✓ Congratulations! You passed!

Grade received 100% **To pass** 100% or higher

Go to next item

1.

1 / 1 point

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.

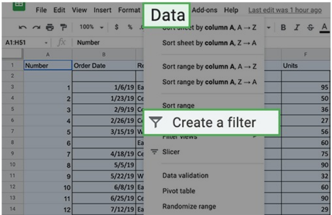
Data Spreadsheet for Cleaning with Spreadsheets
XLSX File

Select and remove blank cells

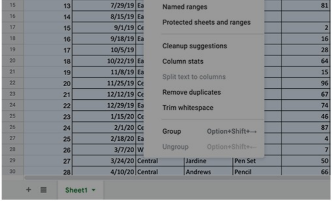
The first technique we'll use is to select and eliminate rows containing blank cells by using filters. To eliminate rows with blank cells:

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**.


File
Edit
View
Insert
Format
Data
Add-ons
Help
Last edit was 1 hour ago
100%
$
%
B
I
A
A1:H51
fx
Number
Sort sheet by column A, A → Z
Sort sheet by column A, Z → A
Sort range by column A, A → Z
Sort range by column A, Z → A
Sort range
Create a filter
Filter views
Slicer
Data validation
Pivot table
Randomize range

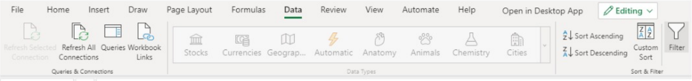
	A	B		F
1	Number	Order Date	Re	Units
2				
3	1	1/6/19	Ea	95
4	2	1/23/19	Ce	50
5	3	2/9/19	Ce	36
6	4	2/26/19	Ce	27
7	5	3/15/19	W	56
8	6		Ea	60
9	7	4/18/19	Ce	75
10	8	5/5/19		90
11	9	5/22/19	W	32
12	10	6/8/19	Ea	60
13	11	6/25/19	Ce	90
14	12	7/12/19	Ea	29



15	13	7/29/19	Ea	81
16	14	8/15/19	Ea	
17	15	9/1/19	Ce	2
18	16	9/18/19	Ea	16
19	17	10/5/19		28
20	18	10/22/19	Ea	64
21	19	11/8/19	Ea	15
22	20	11/25/19	Ce	96
23	21	12/12/19	Ce	67
24	22	12/29/19	Ea	74
25	23	1/15/20	Ce	46
26	24	2/1/20	Ce	87
27	25	2/18/20	Ea	4
28	26	3/7/20	W	7
29	27	3/24/20	Central | Jardine | Pen Set	50
30	28	4/10/20	Central | Andrews | Pencil	66

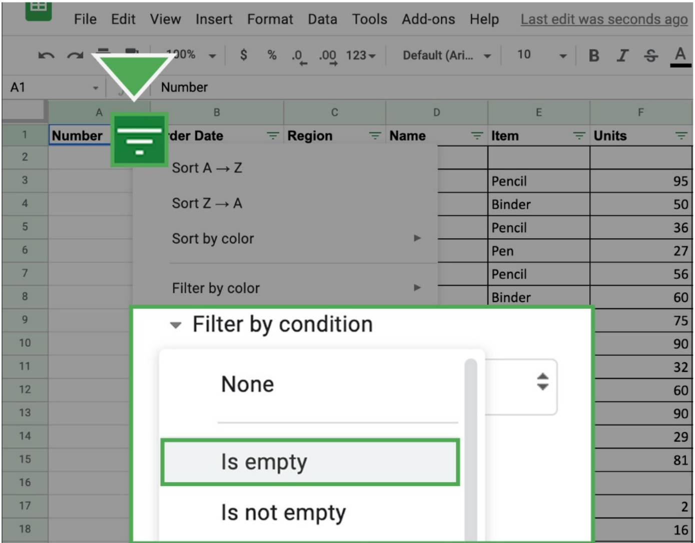
Named ranges
Protected sheets and ranges
Cleanup suggestions
Column stats
Split text to columns
Remove duplicates
Trim whitespace
Group Option+Shift+→
Ungroup Option+Shift+←
+
Sheet1

Excel:


