

Excel Level 1 Workshop

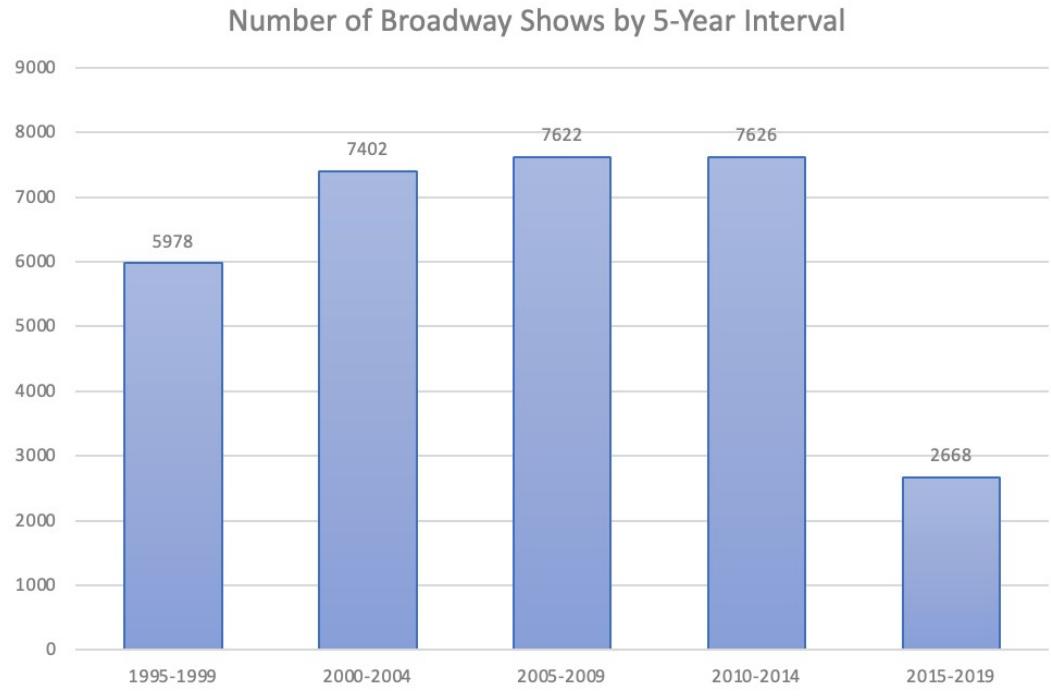
Joan Zheng

Graduate Fellow, Quantitative and Computing Lab at Claremont McKenna College



Let's learn fundamental Excel skills with real-world data analysis! Using actual Broadway box office data from 1995–2019, let's learn to organize, analyze, and visualize theater performance trends with this workshop! This tutorial is beginnerfriendly and designed for users with no prior Excel experience.

By the end of this workshop, we'll create an Excel spreadsheet showcasing historical trends of *Broadway* shows from 1995–2019.



We'll have created a histogram of Broadway attendance information by the end of this workshop.

1. Start Here: Let's Set Up Excel

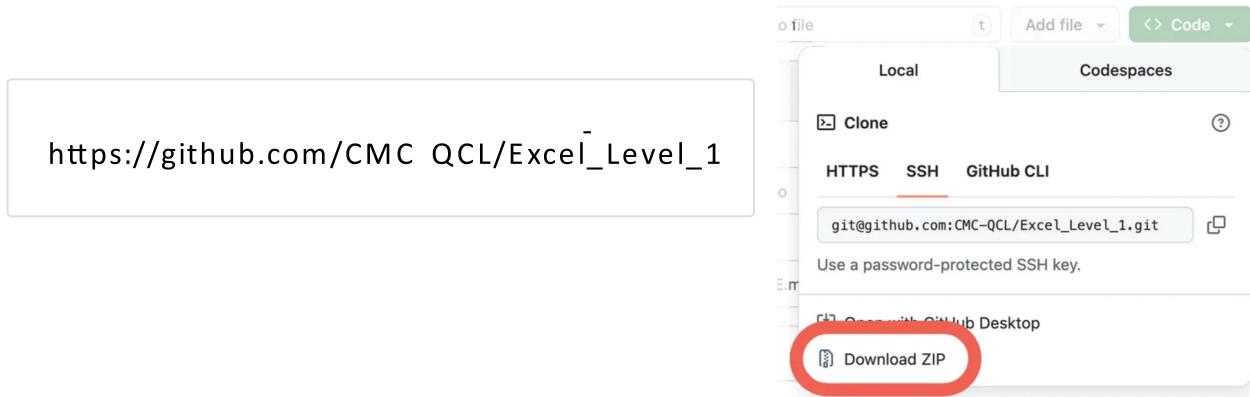
- Don't have Excel? Download
Microsoft Office 365 from CMC IT.

