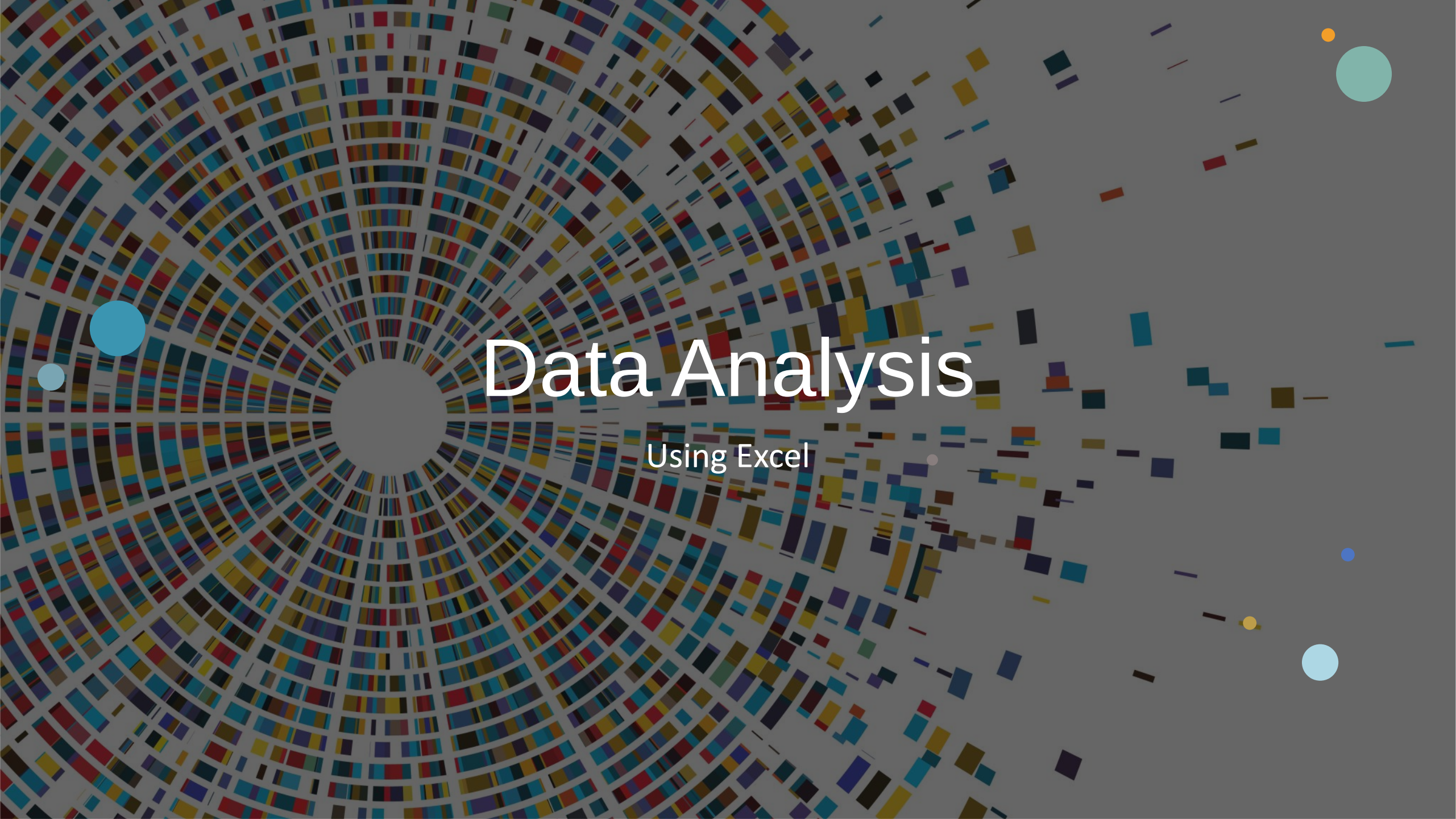

Data Analysis
Using Excel

Responsibilities

Business Analyst	Data Analyst	Data scientist
• Work with clients to understand problem. • Use current data to outline problems. • Outline and communicate client's requirements to team.	• 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.	• 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
• SQL • MS access • Excel • Communication skill • Presentation skill • People skill	• SQL • R/Python(pandas, NumPy, Matplotlib) • Tableau/Power BI • Data Modeling • Excel • AWS/Azure • SAS/SPSS	• 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

