

Data Analysis

Using Excel

Responsibilities

Business Analyst	Data Analyst	Data scientist
<ul style="list-style-type: none">• Work with clients to understand problem.• Use current data to outline problems.• Outline and communicate client's requirements to team.	<ul style="list-style-type: none">• Work with programming team to collect and analyze data.• Use pre-existing data to solve a problem.• Create report and dashboards.• Present analytical finding to team.	<ul style="list-style-type: none">• Use current data to discover opportunities.• Develop analytical methods and machine learning models.• Data cleaning (Lots of it).• Conduct A/B testing.

<https://www.youtube.com/watch?v=G4syHs3M82E>

Skills

Business Analyst	Data Analyst	Data scientist
<ul style="list-style-type: none">• SQL• MS access• Excel• Communication skill• Presentation skill• People skill	<ul style="list-style-type: none">• SQL• R/Python(pandas, NumPy, Matplotlib)• Tableau/Power BI• Data Modeling• Excel• AWS/Azure• SAS/SPSS	<ul style="list-style-type: none">• SQL• R/Python(pandas, NumPy, Scikit, Learn, TensorFlow)• Tableau/Power BI• NLP• Apache Spark• Jupyter Notebooks/PyCharm• SAS/SPSS

<https://www.youtube.com/watch?v=fUpChfNN5Uo>

Python Libraries

- Data collection: Pandas , Requests, BeautifulSoup
- Data Exploration: Panda, NumPy, SciPy
- Data Visualization: Matplotlib, Seaborn, Plotly.

Defining data analysis and data analyst

Process of Data Analysis:

- Inspecting data
- Cleaning data
- Transforming data
- Modelling data
- Using data to inform decision making.

Data Analyst

- Can work with any of these processes or all of them. a person whose job is to examine information in order to find something out, or to help for making decision. Usually, they work with spreadsheets and csv files and extract information from them and show it to others. (Chart and graphs)

Learning to identify data

- Not all data is obvious. Learn how to look for data.



- Serial number
- Warranty date
- Number ordered
- Date ordered
- Date shipped
- Date delivered

Data Type

Text or String

Like first name

Date and/or Time

Date of purchase

Number

How many or how much
of something

Boolean

Yes or no, true or
false

Dealing with the data

- Open sampledatawinterathletes.xlsx
- You can see some Wt kg are empty. For solving this, create new column name it withweightornot and put this formula in G4
 - =IF(ISBLANK(E4), "without weight", "with weight")
 - Apply it to all rows by double click on + sign.
- open sampledatafoodsales.xlsx. Click on D column / right click / insert / column header: DisplayNames
 - Apply this formula in D2 cell : =CONCATENATE(C2," ", B2)
 - Apply it to all rows by double click on + sign.

- open sampledatafoodsales.xlsx.
- Add new column with header: ConvertToWeekNumber
- Write = then click on Insert function on Formula Tab.
- Serial_number : A2
- Return_type : 1 (sunday)
- Add new column with header: ConvertToDateOfWeek
- Write this formula:
- $=WEEKDAY(A2,1)$

The screenshot shows an Excel spreadsheet titled "02_03_FunctionsAndFormulas_Start - Excel". The "Formulas" tab is selected. In cell E2, the formula `=WEEKNUM(serial_number,return_type)` is entered. The "Insert Function" dialog box is open, displaying the "WEEKNUM" function under the "Date & Time" category. The dialog also includes search fields and a help link for "Help on this function". The background of the spreadsheet shows a table with columns "TotalSeconds" and "ConvertToWeekNumber", and a list of rows with values corresponding to the serial number 141.

Best practice for starting any project

- Don't guess at anything. Work with team to confirm your finding.
- Define verification process. Ask for peer review. It is important if you are reporting on information on new system.
- Prepare Agenda: what you need to present to your team.
- Take a million notes: where the data come from, who verified data, where the data is in the process.
- Build documentation
- Training and Skills update

Discovering common mistakes of beginners

- Not asking enough questions.
- Record counts. Basic maths
- Ask for data dictionary : Maps data (where data is located and how is related to other data), defines data (name for each field, data types and data type values).
- Do not assume anything on data.

Important:

- Select a row of data: click on first cell in the row / shift + ctrl + right arrow.
- Ctrl + home : back to beginning
- Select a column of data: click on first cell in the column/ shift + ctrl + down arrow.
- Ctrl + a : just select data
- Hide some columns / select the table / Alt ; (just select visible columns)/ Ctrl C / Ctrl V . Without Alt ; you can see all columns.
- Hide some columns / select the table / press F5 / specials/ visible cells only/ ok / ctrl+C / new sheet / paste

- Open sampledatainsurance.xlsx
- Restructure data to state comma space region. For example: NY, East
- Right click Column C / insert
- Write NY, East for the first row / press Ctrl, Enter (for active cell does not move)/ then go to Data tab / flash fill.

Converting lists to tables

- Open sampledatafoodsales.xlsx
- Before converting data to a table make sure the data is together.
- Ctrl A
- Ctrl . Move the active cell around the four corner of the range. Do it few times and see end rows, this ensure there are no empty rows within the data.
- Home tab/ format as table / read the description.
- Insert tab / table / read the description.
- Which one do you think is better?
- Look at different table style and choose any one you

Prefer.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Employee Name	Building	Department	ID#	Phone	Status	Hire Date	Years Benefits	Comp.	Job Rating	2.3%				
Sanders, Troy	Main	Quality Assurance	975-85-7784	(919) 239-0604	Contract	9-Nov-95	25	77,760	3					
Mills, Melissa	N	Create Table	16-54-6374	(919) 272-7944	Half-Time	2-Jul-95	25	R	26,185	5				
Rhodes, Brenda	M		10-84-3236	(919) 624-5634	Full Time	16-Apr-96	25	R	65,880	5				
Bishop, Juan	W	Where is the data for your table?	13-65-6720	(252) 813-8394	Full Time	28-Nov-96	24	M	72,830	2				
Peters, Robert	M	=S\$A\$1:\$L\$762	6-94-4119	(252) 490-7564	Contract	16-Dec-96	24		28,270	5				
Dorsey, Matthew	T	My table has headers	9-41-5552	(919) 422-1208	Full Time	1-Aug-95	25	DM	73,930	1				
Rose, Mark	W		3-23-8477	(919) 862-4601	Full Time	9-Sep-95	25	R	80,090	2				
Anderson, Teason	W	OK	10-43-2924	(919) 280-4104	Full Time	25-Jun-96	25	DMR	24,550	1				
Bishop, Juan	W		11-65-6720	(252) 813-8394	Full Time	2-Jan-97	24	M	72,830	2				
Potter, Dawn	North	Operations	157-25-7652	(919) 326-2077	Contract	8-Mar-96	25		50,200	4				
Hoffman, Brian D	West	Quality Assurance	426-01-4552	(252) 288-1822	Full Time	30-Nov-03	17	D	62,965	1				
Gates, Anne	West	Project & Contract Services	991-22-1095	(919) 463-0903	Full Time	4-Nov-95	25	D	29,760	2				
Holloway, Chris	West	Peptide Chemistry	550-29-1321	(919) 252-0195	Contract	15-Mar-97	24		72,480	2				
Wilcox, Robert	West	Quality Assurance	855-13-5948	(919) 640-8497	Full Time	3-Nov-95	25	R	72,060	2				
Wilkinson, Gregory	Main	Project & Contract Services	278-12-9861	(919) 856-1246	Contract	16-Mar-97	24		39,550	5				
Thornton, Charles	Main	Quality Control	964-25-5290	(919) 744-6192	Full Time	27-Feb-06	15	R	34,990	3				
Noble, Michael	North	Admin Training	542-05-1793	(252) 731-7354	Full Time	15-Nov-95	25	DMR	75,150	1				
Poole, Tracy	North	Logistics	575-64-8597	(919) 886-5267	Contract	21-Mar-96	25		31,970	5				
Merritt, Kevin	North	Project & Contract Services	938-72-3231	(919) 645-6972	Contract	7-Dec-95	25		89,640	4				
Ray, Reannon	Main	Quality Assurance	836-95-3739	(252) 644-3692	Half-Time	15-Apr-97	24	M	20,990	4				
Estrada, Joan	South	Quality Control	426-81-2736	(919) 839-9625	Contract	20-Dec-95	25		35,240	3				

- Click on table / design tab / Filter button. You can check it or not according to your needs.
- When you scroll down the table you can see the header which is really handy.
- You can add new row or new column end of the table.