Office 365

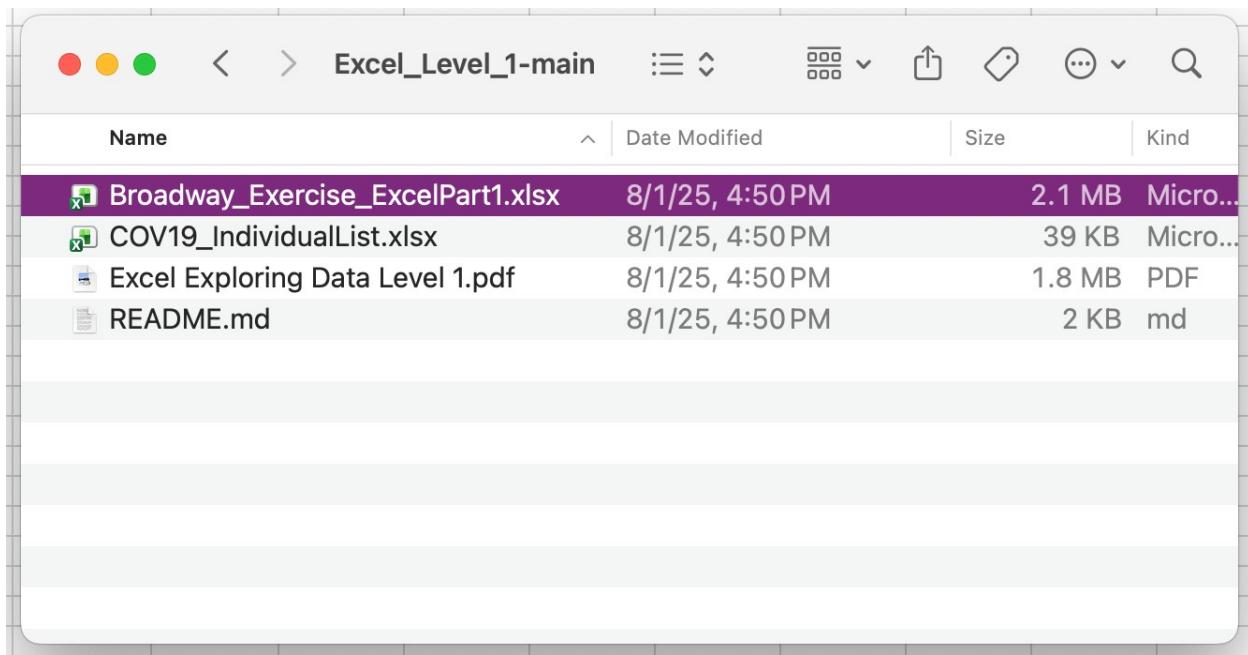
Office 365 is a Microsoft productivity suite that comes with a number of online and local resources, including Word, Excel, PowerPoint,

 <https://www.cmc.edu/information-technology/office-365>

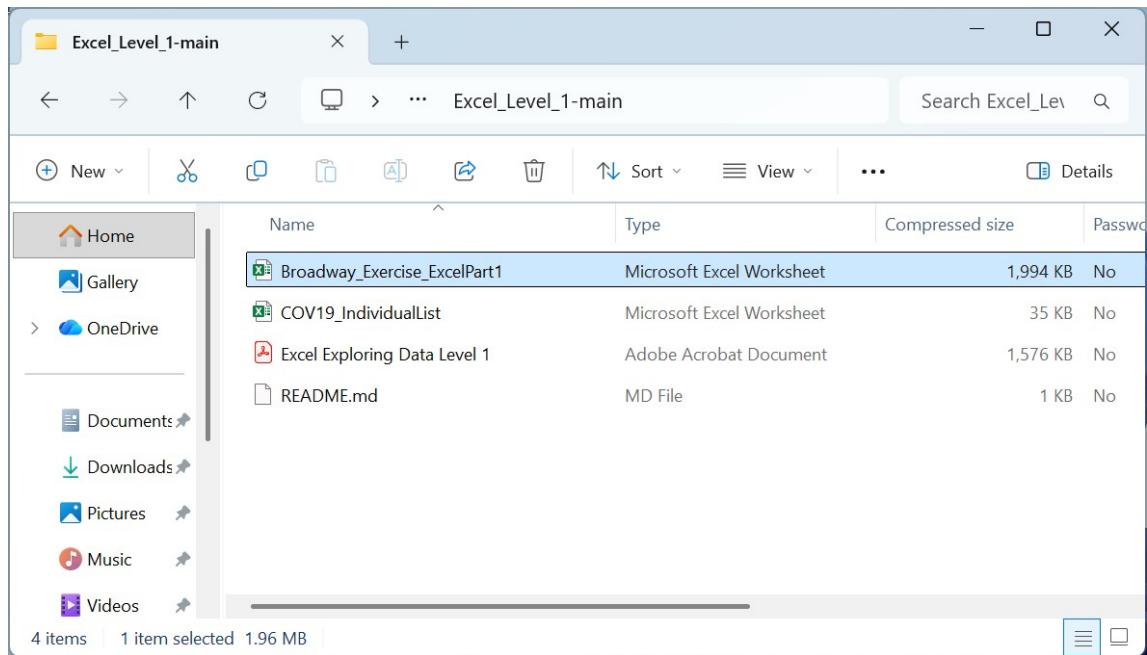
- Once you have Excel, download the ZIP file with our workshop materials.



- Extract workshop by double clicking on the downloaded ZIP file.
- Our extracted folder has four files. The spreadsheet that we'll use first is **Broadway_Exercise_ExcelPart1.xlsx**, which contains attendance statistics of Broadway shows at multiple theatres.



MacOS



Windows

- The file we will be using first is **Broadway_Exercise_ExcelPart1.xlsx**. Double click this file to open it in Excel.

2. A Brief Overview of Excel

The screenshot shows a Microsoft Excel window with the following labels:

- Ribbon:** The top bar containing tabs like Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, and Automate.
- Workbook:** The main area where the data is displayed.
- Column:** A vertical line of data within the Excel worksheet.
- Row:** A horizontal line of data within the Excel worksheet.
- Cell:** An intersection between a horizontal row and a vertical column.
- Tab:** Represents a single spreadsheet within a workbook.