File
Home
Insert
Draw
Page Layout
Formulas
Data
Review
View
Automate
Help
Open in Desktop App
Editing
Refresh Selected Connection
Refresh All Connections
Queries
Workbook Links
Queries & Connections
Stocks
Currencies
Geograp...
Automatic
Anatomy
Animals
Chemistry
Cities
Data Types
Sort Ascending
Sort Descending
Custom Sort
Filter
Sort & Filter

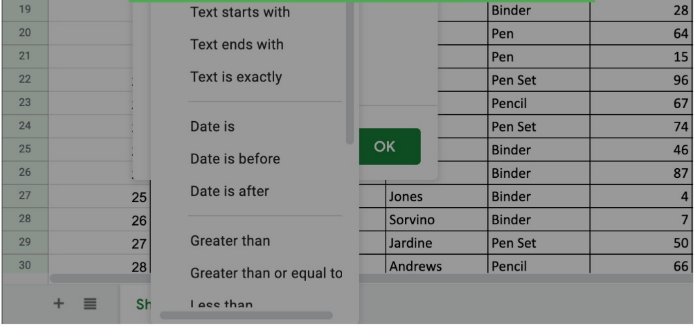
3. Every column now shows a green triangle in the first row next to the column title. Click the green triangle in **Column B** to access a new menu.

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



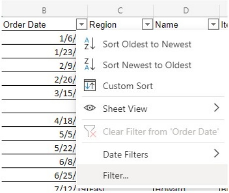
	A	B	C	D	E	F
1	Number	rder Date	Region	Name	Item	Units
2						
3					Pencil	95
4					Binder	50
5					Pencil	36
6					Pen	27
7					Pencil	56
8					Binder	60
9						75
10						90
11						32
12						60
13						90
14						29
15						81
16						
17						2
18						16

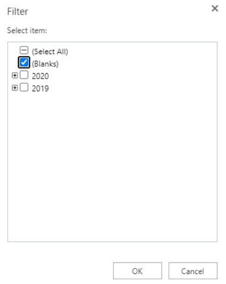
File
Edit
View
Insert
Format
Data
Tools
Add-ons
Help
Last edit was seconds ago
100%
$
%
.0
.00
123
Default (Ari...
10
B
I
S
A
A1
Number
Sort A → Z
Sort Z → A
Sort by color
Filter by color
Filter by condition
None
Is empty
Is not empty


Text starts with
Text ends with
Text is exactly
Date is
Date is before
Date is after
Greater than
Greater than or equal to
Less than
OK
Sh
19
20
21
22
23
24
25
26
27
28
29
30
25
26
27
28
Jones
Sorvino
Jardine
Andrews
Binder
Pen
Pen
Pen Set
Pencil
Pen Set
Binder
Binder
Binder
Binder
Pen Set
Pencil
28
64
15
96
67
74
46
87
4
7
50
66

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

Excel:


B
C
D
Order Date
Region
Name
It
1/6/
1/23/
2/9/
2/26/
3/15/
4/18/
5/5/
5/22/
6/8/
6/25/
7/12/
Sort Oldest to Newest
Sort Newest to Oldest
Custom Sort
Sheet View
Clear Filter from 'Order Date'
Date Filters
Filter...


Filter
×
Select item:
☐ (Select All)
☑ (Blanks)
☐ 2020
☐ 2019
OK
Cancel

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

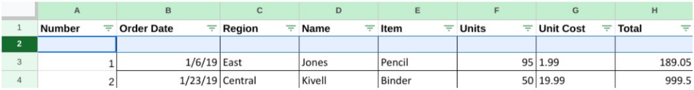
5. Select all these cells and delete the rows except the row of column headers.

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).

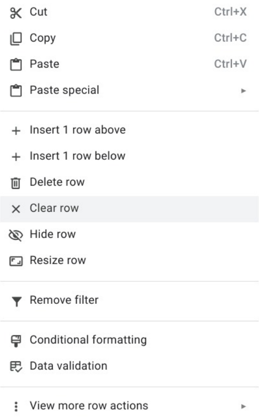
7. Repeat this for **Columns B-H**.

8. **Note:** If you simply deleted the data from the row by tapping the backspace button, you will need to go a step further and *delete the empty row entirely* by left-clicking the row number located on the furthest left side of the screen.