Explore many options in filter for each column such as text filter (just for text columns), Number filter (just for number columns) Top 10, Clear filter and etc. contact number that second character is b. (?b*) Put it in search)

The screenshot shows a Microsoft Excel spreadsheet titled "CH_01 - Excel". The ribbon at the top has the "Design" tab selected. Below the ribbon, there are several buttons for table management: Summarize with PivotTable, Remove Duplicates, Insert Slicer, and Table Style Options. The "Table Tools" ribbon provides more detailed options for the selected table, including Header Row, Total Row, Banded Rows, First Column, Last Column, Banded Columns, and Filter Button. A preview of different table styles is shown on the right side of the ribbon. The main content area displays a table of employee data. The columns are: Employee Name, Building, Department, ID#, Phone, Status, Hire Date, Years, Benefits, Comp., Job Rating, and a percentage column. The data includes rows for employees like Gates, Anne; Holloway, Chris; Wilcox, Robert; Wilkinson, Gregory; Thornton, Charles; Noble, Michael; Poole, Tracy; Merritt, Kevin; Ray, ReAnn; Estrada, Joan; Hickman, John; Ware, David; Barton, Barry; Rodriguez, Denise; Martinez, Kathleen; George, Jessica; Webb, Jim; McClure, Gary; Baker, Barney; Stephenson, Matthew; Parker, Carl; and Watkins, Gary. Each row contains specific details such as their department (e.g., Project & Contract Services, Peptide Chemistry, Quality Assurance, Admin Training, Logistics, Project & Contract Services, Quality Assurance, Process Development, Quality Assurance, Peptide Chemistry, Engineering/Maintenance, Process Development, Manufacturing, Manufacturing, Manufacturing, Operations, Engineering/Maintenance, Quality Control) and their hire date and years of service.

Activity

- Can you find one or more rules from the data in the table?

Copying and printing filtered lists

- You can apply as many filter as you want then copy the data and paste in somewhere else, without worry about copying filtered data.

- Click on table / design tab / convert to a range.

The screenshot shows a Microsoft Excel spreadsheet with a table titled "Employee Data". The "Design" tab is selected in the ribbon. A context menu is open over the first row of the table, specifically over cell C718, which contains the header "Employee ID". The menu item "Convert to Range" is highlighted. A tooltip for "Convert to Range" is displayed, stating: "Convert this table into a normal range of cells. All of the data is preserved." The table data includes columns for Employee ID, Name, Department, ID#, and Phone number.

Employee ID	Name	Department	ID#	Phone
707	Fox, Ellen	& Contract Services	828-71-5080	(252) 361-3
708	Long, Gary	Mfg	559-37-6297	(919) 488-8
709	Phillips, Liesl	Main	625-53-1462	(252) 755-3
710	McDowell, Scott	Main	867-10-0310	(919) 137-6
711	Franklin, Alicia	South	350-10-4448	(919) 388-3
712	Thomas, Shannon	North	458-73-4969	(919) 635-4
713	Clark, William	Main	317-74-9924	(919) 344-1

Using formulas in tables

- Open sample-xlsx-file-for-testing.xlsx
- Add new column "New sales." At the end of table. add 3000 to every sales in this new column.
- Answer: write = / click on J column / write + 3000 / press enter
- Another Answer: write = J2 + 3000 / press enter

Activity:

Every item gets 2.3% sales price increase. Note: write 2.3% in S1 cell.

- Answer: write = / click on G2 cell / write * / click on S1 cell / press function key F4 for absolute address and S1 stay same for all formulas in this column / write + / click again G2 cell / press enter

sample-xlsx-file-for-testing - Excel

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Product	Discount Band	Units Sold	Manufacturi	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year	Column			
Carretera	None	1618.5	\$ 3.00	\$ 20.00	\$ 32,370.00	\$ -	\$ 32,370.00	\$ 16,185.00	\$ 16,185.00	1/01/2014	1	January	2014	\$ 20.46			
Carretera	None	1321	\$ 3.00	\$ 20.00	\$ 26,420.00	\$ -	\$ 26,420.00	\$ 13,210.00	\$ 13,210.00	1/01/2014	1	January	2014	\$ 20.46			
Carretera	None	2178	\$ 3.00	\$ 15.00	\$ 32,670.00	\$ -	\$ 32,670.00	\$ 21,780.00	\$ 10,890.00	1/06/2014	6	June	2014	\$ 15.35			
Carretera	None	888	\$ 3.00	\$ 15.00	\$ 13,320.00	\$ -	\$ 13,320.00	\$ 8,880.00	\$ 4,440.00	1/06/2014	6	June	2014	\$ 15.35			

Activity:

- How many people in this list (sampledatawinterathletes.xlsx) do not have weight?
 - Answer: click on table / Design tab / change the table name to : Winter_Table
 - Write this in empty cell : =COUNTBLANK(Winter_Table[Wt kg]) / press Enter

- Open sample-xlsx-file-for-testing.xlsx
 - Click on table / Design tab / table style options / check the Total Row.
 - Choose which one is appropriate for your job.

Sorting concepts and sort menu options

- Open sampledatasafety.xlsx
- When you work on new data it is good habit to check the data first:
- Ctrl A
- Ctrl . Move the active cell around the four corner of the range. Do it few times and see end rows, this ensure there are no empty rows within the data.
- Data tab/ Sort / always click on my data has headers. You can see it is sorting according to Department, so let sort by Department

The screenshot shows an Excel spreadsheet titled "CH_02 - Excel". The "Data" tab is selected in the ribbon. A "Sort" dialog box is open, showing the "Employee Name" column is selected for sorting, and the "A to Z" option is chosen. The main table below has the following data:

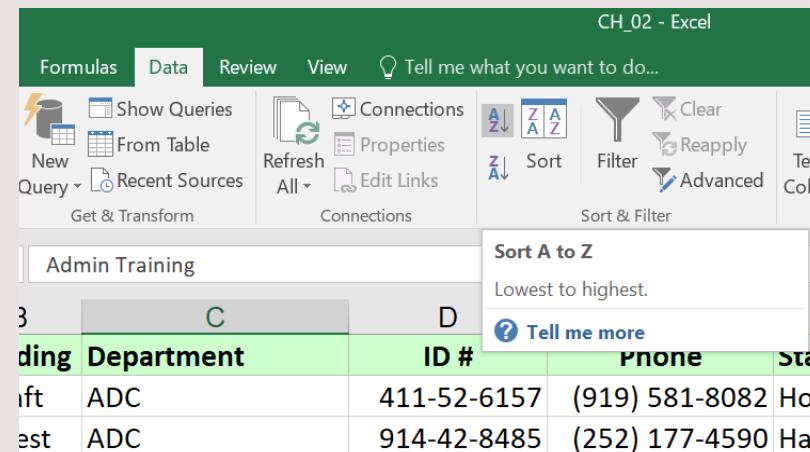
	A	B	C	D	E	F	G	H	I	J	K
1	Employee Name	Building	Department	ID #	Phone	Status	Hire Date	Anniv. Month	Years	Benefits	Job Rating
2	Fleming, Irv	Taft					/1997 August		23		2
3	Flynn, Melissa	West					/1996 February		25	DM	4
4	Owens, Dwight	West					/2012 February		9		5
5	Vasquez, Michael	West					/2004 January		17	DMR	1
6	Woodard, Charles	Watson					/2012 December		8	M	2
7	Ayers, Douglas	Watson					/2005 August		15	D	4
8	Caldwell, Pete	Main					/1996 August		24	DM	1
9	Cameron, John	West					/2013 September		7	D	5
10	Francis, Todd	West					/2005 September		15		3
11	Gallegos, Rick	West					/2014 November		6		4
12	Holt, Robert	Watson					/2011 December		9	M	3
13	Hunter, Lisa	North	Admin Training	542-05-1793	(252) 731-7354	Full Time	3/04/2004 April		17	DMR	1
14	Love, Danny	North	Admin Training	608-79-6012	(919) 407-5460	Full Time	9/08/1997 August		23	DMR	5
15	Lowe, Michelle	Main	Admin Training	202-81-5919	(252) 846-7597	Contract	11/04/1998 April		23		5
16	Martin, Terry	Main	Admin Training	781-91-3936	(919) 788-9149	Half-Time	3/06/2008 June		13	D	3