The data in the worksheet includes columns for Date.Day, Date.Full, Date.Month, Date.Year, Show.Name, Show.Theatre, Show.Type, Statistics.Attendance, and Statistics.Capacity. The data spans from row 1 to 34, showing various Broadway shows and their details.

Terms:

- **Ribbon:** Area that contains the tabs of Home, Insert, Draw, Page Layout, Formulas, Data, Review, View, and Automate.
- **Workbook:** Another word for Excel file that contains at least one worksheet or multiple sheets within the book.
- **Row:** A horizontal line of data within the Excel worksheet.
- **Column:** A vertical line of data within the Excel worksheet.
- **Cell:** An intersection between a horizontal row and a vertical line in a worksheet.
- **Tab:** Represents a single spreadsheet within a workbook.



Exercise 1: The **Key** tab contains information about the columns of our dataset. What is the description of the **Statistics.Performances** column?

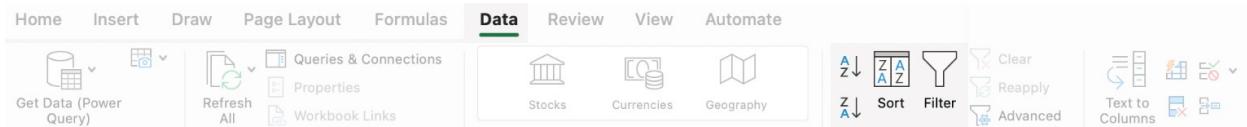


Exercise 2: How many rows are in our *Broadway* spreadsheet?

Hint: Use Ctrl/Cmd (⌘) Down Arrow (↓) to move to the bottom of the sheet.

3. Key Excel Features

- Sort and Filter can be viewed under the *Data* tab on the ribbon.

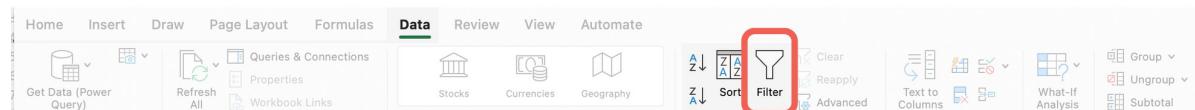


- Conditional Formatting can be accessed under the *Home* tab on the ribbon.

Date.Day	Date.Full	Date.Month	Date.Year	Show.Name
24	1/24/16	1	2016	The Humans
3	4/3/16	4	2016	Long Day's Journey Into Nig
29	5/29/16	5	2016	An Act Of God 2016
7	2/7/16	2	2016	Blackbird
17	4/17/16	4	2016	Paramour

4. Filtering in Excel

- Click on *Filter* under the *Data* tab at the ribbon.



- Funnel icons will appear at the top of each column, indicating that Excel's *Filter* feature has been enabled. Click on the funnel icon under the **Statistics.Performances** column and deselect "0".

The screenshot shows the 'Statistics.Performances' filter dialog in Excel. The dialog includes:

- Sort:** Ascending (selected) and Descending buttons.
- Filter:** By color: None, Choose One dropdown (empty), and a Search bar.
- Value List:** A list of values from 0 to 5, each with a checkbox. The value '1' is checked.
- Buttons:** Auto Apply (checked), Apply Filter, and Clear Filter.

- After step 2, you'll see all shows with more than one performance this week.

	A	B	C	D	E	F	G	H	I	J	K	L
	Date.Day	Date.Full	Date.Month	Date.Year	Show.Name	Show.Theatre	Show.Type	Statistics.Attendance	Statistics.Capacity	Statistics.G=	Statistics.G=	Statistics.Performances
2369	6	3/6/16	3	2016	Eclipsed	Golden	Play	6049	94	340891	42	1
2370	24	4/24/16	4	2016	Waitress	Brooks Atkinson	Musical	8419	101	746871	73	1
2371	31	7/31/16	7	2016	Cats 2016	Nell Simon	Musical	9371	98	755353	72	1
2372	19	4/19/15	4	2015	Fun Home	Circle In The Square	Musical	5813	98	335213	45	1
2373	15	3/15/15	3	2015	On The Twentieth Century	American Airlines	Musical	5936	103	390599	50	1
2374	1	11/1/15	11	2015	King Charles III	Music Box	Play	6168	90	405691	47	1
2375	27	9/27/15	9	2015	Spring Awakening 2015	Brooks Atkinson	Musical	7204	88	375985	37	1
2376	8	11/8/15	11	2015	Allegiance	Longacre	Musical	7529	89	420978	44	1
2377	8	3/8/15	3	2015	The Audience	Schoenfeld	Play	7532	101	861884	84	1
2378	15	11/15/15	11	2015	Misery	Broadhurst	Play	8551	90	822637	69	1
2379	12	7/12/15	7	2015	Penn & Teller On Broadway	Marquis	Special	10276	79	1075289	62	1
2380	6	12/6/15	12	2015	School Of Rock	Winter Garden	Musical	10519	88	669028	46	1
2381	12	4/12/15	4	2015	An American In Paris	Palace	Musical	11559	98	826417	76	1
2382	20	12/20/15	12	2015	Fiddler On The Roof 2015	Broadway	Musical	11821	87	837216	54	1
2383	13	4/13/14	4	2014	Lady Day At Emerson's Bar & Grill	Circle In The Square	Play	4639	97	366156	59	1

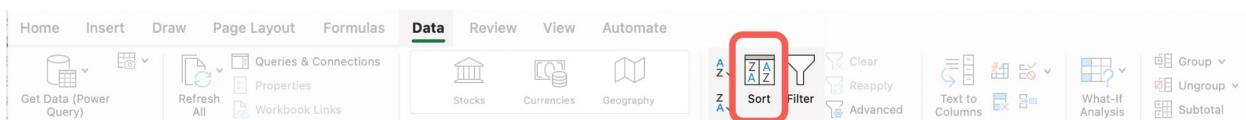


Exercise 3: Using our *Broadway* dataset, find the highest number of performances a show had in a single week.

