Lecture Note #8: Filtering and Sorting Data

BUSI 201: Business Data Analysis

Topic 1. Tables and Data

Although "Data" and "Tables" often look similar, they play different roles in data analysis. A table is a product of data visualization, which aims to give the reader a glimpse into the properties of the data. For an example, please see Figure 1. Figure 1 contains data on three publicly traded stocks, BRK-B, DIS, and TSLA. The data spans three months from June 1st, 2023 up to Aug. 31st 2023, and has information on daily prices and trade volume of each stock.

Figure 1: Data and Tables

The left of Figure 1 is what we call "data." Data can be visualized as "information stacked vertically," where each column contains a different variable. For instance, in the left-hand-side of TABLESvsDATA, we can see that the first column contains information in the ticker, third column containing information on the date of observation, and so on.

Meanwhile, the right hand side of Figure 1 contains information from the data, formatted into a table. Here we use the AVERAGEIFS function to find the relevant information for each cell in the table. Usually, it will be much easier to glean information off from a table compared to the raw data. It would be incredibly difficult to find a clear trend in the average prices or trade volume fron the left hand side entry.

Topic 2. Basic Data Analysis: Filtering

We first look into "filters" in Excel as a means to directly analyze data. Suppose that at the moment we are only interested in stock price information for BRK-B. Filtering offers a handy tool in this situation. To apply filters...
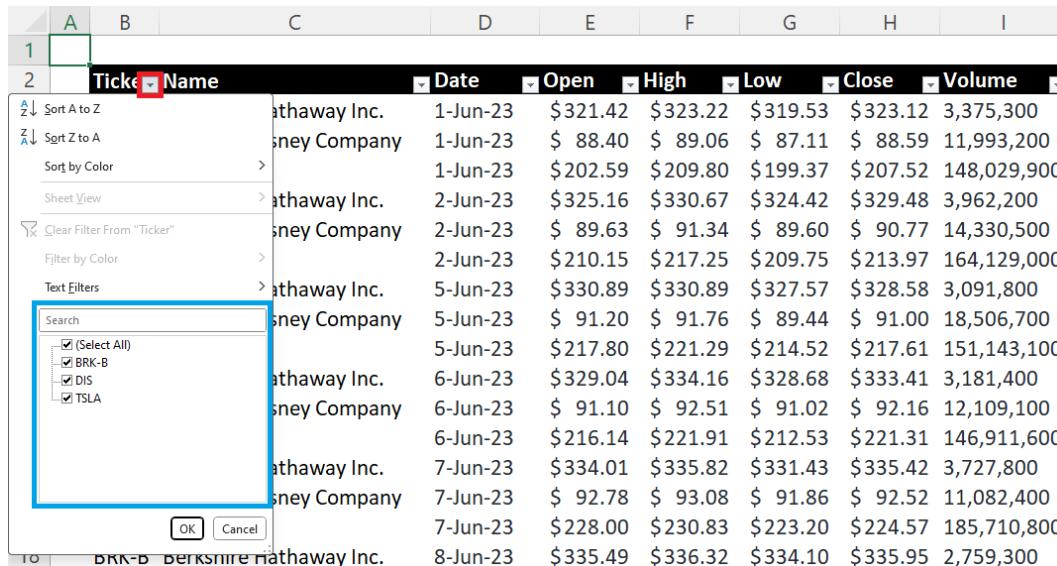
1. Select the variable names of the data to be filtered. In our example, it should be cells B2:I2.

2. On the Quick Access Toolbar, navigate to DATA.

3. Click Filter.

Figure 2: Applying Filters to Data

As an alternate, you can select the variable names, and then use the hotkey ctrl+shift+L. Once the filters are added, you will see a upside-down triangle mark next to each of the variables in the dataset.


Figure 3: Filtering by Ticker Symbol

2

Clicking the upside-down triangle mark in the red box in Figure 3, you will see that currently all three types of tickers have a check mark next to it in the blue of in Figure 3. This blue box serves two main purposes for us at this point:

1. It shows us the list of all entries under the `Ticker` variable in the data, and

2. It shows us which ones are currently "filtered/unfiltered."

Click on the items `DIS` and `TSLA` in the blue box in Figure 4 so that a check mark only appears next to `BRK-B`, and that next to `(Select All)` is now a black square as shown in Figure 4, and click `Ok` in the green box.