Sorting from menu icons

- Open sampledatasafety.xlsx
 - Click on one of the cells in report type, go to data tab

then click on sort A...Z. after that click on one of the cells in Department and then click on sort A...Z.

This is the result: The table first sort according to Department and then report type.



Note: Never click on entire column for sorting, always click on one cell in the column.

#	Employee Name	Region	Training Type	Start Date	End Date	Duration	Initial Status	Last Update	Month	Year	Action
12	Holt, Robert	Watson	Admin Training	840-31-3216	(919) 844-9868	Full Time	12/12/2011	December	9	M	3
13	Hunter, Lisa	North	Admin Training	542-05-1793	(252) 731-7354	Full Time	3/04/2004	April	17	DMR	1
14	Love, Danny	North	Admin Training	608-79-6012	(919) 407-5460	Full Time	9/08/1997	August	23	DMR	5
15	Lowe, Michelle	Main	Admin Training	202-81-5919	(252) 846-7597	Contract	11/04/1998	April	23		5
16	Martin, Terry	Main	Admin Training	781-91-3936	(919) 788-9149	Half-Time	3/06/2008	June	13	D	3
17	Massev, Mark	South	Admin Training	475-25-6935	(252) 785-2326	Full Time	2/11/2002	November	18	R	2

Multiple key sorting

CH_02 - Excel

The screenshot shows an Excel spreadsheet titled "CH_02 - Excel". The data is organized into columns A through K, with rows 1 through 20. Column A contains employee IDs (e.g., 1, 2, 3, ...), B contains names, C contains departments, D contains ID numbers, E contains phone numbers, F contains status, G contains hire dates, H contains months, I contains years, J contains benefits, and K contains job ratings. Row 1 serves as the header. A "Sort" dialog box is displayed, indicating a multi-key sort: first by "Department" (A to Z), then by "Status" (A to Z), then by "Years" (Largest to Smallest), and finally by "Employee Name" (A to Z). The "My data has headers" option is selected. The "OK" button is visible at the bottom of the dialog.

	A	B	C	D	E	F	G	H	I	J	K
1	Employee Name	Building	Department	ID #	Phone	Status	Hire Date	Month	Years	Benefits	Job Rating
2	Abbott, James	North	Quality Control	627-49-4412	(252) 824-9735	Full Time	24/08/1997	August	23	DMR	5
3	Acosta, Robert	West	Research & Development	557-56-8959	(919) 278-3818	Contract	15/02/1996	February	25	DM	5
4	Adams, David	North	Manufacturing	920-47-7476	(252) 316-2442	Contract	02/12/2012	February	9	DMR	5
5	Adkins, Michael	Main	Logistics	916-94-4119	(252) 490-7564	Contract	25/08/2004	January	17	R	4
6	Aguilar, Kevin	West	Quality Control	972-08-6665	(252) 600-7063	Full Time	10/12/2012	December	8	DMR	5
7	Alexander, Charles	North	Process Development	676-02-1140	(610) 322-1145	Full Time	20/05/2005	August	15	DMR	5
8	Allen, Thomas	Watson	Research & Development	557-56-8959	(919) 278-3818	Contract	25/08/1996	August	24		4
9	Allison, Timothy	West	Manufacturing	916-94-4119	(252) 490-7564	Contract	22/05/2013	September	7		4
10	Alvarado, Sonia	West	Logistics	972-08-6665	(252) 600-7063	Full Time	10/09/2005	September	15	DMR	4
11	Alvarez, Steven	Watson	Research & Development	676-02-1140	(610) 322-1145	Full Time	20/11/2014	November	6		2
12	Anderson, Teason	North	Manufacturing	916-94-4119	(252) 490-7564	Contract	25/08/2011	December	9	DMR	3
13	Andrews, Diane	West	Logistics	972-08-6665	(252) 600-7063	Full Time	22/05/2004	April	17		2
14	Anthony, Robert	South	Research & Development	676-02-1140	(610) 322-1145	Full Time	10/04/1997	August	23	DMR	5
15	Armstrong, David	North	Quality Control	627-49-4412	(252) 161-4846	Half-Time	11/04/1998	April	23	M	4
16	Arnold, Cole	North	Peptide Chemistry	557-56-8959	(919) 278-3818	Contract	03/06/2008	June	13		4
17	Ashley, Michael	North	Manufacturing	920-47-7476	(252) 316-2442	Contract	2/11/2002	November	18		3
18	Atkins, Kevin	Main	Quality Control	916-94-4119	(252) 490-7564	Contract	25/08/1996	August	24		5
19	Atkinson, Danielle	Main	Process Development	972-08-6665	(252) 600-7063	Full Time	22/05/2007	May	14	R	3
20	Austin, William	South	Engineering/Operations	676-02-1140	(610) 322-1145	Full Time	20/01/2003	January	10	DMR	4

Menu-Sort Icon-Sort MultipleKeySorting CustomLists HR List-Color ColumnSort RandomSort

Exercise Files

Multiple-key sorting...

Data analysis Quiz ...

CH_02 - Excel

4:22 PM 30/06/2021

Sorting based on the order of data in custom lists

- Open sampledatafoodsales.xlsx
- We want sort according to the city. Write San Diego, Boston, Los Angeles, New York in M column. (each city in one row)
- Choose all Column M data / file / options / advanced / Edit custom lists / click on import / ok / ok

The screenshot shows the 'Excel Options' dialog box. The 'Advanced' tab is selected in the left sidebar. In the main area, there is a 'General' section with several checkboxes and settings, and a 'Data' section with checkboxes related to PivotTables and Data Model operations. At the bottom right of the dialog, there is a green button labeled 'Edit Custom Lists...'. To the right of the dialog, a portion of the Excel ribbon is visible, specifically the 'Data Tools' tab under the 'Data' category.

The 'Custom Lists' dialog box is open, showing a list of items categorized by letter (M, N, O). The items listed under 'M' are:

M	N	O
Full Time	ADC	
Half-Time	Admin Training	
Hourly	Audit Services	
Contract	Compliance	
	Engineering/Maintenance	
	Engineering/Operations	
	Environmental Health/Safety	
	Executive Education	
	International Clinical Safety	
	Logistics	
	Major Mfg Projects	
	Manufacturing	
	Manufacturing Admin	
	Operations	
	Peptide Chemistry	
	Pharmacokinetics	
	Process Development	
	Professional Training Group	

- Back to excel sheet / click on a cell on City column / data tab / sort / for order choose custom list / ok / ok.