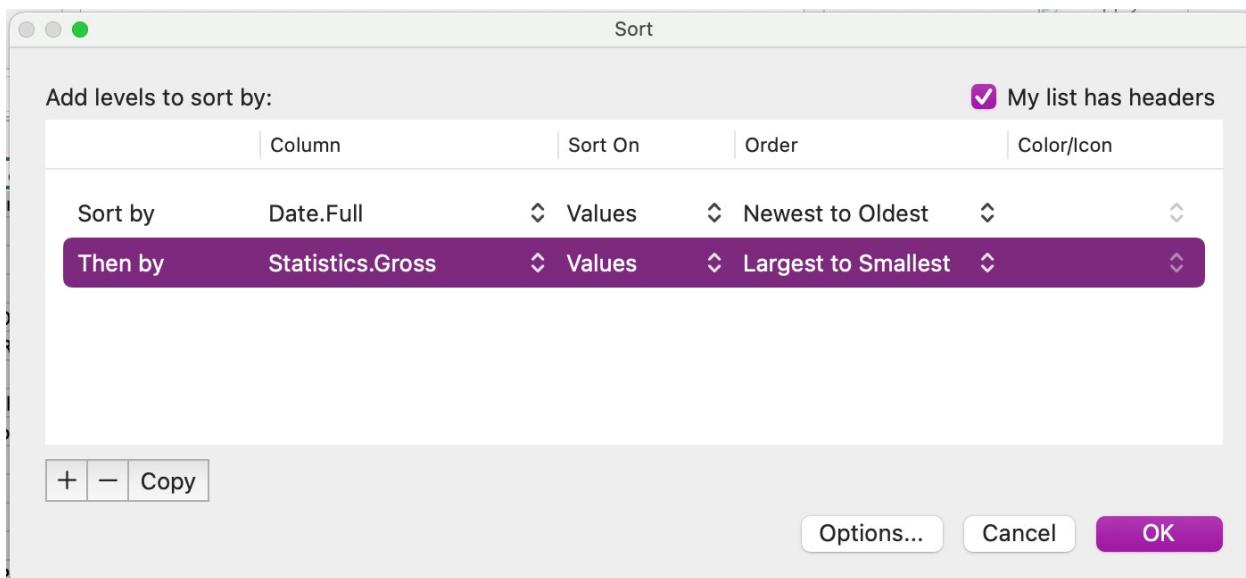
Hint: The Filter interface lists all the unique of a column.

5. Sorting in Excel

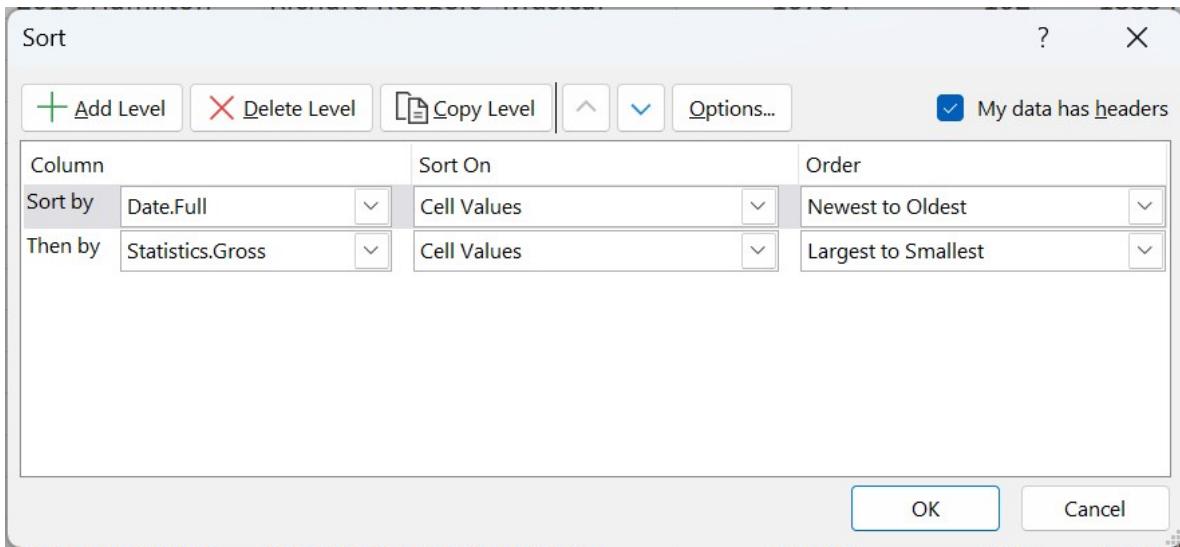
- Click on *Sort* under the *Data* tab at the ribbon.



- Click the plus button (+) to add our desired sorting categories. We want to sort **Date.Full** from “Newest to Oldest” and we want to sort **Statistics.Gross** from “Largest to Smallest”.