	A	B	C	D	E	F	G	H
1	Number	Order Date	Region	Name	Item	Units	Unit Cost	Total
2								
3	1	1/6/19	East	Jones	Pencil	95	1.99	189.05
4	2	1/23/19	Central	Kivell	Binder	50	19.99	999.5

9. Next, right click on the highlighted row to call up the drop down window, and select the **Delete row** option.


Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V
Paste special
Insert 1 row above
Insert 1 row below
Delete row
Clear row
Hide row
Resize row
Remove filter
Conditional formatting
Data validation
View more row actions

10. Continue to do this same operation for the remaining empty rows in the data set.

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

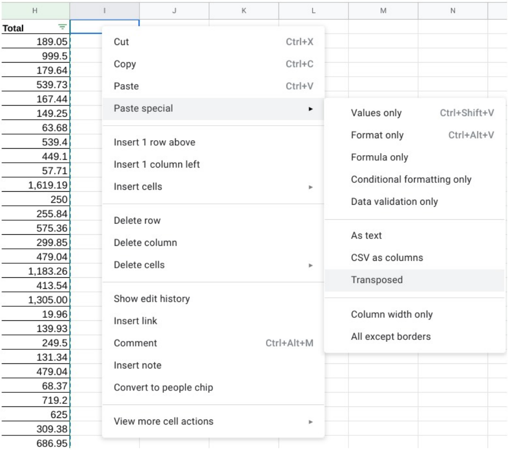
Transpose the data

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.

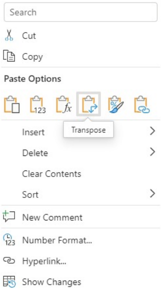
2. Right-click on **cell I1**. This is where you want the transposed data to start.

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
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
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

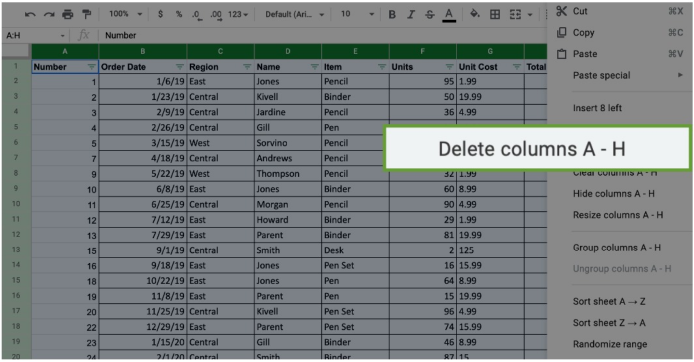
Total
189.05
999.5
179.64
539.73
167.44
149.25
63.68
539.4
449.1
57.71
1,619.19
250
255.84
575.36
299.85
479.04
1,183.26
413.54
1,305.00
19.96
139.93
249.5
131.34
479.04
68.37
719.2
625
309.38
686.95

Excel:


Search
Cut
Copy
Paste Options
Transpose
Insert
Delete
Clear Contents
Sort
New Comment
Number Format...
Hyperlink...
Show Changes

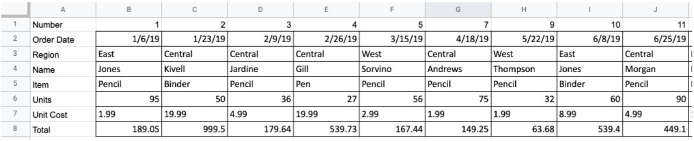
You should now find the data transformed into the new wide format. At this point, you should remove the original long data from the spreadsheet.

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**.


100%
Default (Ari...
10
A:H
Number
Cut ⌘X
Copy ⌘C
Paste ⌘V
Paste special
Insert 8 left
Delete columns A - H
Clear columns A - H
Hide columns A - H
Resize columns A - H
Group columns A - H
Ungroup columns A - H
Sort sheet A → Z
Sort sheet Z → A
Randomize range

Number	Order Date	Region	Name	Item	Units	Unit Cost	Total
1	1/6/19	East	Jones	Pencil	95	1.99	
2	1/23/19	Central	Kivell	Binder	50	19.99	
3	2/9/19	Central	Jardine	Pencil	36	4.99	
4	2/26/19	Central	Gill	Pen			
5	3/15/19	West	Sorvino	Pencil			
7	4/18/19	Central	Andrews	Pencil			
9	5/22/19	West	Thompson	Pencil	32	1.99	
10	6/8/19	East	Jones	Binder	60	8.99	
11	6/25/19	Central	Morgan	Pencil	90	4.99	
12	7/12/19	East	Howard	Binder	29	1.99	
13	7/29/19	East	Parent	Binder	81	19.99	
15	9/1/19	Central	Smith	Desk	2	125	
16	9/18/19	East	Jones	Pen Set	16	15.99	
18	10/22/19	East	Jones	Pen	64	8.99	
19	11/8/19	East	Parent	Pen	15	19.99	
20	11/25/19	Central	Kivell	Pen Set	96	4.99	
22	12/29/19	East	Parent	Pen Set	74	15.99	
23	1/15/20	Central	Gill	Binder	46	8.99	
24	2/1/20	Central	Smith	Binder	87	15	