The screenshot shows a Microsoft Excel spreadsheet titled "sampledatafoodsales(1)(1) - Excel". The "Data" tab is selected in the ribbon. A "Sort" dialog box is displayed, overlaid on the spreadsheet. The dialog has a "Custom Lists" tab selected, showing a list entry "Boston, Los Angeles, New York". There are "Add" and "Delete" buttons next to the list. The "OK" button at the bottom right of the dialog is highlighted with a blue border. The main spreadsheet area shows a table with columns: OrderDate, Region, City, Category, Product. The "City" column contains entries like Boston, Los Angeles, New York, San Diego, etc. The "Sort by" dropdown in the dialog is set to "City".

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	OrderDate	Region	City	Category	Product										
2	1/01/2020	East	Boston	Bars	Carrot										
3	4/01/2020	East	Boston	Crackers	Whole Wheat										
4	13/01/2020	East	Boston	Cookies	Arrowroot										
5	16/01/2020	East	Boston	Bars	Carrot										
6	19/01/2020	East	Boston												
7	31/01/2020	East	Boston												
8	3/02/2020	East	Boston												
9	6/02/2020	East	Boston												
10	18/02/2020	East	Boston												
11	21/02/2020	East	Boston												
12	8/03/2020	East	Boston												
13	11/03/2020	East	Boston												
14	26/03/2020	East	Boston												
15	29/03/2020	East	Boston												
16	16/04/2020	East	Boston												
17	19/04/2020	East	Boston												
18	4/05/2020	East	Boston	Cookies	Oatmeal R										
19	7/05/2020	East	Boston												
20	25/05/2020	East	Boston	Cookies	Arrowroot	27	2.18	90.27							
21	28/05/2020	East	Boston	Bars	Carrot	58	1.77	40.32							

Activity

- Sort the list according to the Category.

Sorting by cell color, font color, or cell icon

- Open sampledatafoodsales.xlsx
- Click on total price column/ home tab / conditional formatting / Icon sets/ choose any directional which you think is appropriate for your data.
- conditional formatting / manage rules / edit rule / try different options.
- Data tab / sort button / try different options sort with cell icon

Sorting columns

- If you have range data (not table) you can sort data left to right, instead of top to bottom.
- Highlight all columns / data tab / sort button / options / choose sort left to right
- Sort by one of the row/ ok

Filtering interactively using slicers with tables

From insert tab / table / insert slicer /

Activity: Can you find one or more rules from the data in the table?

Using Advanced Filter for complex "OR" criteria

- Open sample-xlsx-file-for-testing.xlsx
- Question: Year more than 2013 **OR** unit sold more than 2000.
- Note: you should put two criteria in two separated rows.
- Data tab / sort and filter section/ select advance / select criteria range as below. / OK.

The screenshot shows a Microsoft Excel spreadsheet with data in columns D through R. The first two rows of data are highlighted with a green dashed border. The first row contains criteria: 'Unit Band' (with value '>2000'), 'Year' (with value '>2013'). The second row contains the actual data. A 'Sort & Filter' ribbon tab is selected. An 'Advanced Filter' dialog box is open over the data, with the following settings:

- Action:** Filter the list, in-place
- List range:** \$A\$5:\$P\$705
- Criteria range:** Sheet1!Criteria
- Copy to:** (empty)
- Unique records only:**

At the bottom of the dialog box are 'OK' and 'Cancel' buttons.

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
Unit Band	Units Sold	Manufacturing	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year		
>2000														>2013
2178	\$ 3.00	\$ 15.00	\$ 32,670.00	\$ -	\$ 32,670.00	\$ 21,780.00	\$ 10,890.00	\$ 1/06/2014	6	June	2014			
2470	\$ 3.00	\$ 15.00	\$ 37,050.00	\$ -	\$ 37,050.00	\$ 24,700.00	\$ 12,350.00	\$ 1/06/2014	6	June	2014			
2518	\$ 5.00	\$ 12.00	\$ 30,216.00	\$ -	\$ 30,216.00	\$ 7,554.00	\$ 22,662.00	\$ 1/06/2014	6	June	2014			
2470	\$ 5.00	\$ 15.00	\$ 37,050.00	\$ -	\$ 37,050.00	\$ 24,700.00	\$ 12,350.00	\$ 1/06/2014	6	June	2014			
2665.5	\$ 5.00	\$ 125.00	\$ 333,187.50	\$ -	\$ 333,187.50	#####	\$ 13,327.50	\$ 1/07/2014	7	July	2014			
2146	\$ 5.00	\$ 7.00	\$ 15,022.00	\$ -	\$ 15,022.00	\$ 10,730.00	\$ 4,292.00	\$ 1/09/2014	9	September	2014			
2518	\$ 10.00	\$ 12.00	\$ 30,216.00	\$ -	\$ 30,216.00	\$ 7,554.00	\$ 22,662.00	\$ 1/06/2014	6	June	2014			
2472	\$ 10.00	\$ 15.00	\$ 37,080.00	\$ -	\$ 37,080.00	\$ 12,360.00	\$ 10,720.00	\$ 1/09/2014	9	September	2014			
2152	\$ 10.00	\$ 15.00	\$ 32,280.00	\$ -	\$ 32,280.00	\$ 1,520.00	\$ 10,760.00	\$ 1/12/2013	12	December	2013			
2161	\$ 120.00	\$ 12.00	\$ 25,932.00	\$ -	\$ 25,932.00	\$ 5,483.00	\$ 19,449.00	\$ 1/03/2014	3	March	2014			
2821	\$ 120.00	\$ 125.00	\$ 352,625.00	\$ -	\$ 352,625.00	#####	\$ 14,105.00	\$ 1/08/2014	8	August	2014			
2001	\$ 250.00	\$ 300.00	\$ 600,300.00	\$ -	\$ 600,300.00	#####	\$ 100,050.00	\$ 1/02/2014	2	February	2014			
2838	\$ 250.00	\$ 12.00	\$ 34,056.00	\$ -	\$ 34,056.00	\$ 5,514.00	\$ 25,542.00	\$ 1/04/2014	4	April	2014			
2178	\$ 250.00	\$ 15.00	\$ 32,670.00	\$ -	\$ 32,670.00	\$ 1,780.00	\$ 10,890.00	\$ 1/06/2014	6	June	2014			
2151	\$ 250.00	\$ 300.00	\$ 645,300.00	\$ -	\$ 645,300.00	#####	\$ 107,550.00	\$ 1/09/2014	9	September	2014			
2750	\$ 260.00	\$ 350.00	\$ 962,500.00	\$ -	\$ 962,500.00	#####	\$ 247,500.00	\$ 1/02/2014	2	February	2014			
4219.5	\$ 260.00	\$ 125.00	\$ 527,437.50	\$ -	\$ 527,437.50	#####	\$ 21,097.50	\$ 1/04/2014	4	April	2014			
2141	\$ 260.00	\$ 12.00	\$ 25,692.00	\$ -	\$ 25,692.00	\$ 5,423.00	\$ 19,269.00	\$ 1/08/2014	8	August	2014			
3945	\$ 10.00	\$ 7.00	\$ 27,615.00	\$ -	\$ 27,615.00	\$ 7,725.00	\$ 7,613.85	\$ 1/01/2014	1	January	2014			
2296	\$ 10.00	\$ 15.00	\$ 34,440.00	\$ -	\$ 34,440.00	\$ 2,960.00	\$ 11,135.60	\$ 1/02/2014	2	February	2014			
2529	\$ 3.00	\$ 7.00	\$ 17,703.00	\$ 177.03	\$ 17,525.97	\$ 12,645.00	\$ 4,880.97	\$ 1/07/2014	7	July	2014			
2671	\$ 3.00	\$ 12.00	\$ 32,052.00	\$ 320.52	\$ 31,731.48	\$ 8,013.00	\$ 23,718.48	\$ 1/09/2014	9	September	2014			
2155	\$ 3.00	\$ 350.00	\$ 754,250.00	\$ 7,542.50	\$ 746,707.50	#####	\$ 186,407.50	\$ 1/12/2014	12	December	2014			
2214	\$ 5.00	\$ 15.00	\$ 33,210.00	\$ 332.10	\$ 32,877.90	\$ 22,140.00	\$ 10,737.90	\$ 1/03/2014	3	March	2014			
2301	\$ 5.00	\$ 300.00	\$ 690,300.00	\$ 6,903.00	\$ 683,397.00	#####	\$ 108,147.00	\$ 1/04/2014	4	April	2014			
2102	\$ 5.00	\$ 300.00	\$ 710,400.00	\$ 7,101.00	\$ 701,299.00	#####	\$ 117,106.00	\$ 1/06/2014	6	June	2014			