Figure 4: Filtering by Ticker Symbol (cont'd)

Once the filter is applied, the data will resemble Figure 5. The first thing we can notice is that now we only see the data from `BRK-B`, and the others are "hidden away." Second, notice that the upside-down triangle next to `Ticker` is now slightly difference (in the red box), which tells us that a filter is active on `Ticker`. Finally, the row number indicator to the left hand side (in the blue box) is now blue, and has missing numbers, which also indicates that filters are active on this worksheet.

Figure 5: Filtered by Ticker Symbol

Filtering by Dates

In addition to filtering data out by exact text entries, we can sort data by dates as well. Suppose that we want to filter information so that we only see stock trading information for the month of August 2023.

Figure 6: Filtering by Date

Click on the upside-down triangle to the right of Date in the red box in Figure 5. Then we can head to the region in the green box, and deselect the months of June and July. Alternatively, we can see more options by heading to Date Filters in the blue box. Clicking on Date Filters will lead to Excel showing more advanced options to sort by date as shown in the purple box in Figure 5.

Figure 7: Filtered by Date

Filtering by Numbers

As we were able to filter the data by numbers, it is also possible to filter data by numbers. Suppose for instance, that wer are trying to only view data for stock prices only when the trade volume is above 100,000,000 shares. This time, click on the upside-down triangle next to `Volume`(red box in Figure 8).

Figure 8: Filtering by Numbers

Similar to the previous case when applying filters to data, we can head to the region in the green box and select every entry that is greater than 100,000,000 manually. However, this is a rather tedious process and is prone to mistakes. So we should head to `Number Filters` in the blue box. Clicking on `Number Filters` and `Greater than or Equal to` leads to a new window pop-up as seen in Figure 9.
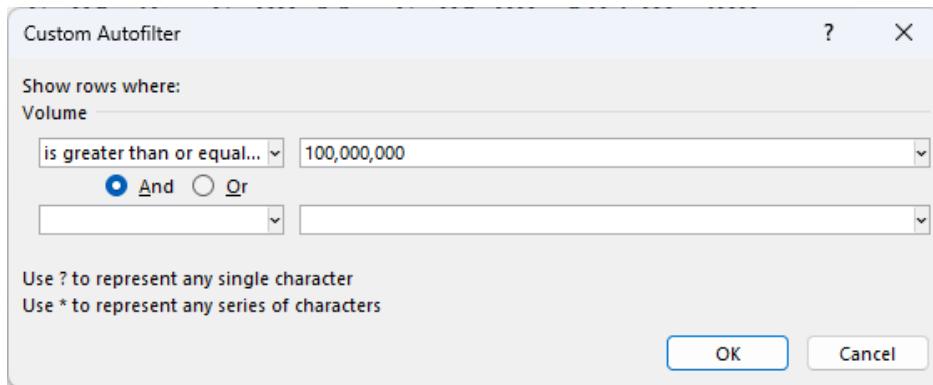
Figure 9: Greater than or Equal to

Inserting 100,000,000 and clicking `OK` will lead to the data being filtered so that only the entries with greater than 100,000,000 shares being traded appear on the worksheet. Note that what we are doing here is not "filtering the stock that had more than 100,000,000 shared being traded at any time." You will notice that while TSLA is the only stock in the current data that has days where more than 100,000,000 shares were traded in the market, there are some days missing from the sorted result.

5

Using Wildcards in Filters

Recall Wildcards that we covered in a previous lecture note. Sometimes, we may want to create a filter that only displays items that contain a certain word or a certain sequence of numbers. We can use wildcard characters to achieve this goal.