MacOS



Windows

- The final result is a spreadsheet that lists shows in chronological order, followed by their ranking by gross revenue.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Date.Day	Date.Full	Date.Month	Date.Year	Show.Name	Show.Theatre	Show.Type	Statistics.Attendance	Statistics.Capacity	Statistics.Gross	Statistics.Gross Potential	Statistics.Performances
2	14	8/14/16	8	2016	The Lion King	Minskoff	Musical	13485	99	2233894	97	8
3	14	8/14/16	8	2016	Hamilton	Richard Rodg	Musical	10756	102	2045095	105	8
4	14	8/14/16	8	2016	Wicked	Gershwin	Musical	13804	95	1779664	100	8
5	14	8/14/16	8	2016	Aladdin	New Amsterd	Musical	13288	96	1663119	90	8
6	14	8/14/16	8	2016	The Book Of f	Eugene O'Nei	Musical	8731	102	1305962	97	8
7	14	8/14/16	8	2016	School Of Roc	Winter Garde	Musical	11342	95	1160694	76	8
8	14	8/14/16	8	2016	Cats 2016	Neil Simon	Musical	10292	94	1108909	78	8
9	14	8/14/16	8	2016	Les Miserable	Imperial	Musical	11244	100	1035397	88	8
10	14	8/14/16	8	2016	The Phantom	Majestic	Musical	10966	85	999632	66	8
11	14	8/14/16	8	2016	Paramour	Lyric	Musical	11360	75	997867	55	8
12	14	8/14/16	8	2016	Waitress	Brooks Atkins	Musical	8058	96	990128	97	8
13	14	8/14/16	8	2016	Matilda	Shubert	Musical	10341	90	876824	73	8
14	14	8/14/16	8	2016	Beautiful	Stephen Sond	Musical	6829	83	728242	72	8
15	14	8/14/16	8	2016	The Color Pur Jacobs		Musical	7367	88	718400	70	8
16	14	8/14/16	8	2016	On Your Feet!	Marquis	Musical	7585	58	703403	45	8
17	14	8/14/16	8	2016	An American	Palace	Musical	9111	68	687762	47	8
18	14	8/14/16	8	2016	Fiddler On Th	Broadway	Musical	8632	63	652283	42	8
19	14	8/14/16	8	2016	Kinky Boots	Al Hirschfeld	Musical	7117	62	651433	50	8
20	14	8/14/16	8	2016	Finding Never Lunt	Fontann	Musical	8775	73	628895	43	8
21	14	8/14/16	8	2016	The Humans	Schoenfeld	Play	7234	87	603770	62	8
22	14	8/14/16	8	2016	Something Rc St.	James	Musical	7239	67	584022	46	8
23	14	8/14/16	8	2016	Chicago	Ambassador	Musical	6929	80	571623	60	8
24	14	8/14/16	8	2016	Jersey Boys	August Wilson	Musical	5991	61	530413	46	8
25	14	8/14/16	8	2016	The Curious I	Ethel Barrym	Play	6407	79	464058	54	8



Exercise 4 What was the highest grossing show on the day of 1/3/16?

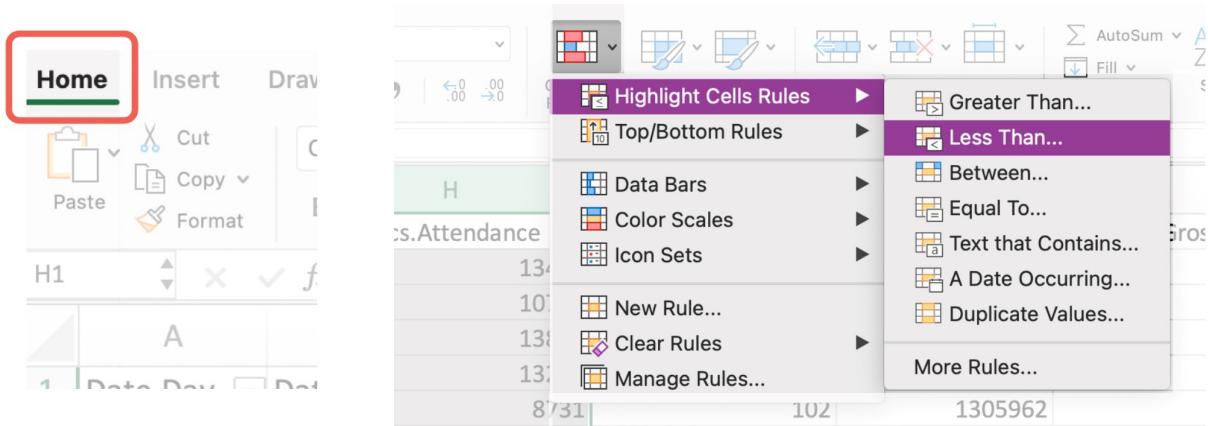
Hint: Now that our spreadsheet is sorted, we can filter by the full date!