Activity

- Find all products with Month number is 9 and gross sale more than 30000 or month number is 11 and gross sale more than 80000

Transform | Connections | Sort & Filter | Data Tools | Forecast | Outline |

D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Discount Band	Units Sold	Manufacturing	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year			
				>30000						9					
				>80000						11					
Discount Band	Units Sold	Manufacturing	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year			
None	788	\$ 10.00	\$ 300.00	\$ 236,400.00	\$ -	\$ 236,400.00	\$ 39,400.00	\$ 197,000.00	1/09/2013	9	September	2013			
None	2472	\$ 10.00	\$ 15.00	\$ 37,080.00	\$ -	\$ 37,080.00	\$ 24,720.00	\$ 12,360.00	1/09/2014	9	September	2014			
None	1725	\$ 10.00	\$ 350.00	\$ 603,750.00	\$ -	\$ 603,750.00	\$ 155,250.00	\$ 448,500.00	1/11/2013	11	November	2013			
None	1527	\$ 250.00	\$ 350.00	\$ 534,450.00	\$ -	\$ 534,450.00	\$ 137,430.00	\$ 407,020.00	1/09/2013	9	September	2013			
None	2151	\$ 250.00	\$ 300.00	\$ 645,300.00	\$ -	\$ 645,300.00	\$ 107,550.00	\$ 537,750.00	1/09/2014	9	September	2014			
Low	330	\$ 3.00	\$ 125.00	\$ 41,250.00	\$ 412.50	\$ 40,837.50	\$ 39,600.00	\$ 1,237.50	1/09/2013	9	September	2013			
Low	2671	\$ 3.00	\$ 12.00	\$ 32,052.00	\$ 32,052.00	\$ 32,052.00	\$ 8,013.00	\$ 23,718.48	1/09/2014	9	September	2014			
Low	2498	\$ 5.00	\$ 300.00	\$ 749,400.00	\$ 749,400.00	\$ 749,400.00	\$ 117,406.00	\$ 632,000.00	1/09/2013	9	September	2013			
Low	2905	\$ 10.00	\$ 300.00	\$ 871,500.00	\$ 871,500.00	\$ 871,500.00	\$ 136,535.00	\$ 735,000.00	1/11/2014	11	November	2014			
Low	986	\$ 250.00	\$ 300.00	\$ 295,800.00	\$ 295,800.00	\$ 295,800.00	\$ 46,342.00	\$ 249,458.00	1/09/2014	9	September	2014			
Low	1744	\$ 250.00	\$ 125.00	\$ 218,000.00	\$ 218,000.00	\$ 218,000.00	\$ 6,540.00	\$ 211,460.00	1/11/2014	11	November	2014			
Low	1660	\$ 5.00	\$ 125.00	\$ 207,500.00	\$ 207,500.00	\$ 207,500.00	\$ 4,150.00	\$ 203,350.00	1/11/2013	11	November	2013			
Low	1797	\$ 5.00	\$ 350.00	\$ 628,950.00	\$ 628,950.00	\$ 628,950.00	\$ 142,861.50	\$ 486,088.50	1/09/2013	9	September	2013			
Low	2074	\$ 10.00	\$ 20.00	\$ 41,480.00	\$ 41,480.00	\$ 41,480.00	\$ 7,400.00	\$ 19,080.80	1/09/2014	9	September	2014			
Low	2646	\$ 120.00	\$ 20.00	\$ 52,920.00	\$ 52,920.00	\$ 52,920.00	\$ 5,460.00	\$ 47,460.00	1/09/2013	9	September	2013			
Low	349	\$ 250.00	\$ 350.00	\$ 122,150.00	\$ 122,150.00	\$ 122,150.00	\$ 7,400.00	\$ 26,524.00	1/09/2013	9	September	2013			
Low	1907	\$ 260.00	\$ 350.00	\$ 667,450.00	\$ 667,450.00	\$ 667,450.00	\$ 144,932.00	\$ 523,518.00	1/09/2014	9	September	2014			
Medium	720	\$ 5.00	\$ 350.00	\$ 252,000.00	\$ 252,000.00	\$ 252,000.00	\$ 52,200.00	\$ 200,000.00	1/09/2013	9	September	2013			
Medium	2620	\$ 10.00	\$ 15.00	\$ 39,300.00	\$ 39,300.00	\$ 39,300.00	\$ 5,200.00	\$ 11,135.00	1/09/2014	9	September	2014			
Medium	1679	\$ 260.00	\$ 350.00	\$ 587,650.00	\$ 587,650.00	\$ 587,650.00	\$ 115,851.00	\$ 471,800.00	1/09/2014	9	September	2014			
Medium	1834	\$ 3.00	\$ 20.00	\$ 36,680.00	\$ 2,567.60	\$ 34,112.40	\$ 18,340.00	\$ 15,772.40	1/09/2013	9	September	2013			
Medium	2500	\$ 5.00	\$ 125.00	\$ 312,500.00	\$ 21,875.00	\$ 290,625.00	\$ 11,567.50	\$ (9,375.00)	1/11/2013	11	November	2013			
Medium	2931	\$ 10.00	\$ 15.00	\$ 43,965.00	\$ 3,077.55	\$ 40,887.45	\$ 29,310.00	\$ 11,577.45	1/09/2013	9	September	2013			
Medium	1535	\$ 10.00	\$ 20.00	\$ 30,700.00	\$ 2,149.00	\$ 28,551.00	\$ 15,350.00	\$ 13,201.00	1/09/2014	9	September	2014			
Medium	1123	\$ 10.00	\$ 300.00	\$ 336,900.00	\$ 23,583.00	\$ 313,317.00	\$ 32,567.00	\$ 282,750.00	1/09/2013	9	September	2013			
Medium	1404	\$ 10.00	\$ 200.00	\$ 421,200.00	\$ 20,404.00	\$ 401,796.00	\$ 40,716.00	\$ 361,080.00	1/11/2013	11	November	2013			

OK Cancel

Unique records only

List range: \$A\$5:\$P\$705

Criteria range: Sheet1!Criteria

Copy to:

Unique records only

Action

Filter the list, in-place

Copy to another location

Activity

- Show me three more queries.

Using the Remove Duplicates command

- Go to Data tab / Data tools section / remove duplicates

Using an array formula to count the number of unique items in a list

- Open sampledatainsurance.xlsx
- Insert new row at top by right click / insert . Add this formula at the top of Region `=sum(1/COUNTIF(E3:E502,E3:E502))` / ctrl+shift+enter/ apply formula to all cells / so we have 4 unique region, 13 unique business type, 4 construction and etc.
- Note: Countif (counts the number of entries in a location)

Using SUMIF, COUNTIF, and related functions for quick data analysis

- Open sampledatainsurance.xlsx
- Question: How many frame construction we have?