Wildcard	Function
*	Replaces an arbitrary number of characters in the search. (i.e. North* will return North Carolina and North Dakota)
?	Replaces one character in the search. (i.e. N? returns NE, NV, NH, NJ, NM, NY, NC, ND)
~	Used to override other wildcards. (i.e. N~* returns N*)

An example of using wildcard characters in filters can be found in Figure 10. Using the term `*Inc*`, we are able to find any entries under `Name` that has the term `Inc` in the official company name listed in the NYSE.
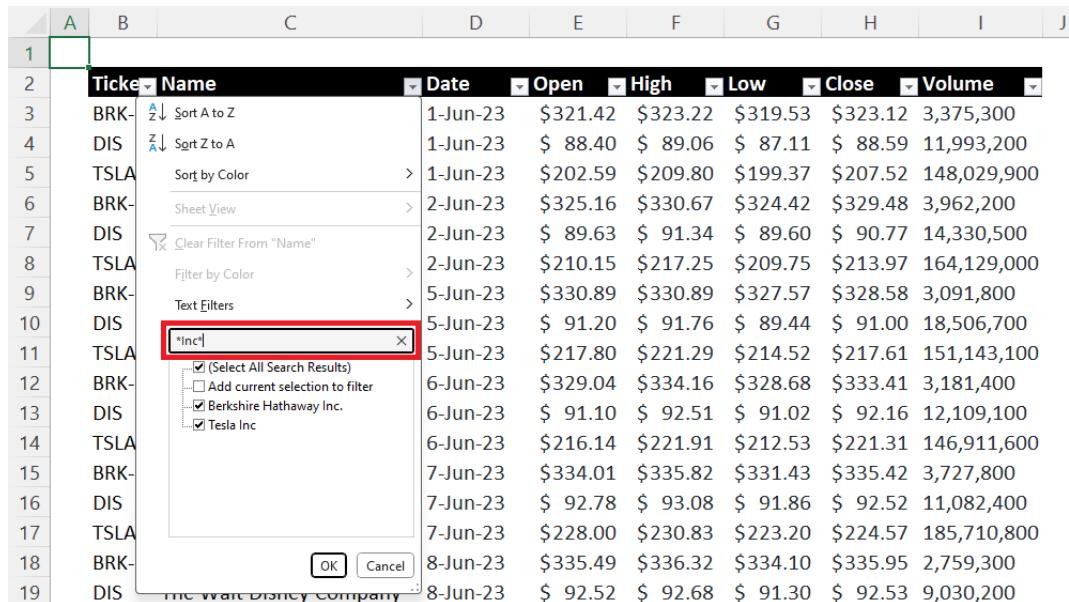
Figure 10: Wildcards in Filters

Topic 3. Basic Data Analysis: Sorting using Filters

Please navigate to worksheet `FILTER` in the workbook. This sheet contains a larger dataset that is similar to the previous one in the `TABLESvsDATA` sheet. Here we have 8 different assets' price information for the month of August 2023. We now look into filters as a tool in Excel as a means to sort data, which will allow us to directly analyze data. As it stands, the data is filtered in the following order:

1. Date (Earlier date first)

2. Ticker (Alphabetical)

3. Type of Asset (Alphabetical)

To see how filters can work to sort data, select the cells that contain variable names (B2:J2), and apply filters. Suppose that we want sort the data so that we have the data sorted by `Ticker` alphabetically so that all entries of BRK-B appear before any entry of DAL, which comes before any entry of DIS, and so forth...

Figure 11: Sorting by Ticker Symbol

To acheive this task, click on the upside-down triangle in the red box in Figure 11, and select `Sort A to Z` to sort alphabetically.[1] However, we may want to sort the data by multiple variables...

---
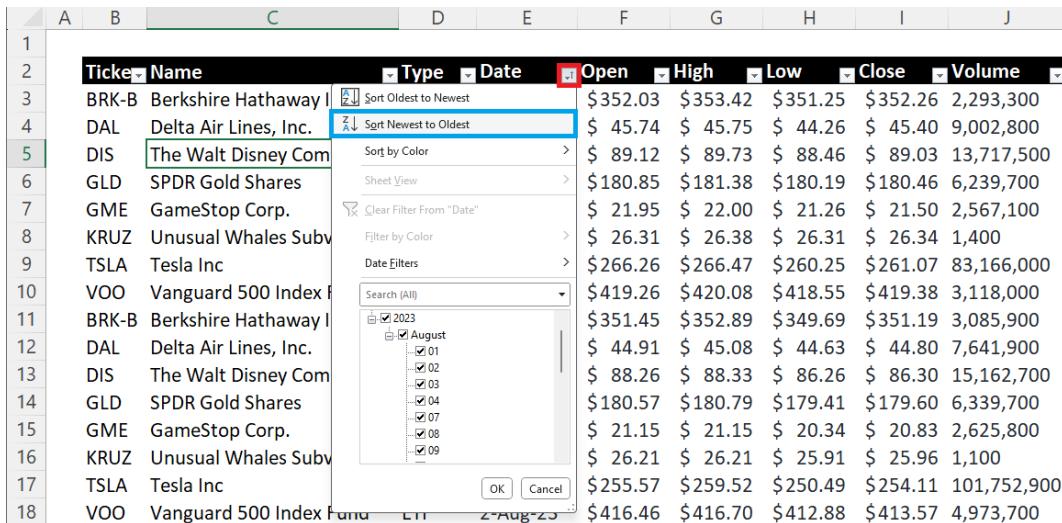
[1]Choosing `Sort Z to A` will result in the opposite result.