6. Conditional Formatting in Excel

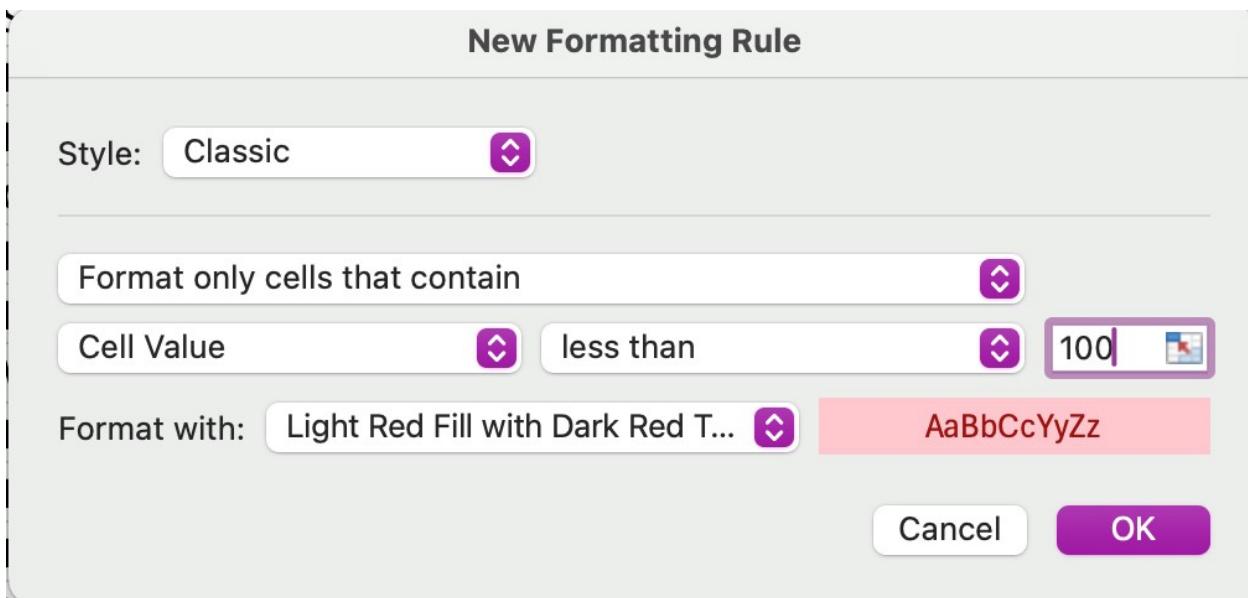
- Let's highlight the showings that had less than 100 attendants. Select column H by clicking on its column header.

E	F	G	H	I	J
Show.Nam	Show.Thea	Show.Type	Statistics.Attendance	Statistics.Capacity	Statistics.Grc
The Lion King	Minskoff	Musical	13485	99	22
...

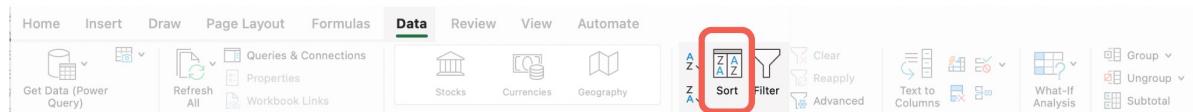
- Select *Home* in the ribbon, and access *Conditional Formatting > Highlight Cells Rules > Less Than...*



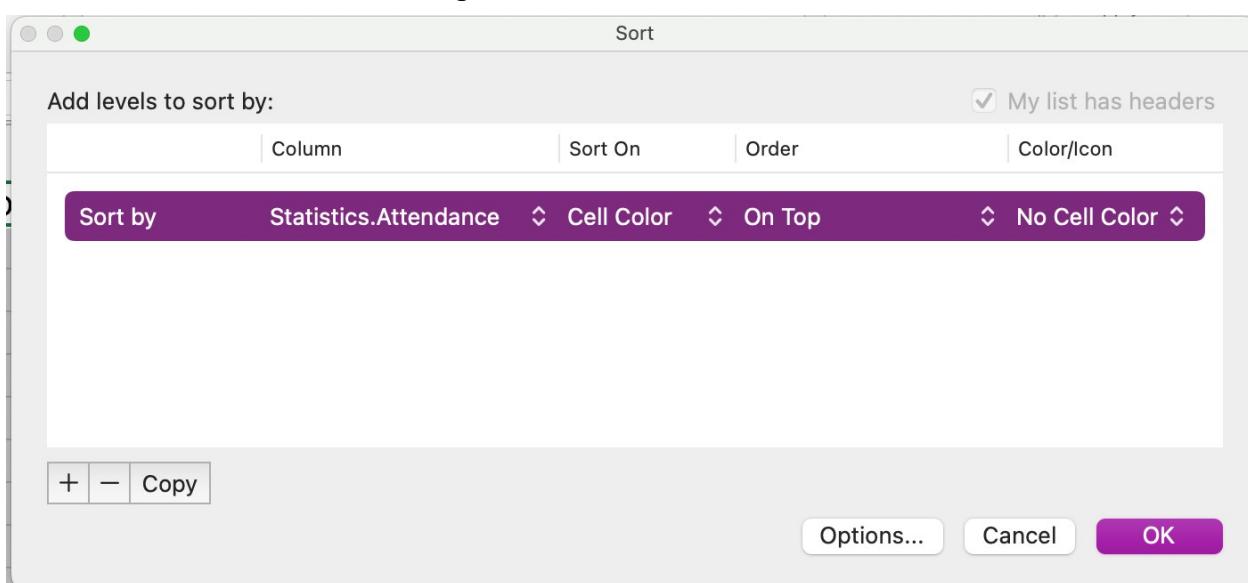
- Let's create a New Formatting Rule that format cells with a value less than 100.



- Click on *Sort* under the *Data* tab at the ribbon.



- We can use Cell Color as a sorting condition under the *Sort* Menu.





