Keeping data organized with sorting and filters

You have learned about four phases of analysis:

- Organize data

- Format and adjust data

- Get input from others

- Transform data

The organization of datasets is really important for data analysts. Most of the datasets you will use will be organized as tables. Tables are helpful because they let you manipulate your data and categorize it. Having distinct categories and classifications lets you focus on, and differentiate between, your data quickly and easily.

Data analysts also need to format and adjust data when performing an analysis. **Sorting** and **filtering** are two ways you can keep things organized when you format and adjust data to work with it. For example, a filter can help you find errors or outliers so you can fix or flag them before your analysis. **Outliers** are data points that are very different from similarly collected data and might not be reliable values. The benefit of filtering the data is that after you fix errors or identify outliers, you can remove the filter and return the data to its original organization.

In this reading, you will learn the difference between sorting and filtering. You will also be introduced to how a particular form of sorting is done in a pivot table.

Sorting versus filtering

Sorting vs Filtering
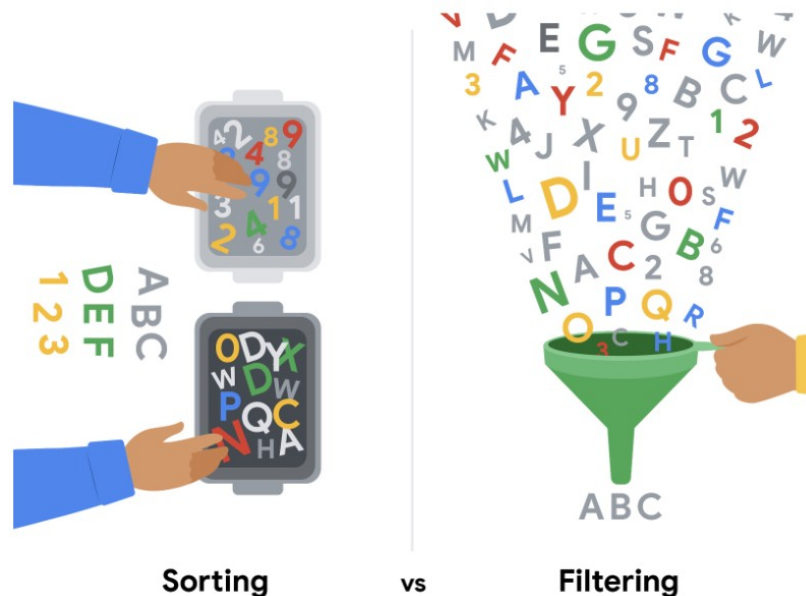
Sorting is when you arrange data into a meaningful order to make it easier to understand, analyze, and visualize. It ranks your data based on a specific metric you choose. You can sort data in spreadsheets, SQL databases (when your dataset is too large for spreadsheets), and tables in documents.

For example, if you need to rank things or create chronological lists, you can sort by ascending or descending order. If you are interested in figuring out a group's favorite movies, you might sort by movie title to figure it out. Sorting will arrange the data in a meaningful way and give you immediate insights. Sorting also helps you to group similar data together by a classification. For movies, you could sort by genre -- like action, drama, sci-fi, or romance.

Filtering is used when you are only interested in seeing data that meets a specific criteria, and hiding the rest. Filtering is really useful when you have lots of data. You can save time by zeroing in on the data that is really important or the data that has bugs or errors. Most spreadsheets and SQL databases allow you to filter your data in a variety of ways. Filtering gives you the ability to find what you are looking for without too much effort.

For example, if you are only interested in finding out who watched movies in October, you could use a filter on the dates so only the records for movies watched in October are displayed. Then, you could check out the names of the people to figure out who watched movies in October.

To recap, the easiest way to remember the difference between sorting and filtering is that you can use sort to quickly order the data, and filter to display *only* the data that meets the criteria that you have chosen. Use filtering when you need to reduce the amount of data that is displayed.

It is important to point out that, after you filter data, you can **sort the filtered data**, too. If you revisit the example of finding out who watched movies in October, after you have filtered for the movies seen in October, you can then sort the names of the people who watched those movies in alphabetical order.

Sorting in a pivot table

Items in the row and column areas of a pivot table are sorted in ascending order by any custom list first. For example, if your list contains days of the week, the pivot table allows weekday and month names to sort like this: Monday, Tuesday, Wednesday, etc. rather than alphabetically like this: Friday, Monday, Saturday, etc.

If the items aren't in a custom list, they will be sorted in ascending order by default. But, if you sort in descending order, you are setting up a rule that controls how the field is sorted even after new data fields are added.