Sorting by Multiple Variables

Now suppose that we want to sort the data in the following order:

1. Ticker (Alphabetical)

2. Date (Earlier date first)

That is, we want to sort by `Ticker` alphabetically as we did in the previous example, but within the alphabetically sorted `Ticker`, have the most recent trade information on top. To get this result, we must first sort by `Date`, and then sort by `Ticker`.

Figure 12: Sorting by Ticker-Date: Date

Figure 13: Sorting by Ticker-Date: Ticker

The results of the sorting from Figures 12 and 13 is displayed in Figure 14 below. You can see that all entries of BRK-B comes first, with each entry of BRK-B being sorted from newest to oldest information. Scroll down a bit, and you will see that the next entries will be DAL sorted from newest to oldest stock market data.

	A	B	C	D	E	F	G	H	I	J
1										
2		Ticker	Name	Type	Date	Open	High	Low	Close	Volume
3		BRK-B	Berkshire Hathaway Inc.	Stock	31-Aug-23	$362.18	$362.47	$359.25	$360.20	2,842,300
4		BRK-B	Berkshire Hathaway Inc.	Stock	30-Aug-23	$358.63	$362.68	$358.60	$361.06	3,058,300
5		BRK-B	Berkshire Hathaway Inc.	Stock	29-Aug-23	$355.04	$358.59	$354.01	$358.29	2,285,600
6		BRK-B	Berkshire Hathaway Inc.	Stock	28-Aug-23	$357.89	$358.41	$354.53	$355.55	1,728,000
7		BRK-B	Berkshire Hathaway Inc.	Stock	25-Aug-23	$354.99	$357.35	$352.92	$355.93	2,136,800
8		BRK-B	Berkshire Hathaway Inc.	Stock	24-Aug-23	$354.35	$357.23	$354.13	$354.30	2,521,100
9		BRK-B	Berkshire Hathaway Inc.	Stock	23-Aug-23	$351.63	$354.32	$351.54	$354.26	2,239,500
10		BRK-B	Berkshire Hathaway Inc.	Stock	22-Aug-23	$353.01	$353.50	$349.66	$350.57	2,363,300

Figure 14: Sorted by Ticker-Date

Order of Sorting

Reversing the order of sorting, that is by sorting on `Ticker` first, and then `Date`, we find the following incorrect result in Figure 15. We can see that while the `Ticker` is indeed alphabetically sorted, the sorting in the `Date` variable is of a higher "priority," as it was sorted last.