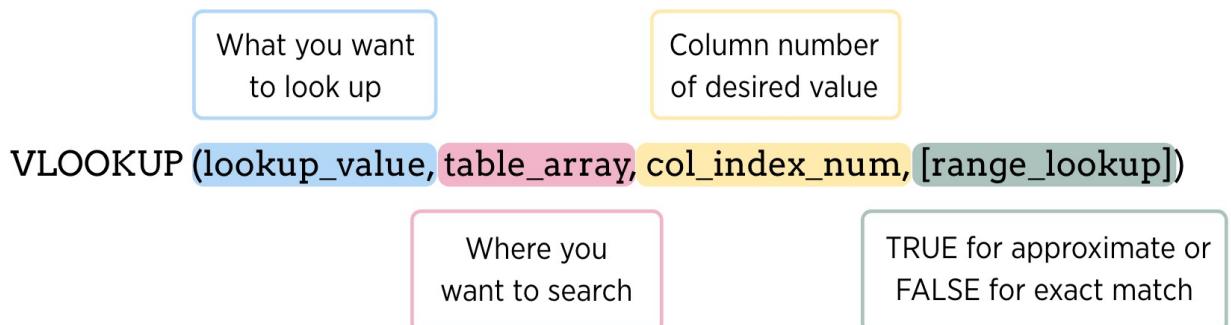
Exercise 5: How many shows had less than 100 attendees?



Exercise 6: Which two shows historically had more than 20,000 attendees?

There are several ways to answer this question with Excel. What is the most visually impactful way to present this information?

7. VLOOKUP Function in Excel



Scenario: You are the theater manager for the Richard Rodgers Theatre, where Hamilton is playing. You need to quickly look up show statistics for specific dates. Let's learn to use the VLOOKUP function to retrieve attendance, capacity, and performance data for the date August 30, 2015.

Statistics for 8/30/15	
Attendance	?
Capacity	?
Performances	?

1. Open the Hamilton Spreadsheet

The VLOOKUP function lets users look up specific values in a table. Let's put this function to use with our *Hamilton* spreadsheet. Navigate to the Hamilton box office data spreadsheet.



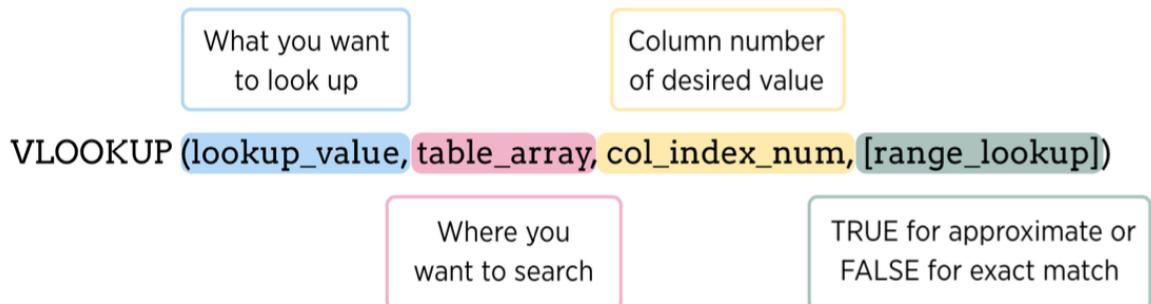
2. Review Column Structure

Your spreadsheet contains these columns:

- Column B: Date.Full (e.g., "8/30/15")
- Column H: Statistics.Attendance
- Column I: Statistics.Capacity
- Column L: Statistics.Performances

3. Practice Exercise

Let's make a VLOOKUP command for **Statistics.Attendance** for the date 8/30/2015.



- **lookup_value:** `DATE(2015,8,30)`

The date we're searching for (8/30/2015)

- **table_array:** `B:H`

The range of data from column B (Date.Full) through column H (Statistics.Attendance)

- **col_index_num:** `7`

Statistics.Attendance is the 7th column counting from B (B=1, C=2, D=3, E=4, F=5, G=6, H=7)

- **[range_lookup]:** `FALSE`

We want an exact match for this specific date



Exercise 7: Find *Hamilton* statistics for the day 8/30/15. Write three VLOOKUP formulas to retrieve:

- Statistics.Attendance
- Statistics.Capacity
- Statistics.Performances

8. Calculations in Excel

Let's create an "at-a-glance" overview of *Broadway* show attendance, capacity, and performances. We can create this mini-table starting in cell **N2**.

N	O	P	Q
Statistics at a Glance			
	Minimum	Maximum	Average
Attendance			
Capacity			
Performances			

Let's start with core statistics for minimum, maximum, and average for *Broadway* attendance, which is stored in Column H.

Minimum

=MIN(H:H)

Maximum

=MAX(H:H)

Average

=AVERAGE(H:H)



Exercise 8: Calculate minimum, maximum and average values for **Statistics.Attendance**, and **Statistics.Capacity**, and **Statistics.Performances** using the whole *Broadway* spreadsheet.

	Minimum	Maximum	Average
Statistics.Attendance	_____	_____	_____
Statistics.Capacity	_____	_____	_____
Statistic.Performances	_____	_____	_____



Exercise 9: Find which show had the highest attendance, and then look up its gross earnings with VLOOKUP.

Hint: You'll need to use a calculation function as well as VLOOKUP.

9. Frequency and Histograms in Excel

The FREQUENCY function calculates how often values occur within a range of values, which is considerably helpful for creating histograms.