The screenshot shows a Microsoft Excel spreadsheet titled "sampledatainsurance(1) - Excel". The ribbon menu is visible at the top, with the "Data" tab selected. In the formula bar, the formula `=COUNTIF(H:H,N7)` is entered. The main table contains data with columns: Policy, Expiry, Location, State, Region, InsuredValue, Construction, BusinessType, Earthquake, Flood, Column, and others. A red box highlights the word "Frame" in the "Construction" column of the 13th row. A green box highlights the formula `=COUNTIF` in the formula bar.

Policy	Expiry	Location	State	Region	InsuredValue	Construction	BusinessType	Earthquake	Flood	Column			
00242	2-Jan-21	Urban	NY	East	1,617,630	Frame	Retail	N	N	unique			
00314	2-Jan-21	Urban	NY	East	8,678,500	Fire Resist	Apartment	Y	Y	unique			
00359	2-Jan-21	Rural	WI	Midwest	2,052,660	Frame	Farming	N	N	unique			
00315	3-Jan-21	Urban	NY	East	17,580,000	Frame	Apartment	Y	Y	unique			
00385	3-Jan-21	Urban	NY	East	1,925,000	Masonry	Hospitality	N	N	unique			
00388	4-Jan-21	Urban	IL	Midwest	12,934,500	Frame	Apartment	Y	Y	unique			
00358	5-Jan-21	Urban	WI	Midwest	928,300	Masonry	Office Bldg	N	N	unique			
00264	7-Jan-21	Rural	NY	East	2,219,900	Frame	Farming	N	N	unique			
00265	7-Jan-21	Urban	NY	East	14,100,000	Frame	Apartment	Y	Y	unique			
00357	8-Jan-21	Urban	NY	East	4,762,808	Masonry	Other	Y	Y	unique			
00399	8-Jan-21	Urban	NY	East	13,925,190	Frame	Apartment	Y	Y	unique			
00329	9-Jan-21	Urban	NY	East	6,350,000	Frame	Apartment	Y	Y	unique			
00429	9-Jan-21	Urban	WI	Midwest	4,036,000	Masonry	Medical	Y	Y	unique			
00441	10-Jan-21	Urban	NJ	East	472,800	Masonry	Retail	Y	Y	unique			
00442	10-Jan-21	Urban	WI	Midwest	11,710,880	Masonry	Apartment	Y	Y	unique			
00372	12-Jan-21	Urban	NY	East	1,370,300	Frame	Apartment	Y	Y	unique			

- How much is insured value for retail type?

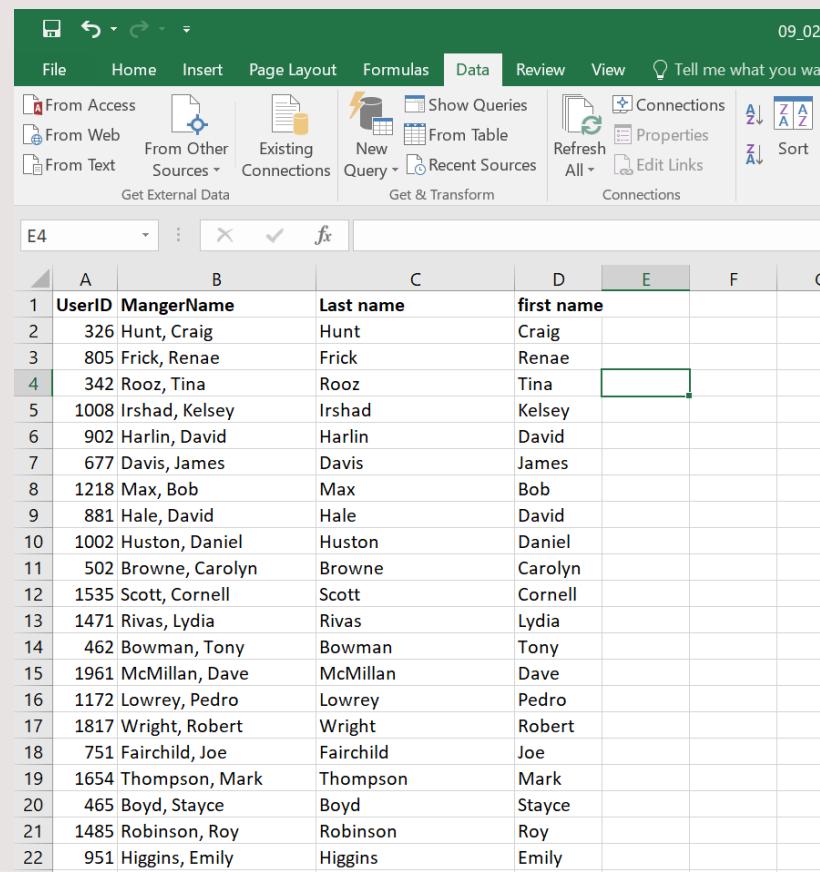
The screenshot shows the Microsoft Power BI Data Editor interface. The top navigation bar includes 'From Other', 'Existing Sources', 'Query', 'Recent Sources', 'Refresh', 'Edit Links', 'Advanced', 'Columns', 'Data Validation', 'Manage Data Model', 'Analysis', 'Sheet', 'Forecast', and 'Outline'. The main area displays a table with the following data:

A	B	C	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
	Expiry	Location	State	Region	InsuredValue	Construction	BusinessType	Earthqua	Flood	Column						
	2-Jan-21	Urban	NY	East	1,617,630	Frame	Retail	N	N	unique						
	2-Jan-21	Urban	NY	East	8,678,500	Fire Resist	Apartment	Y	Y	unique						
	2-Jan-21	Rural	WI	Midwest	2,052,660	Frame	Farming	N	N	unique						
	3-Jan-21	Urban	NY	East	17,580,000	Frame	Apartment	Y	Y	unique						
	3-Jan-21	Urban	NY	East	1,925,000	Masonry	Hospitality	N	N	unique						
	4-Jan-21	Urban	IL	Midwest	12,934,500	Frame	Apartment	Y	Y	unique				Frame		
	5-Jan-21	Urban	WI	Midwest	928,300	Masonry	Office Bldg	N	N	unique						
	7-Jan-21	Rural	NY	East	2,219,900	Frame	Farming	N	N	unique				322		
	7-Jan-21	Urban	NY	East	14,100,000	Frame	Apartment	Y	Y	unique						

- Activity: how many policies have more than 2000000 insured value
- Activity: sum of the insured value for NY.
- Activity: create 3 more questions and see the results.

Learning to split text with delimiters

- Put first names and last names in different columns.
- Right click on column / copy / paste to the next column / data tab/ text to columns / next / choose comma and space / next / finish



	A	B	C	D	E	F	G
1	UserID	MangerName	Last name	first name			
2	326	Hunt, Craig	Hunt	Craig			
3	805	Frick, Renae	Frick	Renae			
4	342	Rooz, Tina	Rooz	Tina			
5	1008	Irshad, Kelsey	Irshad	Kelsey			
6	902	Harlin, David	Harlin	David			
7	677	Davis, James	Davis	James			
8	1218	Max, Bob	Max	Bob			
9	881	Hale, David	Hale	David			
10	1002	Huston, Daniel	Huston	Daniel			
11	502	Browne, Carolyn	Browne	Carolyn			
12	1535	Scott, Cornell	Scott	Cornell			
13	1471	Rivas, Lydia	Rivas	Lydia			
14	462	Bowman, Tony	Bowman	Tony			
15	1961	McMillan, Dave	McMillan	Dave			
16	1172	Lowrey, Pedro	Lowrey	Pedro			
17	1817	Wright, Robert	Wright	Robert			
18	751	Fairchild, Joe	Fairchild	Joe			
19	1654	Thompson, Mark	Thompson	Mark			
20	465	Boyd, Stayce	Boyd	Stayce			
21	1485	Robinson, Roy	Robinson	Roy			
22	951	Higgins, Emily	Higgins	Emily			