	A	B	C	D	E	F	G	H	I	J
1										
2		Ticker	Name	Type	Date	Open	High	Low	Close	Volume
3		BRK-B	Berkshire Hathaway Inc.	Stock	31-Aug-23	$362.18	$362.47	$359.25	$360.20	2,842,300
4		DAL	Delta Air Lines, Inc.	Stock	31-Aug-23	$43.54	$43.58	$42.43	$42.88	6,506,500
5		DIS	The Walt Disney Company	Stock	31-Aug-23	$84.46	$84.61	$83.64	$83.68	10,448,000
6		GLD	SPDR Gold Shares	ETF	31-Aug-23	$180.37	$180.49	$179.94	$180.02	4,795,600
7		GME	GameStop Corp.	Stock	31-Aug-23	$18.41	$19.42	$18.15	$18.55	4,674,100
8		KRUZ	Unusual Whales Subversive	ETF	31-Aug-23	$25.70	$25.70	$25.70	$25.70	400
9		TSLA	Tesla Inc	Stock	31-Aug-23	$255.98	$261.18	$255.05	$258.08	108,861,700
10		VOO	Vanguard 500 Index Fund	ETF	31-Aug-23	$414.98	$416.12	$413.66	$413.83	4,751,100
11		BRK-B	Berkshire Hathaway Inc.	Stock	30-Aug-23	$358.63	$362.68	$358.60	$361.06	3,058,300
12		DAL	Delta Air Lines, Inc.	Stock	30-Aug-23	$43.78	$43.85	$43.22	$43.23	5,511,900
13		DIS	The Walt Disney Company	Stock	30-Aug-23	$84.40	$84.68	$83.83	$84.28	11,816,600
14		GLD	SPDR Gold Shares	ETF	30-Aug-23	$180.58	$180.88	$180.29	$180.32	6,593,700
15		GME	GameStop Corp.	Stock	30-Aug-23	$18.20	$18.52	$17.91	$18.37	1,884,500
16		KRUZ	Unusual Whales Subversive	ETF	30-Aug-23	$25.60	$25.74	$25.60	$25.71	4,000
17		TSLA	Tesla Inc	Stock	30-Aug-23	$254.20	$260.51	$250.59	$256.90	121,988,400
18		VOO	Vanguard 500 Index Fund	ETF	30-Aug-23	$413.03	$415.03	$412.40	$414.36	3,004,500

Figure 15: Sorted by Date-Ticker (Incorrect)

Thus, if you want the data should be sorted by variables A first, B second, and then C in that order, you must apply the sorting via filters in the opposite order of C > B > A.

Range of Sorting

Selecting the correct range to apply filters is critical when it comes to using the filters to sort data. Return to worksheet FILTER and remove the filter from cells B2:J2 by selecting this region, and clicking Data > Filter. Now apply the filter for cells B2:I2, leaving Volume out when we are setting filters as shown in Figure 16.

	A	B	C	D	E	F	G	H	I	J
1										
2		Ticker	Name	Type	Date	Open	High	Low	Close	Volume
3		BRK-B	Berkshire Hathaway Inc.	Stock	1-Aug-23	$352.03	$353.42	$351.25	$352.26	2,293,300
4		DAL	Delta Air Lines, Inc.	Stock	1-Aug-23	$45.74	$45.75	$44.26	$45.40	9,002,800
5		DIS	The Walt Disney Company	Stock	1-Aug-23	$89.12	$89.73	$88.46	$89.03	13,717,500
6		GLD	SPDR Gold Shares	ETF	1-Aug-23	$180.85	$181.38	$180.19	$180.46	6,239,700
7		GME	GameStop Corp.	Stock	1-Aug-23	$21.95	$22.00	$21.26	$21.50	2,567,100
8		KRUZ	Unusual Whales Subversive	ETF	1-Aug-23	$26.31	$26.38	$26.31	$26.34	1,400
9		TSLA	Tesla Inc	Stock	1-Aug-23	$266.26	$266.47	$260.25	$261.07	83,166,000

Figure 16: Incorrect Filter Applied

If we sort this data using the filters, we will see that the variable Volume is left out when sorting the data. Notice that the Volume data did not change when we sorted Ticker alphabetically.

	A	B	C	D	E	F	G	H	I	J
1										
2		Ticker	Name	Type	Date	Open	High	Low	Close	Volume
3		BRK-B	Berkshire Hathaway Inc.	Stock	1-Aug-23	$352.03	$353.42	$351.25	$352.26	2,293,300
4		BRK-B	Berkshire Hathaway Inc.	Stock	2-Aug-23	$351.45	$352.89	$349.69	$351.19	9,002,800
5		BRK-B	Berkshire Hathaway Inc.	Stock	3-Aug-23	$350.29	$354.47	$349.42	$353.81	13,717,500
6		BRK-B	Berkshire Hathaway Inc.	Stock	4-Aug-23	$353.99	$355.11	$349.39	$349.99	6,239,700
7		BRK-B	Berkshire Hathaway Inc.	Stock	7-Aug-23	$355.73	$364.63	$355.15	$362.58	2,567,100
8		BRK-B	Berkshire Hathaway Inc.	Stock	8-Aug-23	$359.42	$364.25	$358.85	$363.73	1,400

Figure 17: Consequence of Incorrect Filters