Your screen should now appear like this:



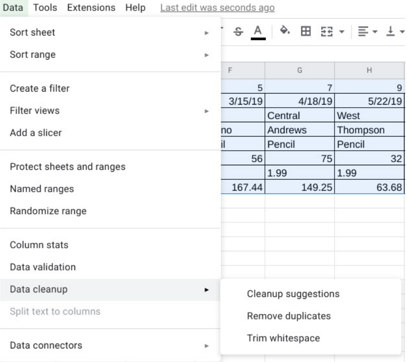
	A	B	C	D	E	F	G	H	I	J
1	Number	1	2	3	4	5	7	9	10	11
2	Order Date	1/6/19	1/23/19	2/9/19	2/26/19	3/15/19	4/18/19	5/22/19	6/8/19	6/25/19
3	Region	East	Central	Central	Central	West	Central	West	East	Central
4	Name	Jones	Kivell	Jardine	Gill	Sorvino	Andrews	Thompson	Jones	Morgan
5	Item	Pencil	Binder	Pencil	Pen	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.

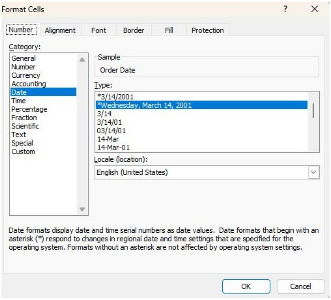
2. Click on the **Data** tab, then hover over **Data cleanup** and select **Trim whitespace**.


Data
Tools
Extensions
Help
Last edit was seconds ago
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
Cleanup suggestions
Remove duplicates
Trim whitespace

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

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 auto-fill square handle at the bottom right corner of the cell and pull it downward to call the rest of the data in the column without the white spaces.

Note: It's important to carry out the TRIM formula for each column individually, as it may affect certain types of cell values (like dates) undesirably. One option is right-clicking the "Order Date" column at the top, and selecting the **Format Cells** option to change the dates to string values before applying the TRIM formula.


Format Cells
?
Number
Alignment
Font
Border
Fill
Protection
General
Number
Currency
Accounting
Time
Percentage
Fraction
Scientific
Text
Special
Custom
*3/14/2001
3/14
3/14/01
03/14/01
14-Mar
14-Mar-01
OK
Cancel

Category: Date

Sample: Order Date

Type: *Wednesday, March 14, 2001

Locale (location): English (United States)

Date formats display date and time serial numbers as date values. Date formats that begin with an asterisk (*) respond to changes in regional date and time settings that are specified for the operating system. Formats without an asterisk are not affected by operating system settings.

For additional understanding of using the TRIM formula in Excel, please explore the Microsoft Excel guide on using the formula correctly.

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

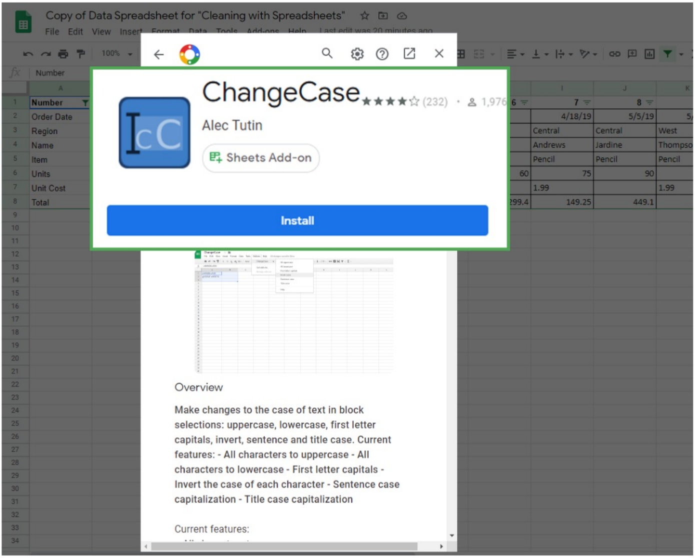
If you are using Microsoft Excel, this documentation explains how to use a formula to change the case of a text string. Follow these instructions to clean the string text and then move on to the confirmation and reflection section of this activity.