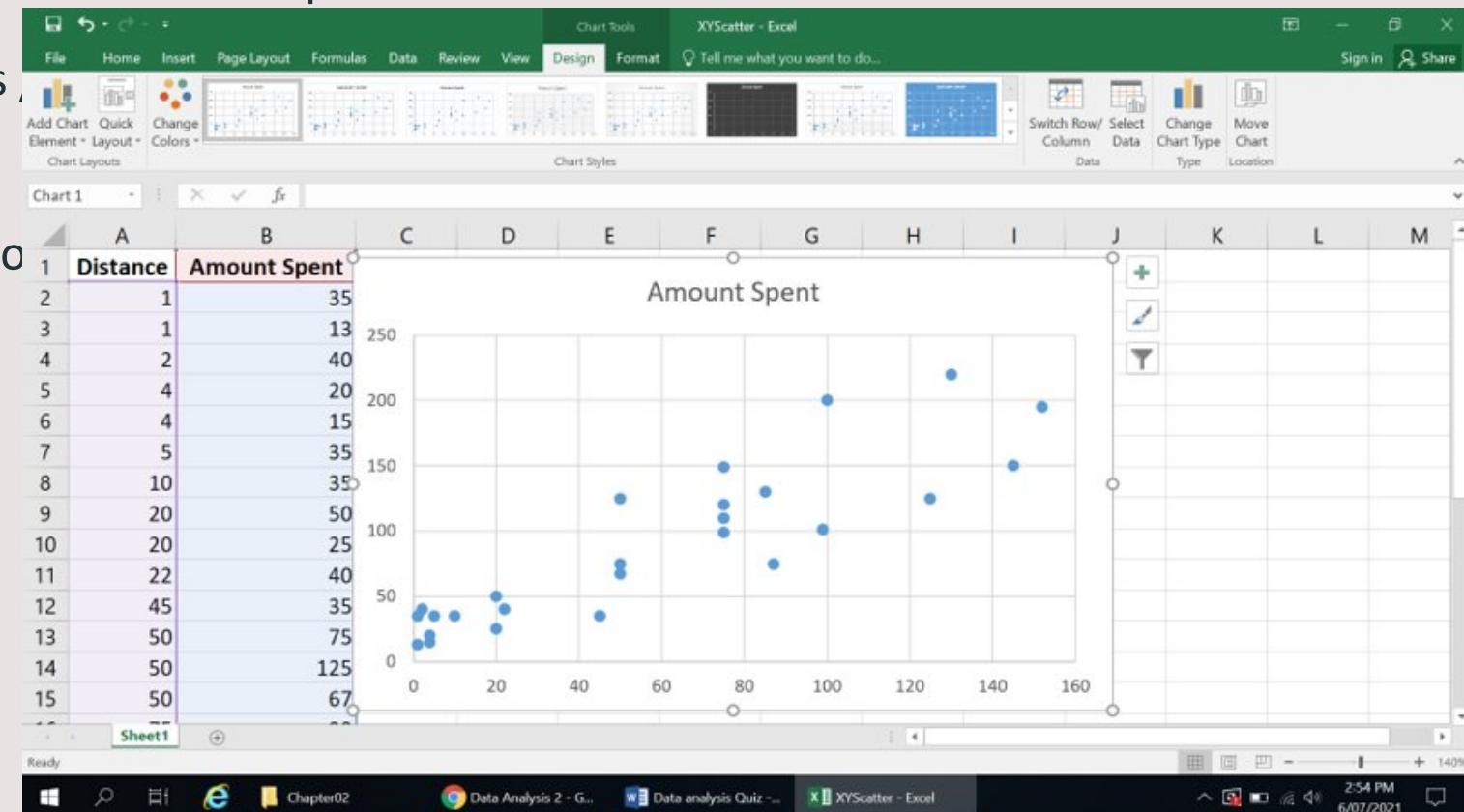
Calculate mean and median values

- CheckMean = average value
- Median = the value that would appear in the middle of the list if it were sorted in order.
- Mode = the most common value. (for example, many insured value with 847300)
- Open sampledatainsurance.xlsx
- If you are suspect, there are several values that accrue most frequently. To detect those values you can use MODE.MULT. Select 3 cells (M3 to M5) / type =MODE.MULT(F2:F501) / Press Ctrl Shift Enter (because it is array formula)

Mean	=AVERAGE(F2:F501)	
Median	=MEDIAN(F2:F501)	
Mode	=MODE.SNGL(F2:F501)	

Identify relationships using XY scatter charts

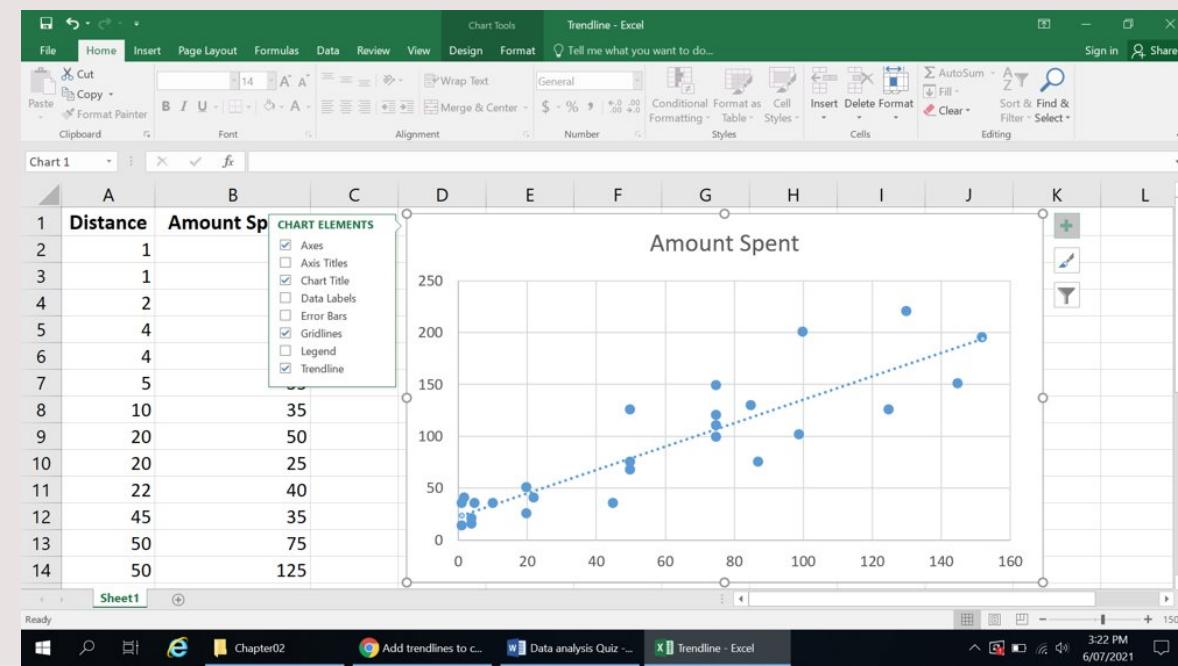
- Analysis to Discover relationship between values.
- Click on any cells to select any chart you want.
- Note: You have to



select any

Add trendlines to charts

- How to examine your data visually by adding Trendline into a chart.
- Click on chart / click on + button / select Trendline on chart elements
- You can use Trendline to estimate how much a particular customer would spend based on a trip of a given length. So, if customer drove 120 miles, you can see that it would cross at approximately 160.



