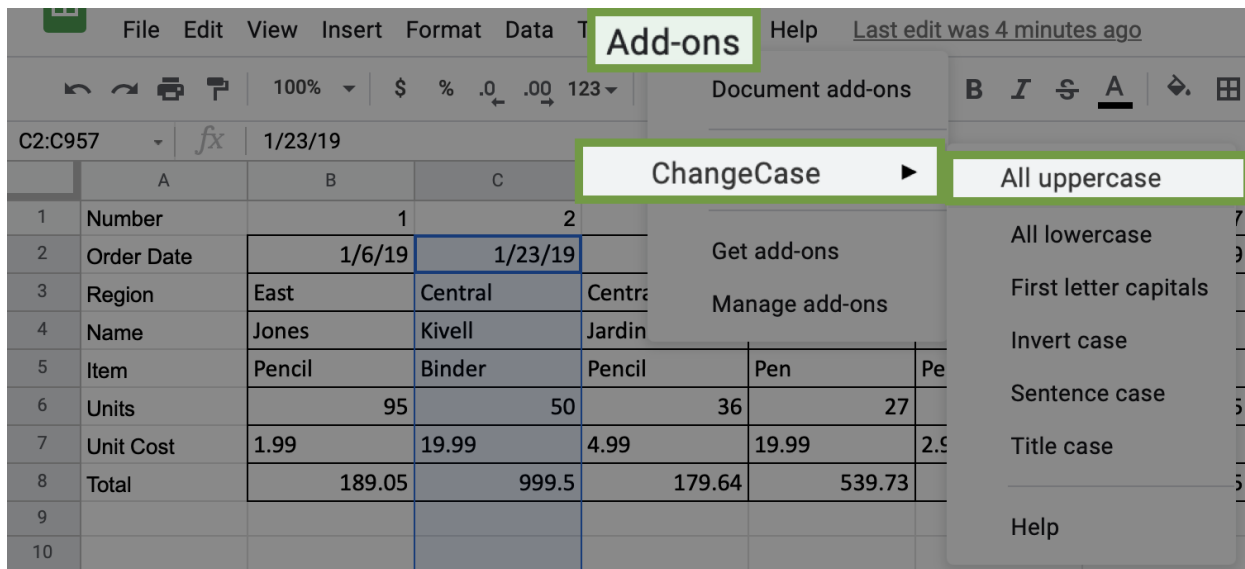3. Search for **ChangeCase**. It should appear like this:
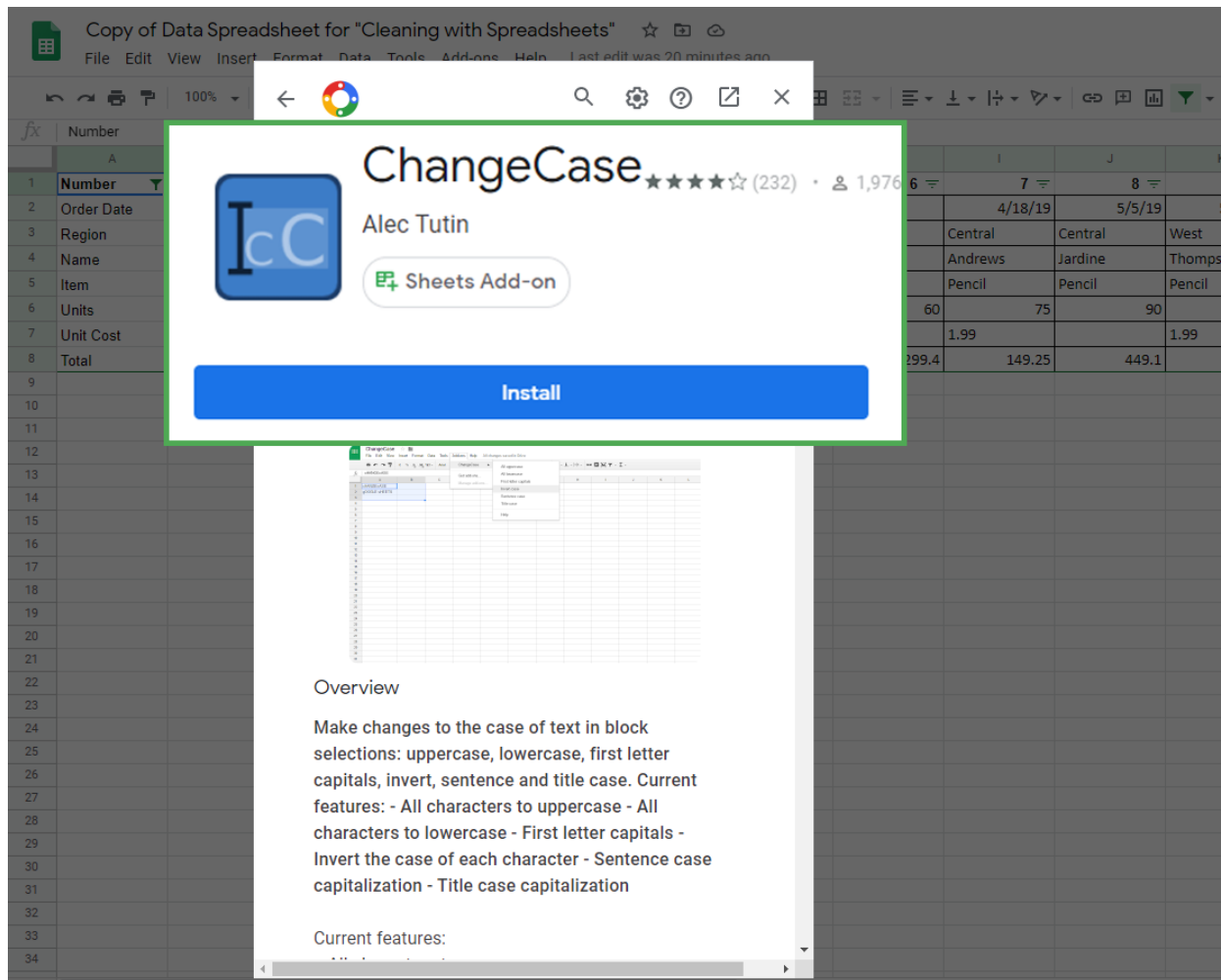
4. Click on **Install** to install the add-on. It may ask you to login or verify the installation permissions.

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.

	A	B	C			
1	Number	1	2			7
2	Order Date	1/6/19	1/23/19			9
3	Region	East	Central	Centra		
4	Name	Jones	Kivell	Jardin		
5	Item	Pencil	Binder	Pencil	Pen	Pe
6	Units	95	50	36	27	5
7	Unit Cost	1.99	19.99	4.99	19.99	2.9
8	Total	189.05	999.5	179.64	539.73	5
9						
10						

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.

In Excel, go to the **Home** tab, then hover over **Clear** and select **Clear Formats**.

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