COLLECT MEANINGFUL DATA: Data Quality – Excel Tips and Tricks

More emPower Tools
+ learn more about each topic
thecapacitycollective.org/resources

Purpose: Data are crucial for making program decisions, reporting to funders, tracking emerging client issues and delivering high quality services. You want your data quality to be as high as possible so you can make decisions with confidence. In this tool, we share our favorite tips for using Excel to check data quality. The same concepts work for Google Sheets. *For more details on what makes data high quality, and for tips and tricks on collecting high quality data, also see the* ***Collect Meaningful Data: Data Quality Tips and Tools*** *emPower Tool.*

How do I get started?

1. Export the Data as .csv or .xlsx (if needed)
2. Open file in Excel
3. Save a new copy of the file so no data are lost while you work
4. Clean up data (remove columns you do not need)

What Data Quality Issues Do I Look For?

Blanks	Outliers	Duplicates	Spaces	Variations	Error Codes
Data that are missing	Numbers higher or lower than reasonable	Same data point (like a person's name) listed more than once	Extra spaces before, after or between words	Using different words or spellings for the same thing	Formulas that are broken or not working properly

How Do I Find Data Quality Issues?

Sort + Filter the Data

(Home Ribbon > Sort and Filter)

Sorting: The *sorting* tool allows you to put the data into a particular order by date, number, alphabetic order and more.

Data Quality: Look at the top and bottom of the sorted data. Look for outliers that are lower/higher than expected numbers (like a 1902 or 2118 birth year).

Filtering: The *filtering* tool allows you to isolate particular data based on criteria you choose, and hides the rest, so you can focus on just that data (like particular clients, or a specific group or date).

Data Quality: Filtering will cluster similar answers together, so you can see if there are duplicates, misspellings, spaces, and so forth. Click "(Select All)" to deselect all of the options, then click on the checkbox of the data you want to isolate.

Conditional Formatting

(Home Ribbon > Conditional Formatting)

- *Conditional Formatting* allows you to apply specific formatting (such as fill color and/or font color) to cells that meet certain criteria (such as above, below or equal to a value you define)
- For example, use color to highlight, emphasize or differentiate among data in a spreadsheet
- Once you apply the conditional formatting, all cells with values outside the criteria range you set will appear in the formatting you specified so you can see them.
- For example, conditional formatting can show you all duplicate participant names

How Do I Apply Conditional Formatting?

1. Select the column(s) you want to format
2. Home Ribbon > Conditional Formatting
3. Hover on type of rule (Highlight or Top/Bottom)
4. Select the rule that best applies
5. Specify the criteria (threshold value)
6. Select the formatting you want to apply

"Highlight Cells" Rules		Conditional Formatting Options
Rule Type	**Example Use**	**Choose the text and fill colors**
Greater Than…	People with income over $xxx may not be eligible for services	Light Orange Fill with Dark Orange Text
Less Than…	Anyone with fewer than 1 home visit/month may not be active	Red Border
Between…	Looking for all the children between ages 0 and 3 for outreach	Light Orange Fill
Equal To…	Specifically highlight clients receiving TANF funding	Solid Fill
Text That Contains…	Any case notes that mention "success" to pull out for reporting	Gradient Fill
A Date Occurring…	Children's birth dates should logically fall between 2015-2020	***Note: You can apply multiple conditional formatting rules to the same data so you can see various patterns at the same time!***
Duplicate Values…	Client seems to have two kids with the same name: possible error	

How Do I FIX Data Quality Issues?

Manually Fix Data

If you are *sure* you know what the data should be, manually enter the correct info into the cell.

Find + Replace

If you need to make more than a few of the same changes (like a common spelling error) use *Find* (Ctrl+F) to locate all of the incorrect data and *Replace* to change all of the same error at once.

Triple check for accuracy!

How Do I PREVENT Data Quality Issues?

Data Culture

Get buy-in: *show* staff why data quality matters and involve them in data decisions *(see the **Data Quality Tips and Tools** and **Create a Data Culture** emPower Tools).*

Spot Checks

Do not wait for reporting time. Check early and often and give feedback as needed. **Integrate data checks into your routine.**

Excel: Data Validation

Excel allows you to create dropdown options for cells so you can choose an answer rather than typing from scratch each time.

Other Excel Tricks

Formulas

=COUNTBLANK counts the number of blank cells in the selected data.

=MAX/=MIN shows you the high/low numbers in selected data.

Error Codes

Find and resolve all of the error messages in your data and formulas. Look for red triangles or error codes in cells to identify potential issues.

Cell Formats (*Home Ribbon*)

If anything seems strange, check the *cell formats*. Make sure Excel is reading dates as dates, numbers as numbers and text as text.

We encourage you to share these resources with your organization, and other local social service organizations. PLEASE NOTE this handout is the intellectual property of The Capacity Collective. Please do not duplicate parts, or adapt, without the express permission of The Capacity Collective. Thank you for supporting our work!