FREQUENCY(data_array, bins_array)

Data that you want to categorize into “bins”

Bins that you want to sort your data into

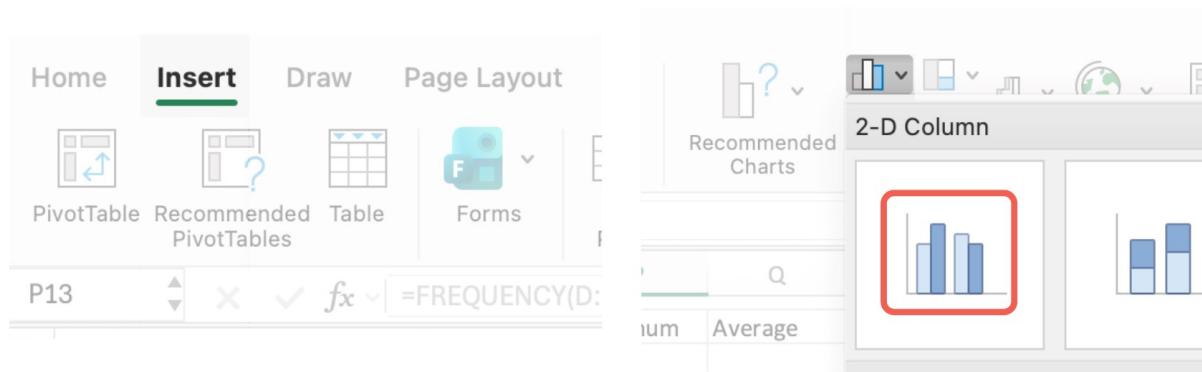
- Let's use the FREQUENCY formula to find out how many Broadway shows took place during the years of 1995-1999, 2000-2004, 2005-2009, 2010-2014, and 2015-2019.

N	O	P	Q
Broadway Show Frequency by Five-Year Period			
Actual Bin	Upper Bin	Frequency	
1995-1999	1999	=FREQUENCY(D:D, O9:O13)	
2000-2004	2004		
2005-2009	2009		
2010-2014	2014		
2015-2019	2019		

- Let's create our histogram figure. Select the “Frequency” column of our overview.

Broadway Show Frequency by Five-Year Period		
Actual Bin	Upper Bin	Frequency
1995-1999	1999	5978
2000-2004	2004	7402
2005-2009	2009	7622
2010-2014	2014	7626
2015-2019	2019	2668

- Under *Insert*, we'll select the first available 2-D Column chart.



- Our chart won't be perfect immediately. We need to select the data that we want to make up the X and Y axis. From the ribbon, click on *Chart Design* > *Select Data*.



- The window pop up allows to select our chart's name, Y values, and horizontal axis values. It's useful to click on the pop-out icon at each text field so we can select a data range.

Select Data Source

Range Details

Chart data range: `=broadway!N8:P13`

Legend entries (Series): `Frequency` Name: `=broadway!P8`

Y values: `=broadway!P9:P13`

+ - Switch Row/Column

Horizontal (Category) axis labels: `=broadway!N9:O13`

Hidden and Empty Cells

Show empty cells as: `Gaps` ↻

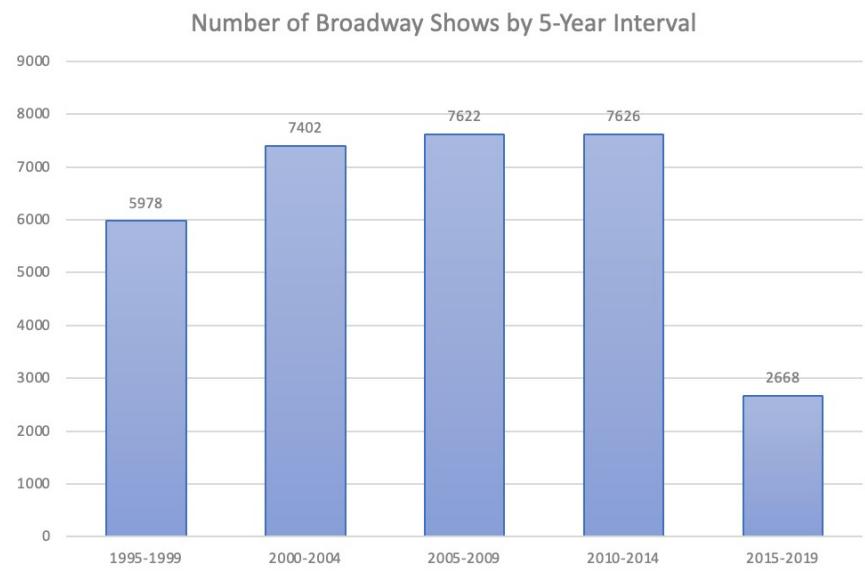
Show #N/A as an empty cell

Show data in hidden rows and columns

Cancel OK

 **Tip:** Click this “pop-out” icon to easily select existing data in a spreadsheet.

Your result should look like this.





Exercise 10: We're creating an account of Broadway shows since 1995. Use the FREQUENCY command to fill out the table below.

Actual Bin	Upper Bin	Frequency
1995 1999	1999	_____
2000 2004	2004	_____
2005 2009	2009	_____
2010 2014	2014	_____
2015 2019	2019	_____



Exercise 11: Our final exercise! Use all the skills we've learned to create a histogram of Broadway show frequency in five-year intervals.

Tip: Choose your own chart design from the *Chart Design* tab.

10. Conclusion

We've made it! Here's a recap of the skills we've learned.

- Learn how to move data around with **Sort**.
- Learn how to highlight and hide data with **Filter and Conditional Formatting**.
- Make a simple search feature with **VLOOKUP (exact match)**.
- Make simple **Calculations** about our data.
- Put data in “bins” for visuals with **FREQUENCY** and bar charts.

Thanks for attending Excel Level 1! I may be reached below for any follow up questions.

Joan Zheng, QCL Graduate Fellow

joanzheng [at] claremontmckenna [dot] edu