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.

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:


Copy of Data Spreadsheet for "Cleaning with Spreadsheets"
File Edit View Insert Format Data Tools Add-ons Help Last edit was 20 minutes ago
100%
Number
ChangeCase
★★★★☆ (232) • 1,976
Alec Tutin
Sheets Add-on
Install
Number
Order Date
Region
Name
Item
Units
Unit Cost
Total
Current features:

Overview

Make changes to the case of text in block selections: uppercase, lowercase, first letter capitals, invert, sentence and title case. Current features: - All characters to uppercase - All characters to lowercase - First letter capitals - Invert the case of each character - Sentence case capitalization - Title case capitalization

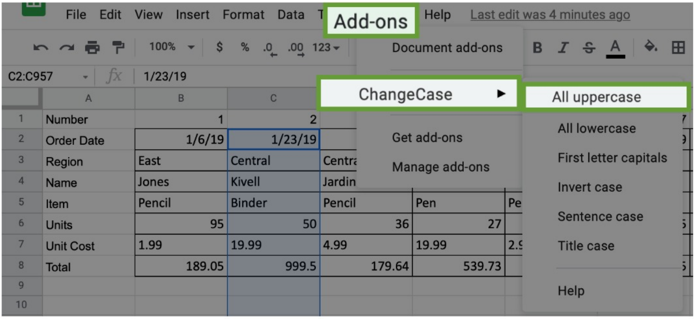
4. Click on **Install** to install the add-on. It may ask you to login or verify the installation permissions.

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.


File Edit View Insert Format Data
Add-ons
Help
Last edit was 4 minutes ago
100%
C2:C957
1/23/19
Document add-ons
ChangeCase
All uppercase
All lowercase
First letter capitals
Invert case
Sentence case
Title case
Help
Get add-ons
Manage add-ons
Number 1 2
Order Date 1/6/19 1/23/19
Region East Central Centra
Name Jones Kivell Jardin
Item Pencil Binder Pencil Pen Pe
Units 95 50 36 27
Unit Cost 1.99 19.99 4.99 19.99 2.9
Total 189.05 999.5 179.64 539.73

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.

Confirmation and reflection

Review the final product of the spreadsheet you cleaned during this activity. Which of the following is the rightmost column?

○ Column AZ
● Column AS
○ Column AA
○ Column Z

✓ **Correct**
In the final product of this activity, the rightmost column is Column AS. You are able to find this information after you properly transpose the data. Going forward, you can apply what you learned about data cleaning and transposing to work with data in the future.

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:

1 / 1 point

- 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, what would you do differently? Why?

1. **Most Challenging Part:** Identifying and removing extra spaces in text strings was the most challenging part for me. While the TRIM function exists, it could potentially affect other types of cell values (dates) undesirably. It required additional research and understanding of the data types to apply the solution effectively.

2. **Importance of Cleaning and Transposing:** Cleaning data ensures accuracy and reliability, eliminating errors and inconsistencies that can skew analysis. Transposing helps reshape data into a format more suitable for specific tasks, like comparing across different categories or visualizing trends. Both are crucial for drawing meaningful conclusions from the data.

3. **Doing it Differently:** If I had to clean this data again, I would:

* **Prioritize data format:** Before diving into cleaning, I'd clarify the target format needed for analysis, guiding the cleaning process and potentially avoiding unnecessary transposing steps later.
* **Explore alternative tools:** Depending on the data complexity, I might utilize dedicated data cleaning tools like OpenRefine instead of relying solely on spreadsheets for better efficiency and handling more intricate issues.
* **Document cleaning steps:** Keeping a detailed record of cleaning actions, decisions, and challenges would save time and effort in future revisions or collaborations.

These learnings would allow me to approach future data cleaning tasks with more strategic planning and efficient techniques, resulting in cleaner and more readily usable data for analysis.

✓ **Correct**
Congratulations on completing this hands-on activity! In this activity, you cleaned and transposed data on a spreadsheet. A good response would include 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.