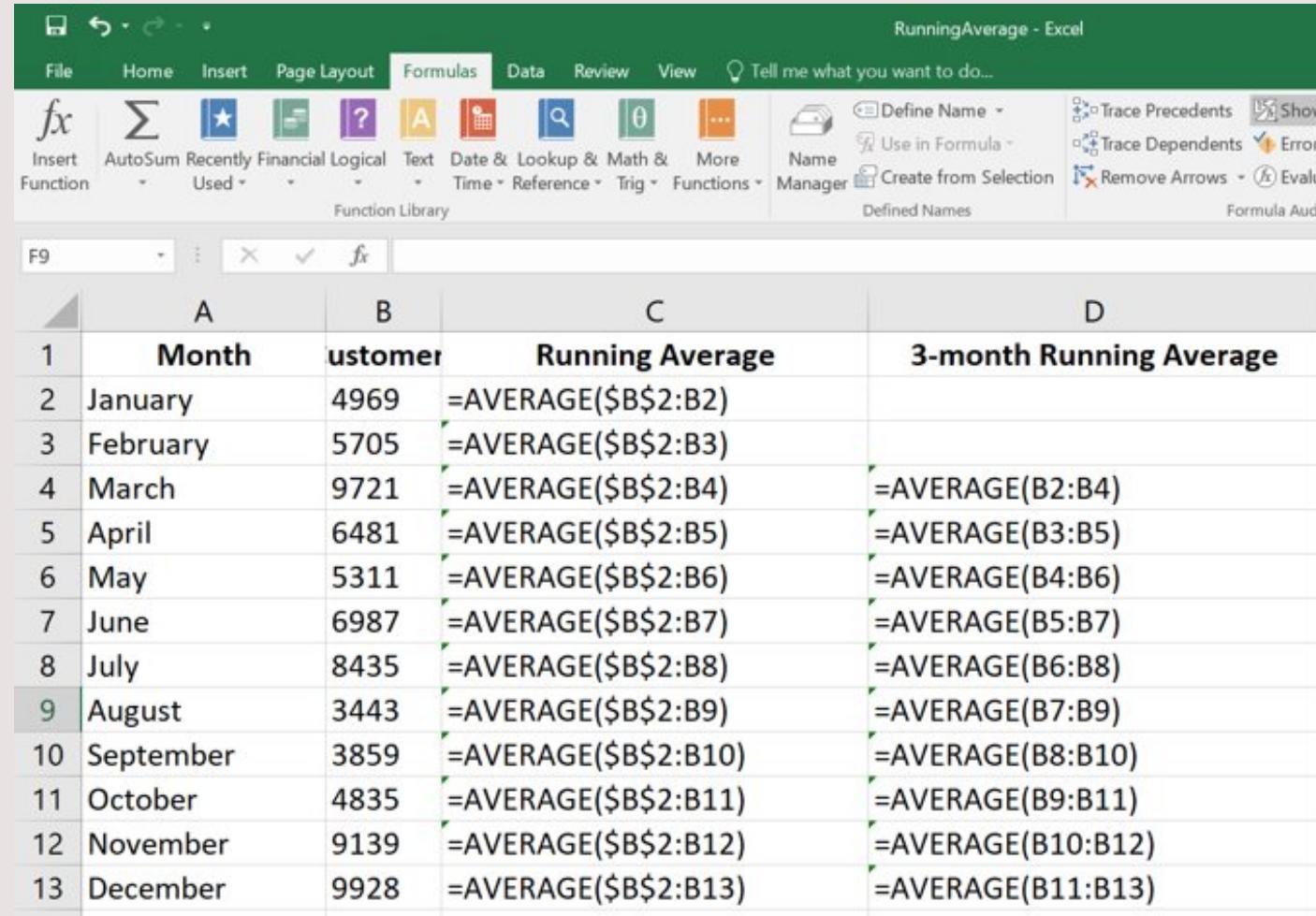
Forecast future results

- Lets predict how much a customer will spend if they drive 30 miles to my store.
- Predict inside data range is more accurate in this situation. Data range here is from 1 mile to 152 miles.

The screenshot shows a Microsoft Excel spreadsheet titled "Forecast - Excel". The ribbon menu is visible at the top, with "Home" selected. The formula bar displays the formula =FORECAST.LINEAR(D2, B2:B27, A2:A27). The spreadsheet contains two data tables:

	A	B	C	D	E	F	G
1	Distance	Amount Spent		Miles Driven	Predicted Amount Spent		
2	1	\$ 35.00		30	=FORECAST.LINEAR(D2, B2:B27, A2:A27)		
3	1	\$ 13.00					
4	2	\$ 40.00					
5	4	\$ 20.00					
6	4	\$ 15.00					

Activity: Calculate running averages



The screenshot shows a Microsoft Excel spreadsheet titled "RunningAverage - Excel". The ribbon menu is visible at the top, with the "Formulas" tab selected. The main area displays a table with four columns: Month, Customer, Running Average, and 3-month Running Average. The "Running Average" column contains formulas starting with =AVERAGE(\$B\$2:B2) for January and extending down to =AVERAGE(\$B\$2:B13) for December. The "3-month Running Average" column contains formulas starting with =AVERAGE(B2:B4) for March and extending down to =AVERAGE(B11:B13) for December.

	A	B	C	D
1	Month	Customer	Running Average	3-month Running Average
2	January	4969	=AVERAGE(\$B\$2:B2)	
3	February	5705	=AVERAGE(\$B\$2:B3)	
4	March	9721	=AVERAGE(\$B\$2:B4)	=AVERAGE(B2:B4)
5	April	6481	=AVERAGE(\$B\$2:B5)	=AVERAGE(B3:B5)
6	May	5311	=AVERAGE(\$B\$2:B6)	=AVERAGE(B4:B6)
7	June	6987	=AVERAGE(\$B\$2:B7)	=AVERAGE(B5:B7)
8	July	8435	=AVERAGE(\$B\$2:B8)	=AVERAGE(B6:B8)
9	August	3443	=AVERAGE(\$B\$2:B9)	=AVERAGE(B7:B9)
10	September	3859	=AVERAGE(\$B\$2:B10)	=AVERAGE(B8:B10)
11	October	4835	=AVERAGE(\$B\$2:B11)	=AVERAGE(B9:B11)
12	November	9139	=AVERAGE(\$B\$2:B12)	=AVERAGE(B10:B12)
13	December	9928	=AVERAGE(\$B\$2:B13)	=AVERAGE(B11:B13)

Analyze Data Button in Excel 365.

- Open your data in Excel 365.
- Simply select a cell in a data range > select the **Analyze Data** button on the **Home** tab. Analyze Data in Excel will analyze your data, and return interesting visuals about it in a task pane.
- If you're interested in more specific information, you can enter a question in the query box at the top of the pane, and press **Enter**. Analyze Data will provide answers with visuals such as tables, charts or PivotTables that can then be inserted into the workbook.
- If you are interested in exploring your data, or just want to know what is possible, Analyze Data also provides personalized suggested questions which you can access by selecting on the query box.
- **Note:** The **Not a value** option in the field list refers to fields that are not normally summed or averaged. For example, you wouldn't sum the years displayed, but you might sum the values of the years displayed. If used with another field that is summed or averaged, **Not a value** works like a row label, but if used by itself, **Not a value** counts unique values of the selected field.

- Analyze Data works best with data that's formatted as an Excel table. To create an Excel table, click anywhere in your data and then press **Ctrl+T**.
- Make sure you have good headers for the columns. Headers should be a single row of unique, non-blank labels for each column. Avoid double rows of headers, merged cells, etc.
- **Activity:** choose two Excel file and work on analyse button.

Pivot Table

- Open sampledatafoodsales.xlsx file
- Select any cell in the source data table.
- On the Ribbon, click the Insert tab.
- In the Tables group, click Recommended PivotTables
- In the Recommended PivotTables window, scroll down the list, to see the suggested layouts.
- Click on the layout that you want to use, then click OK.