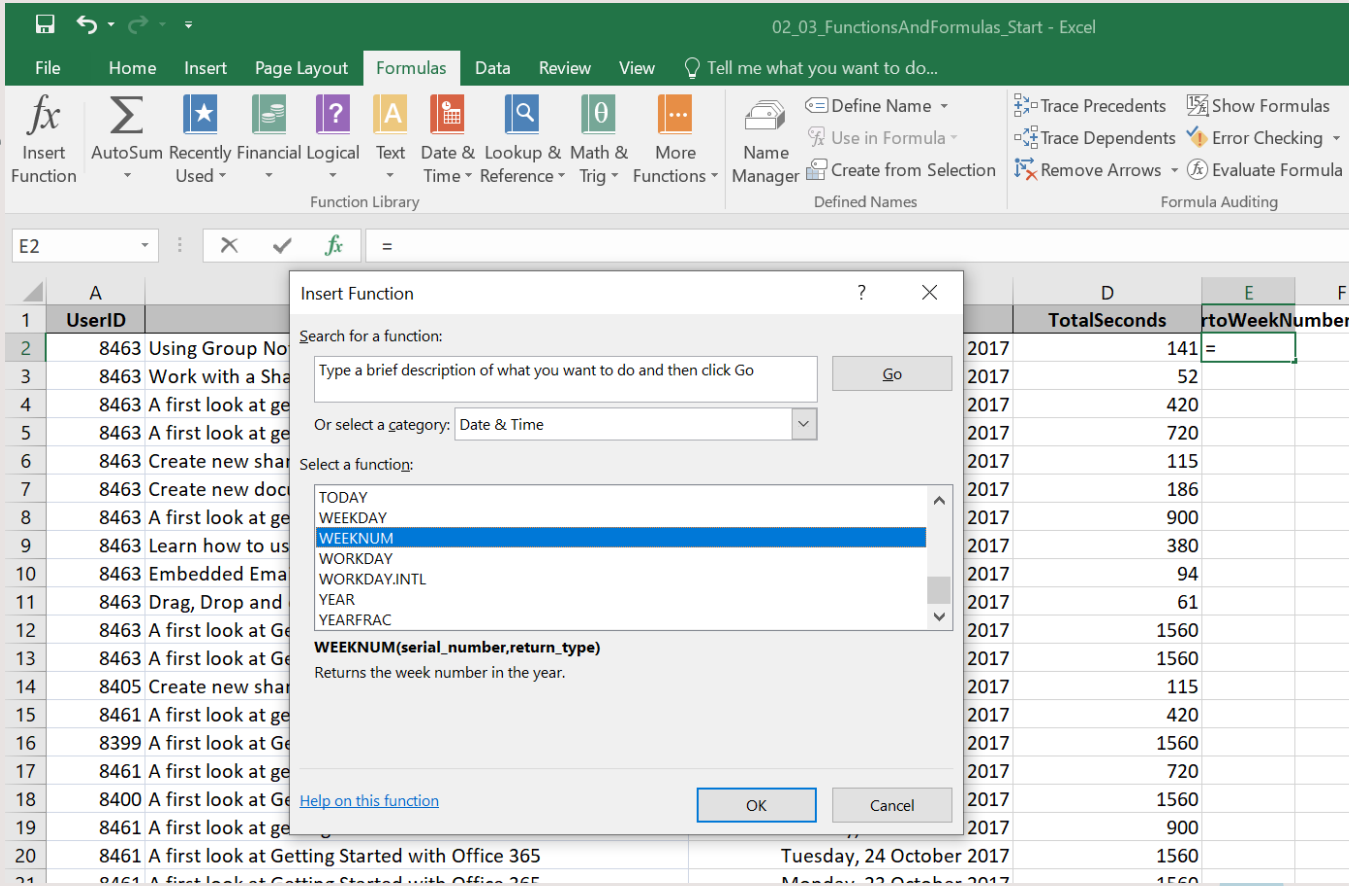
Dealing with the data

- 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: ConvertoWeekNumbe
- Write = then click on Insert function on Formula Tab.
- Serial_number : A2
- Return_type : 1 (sunday)

- Add new column with header: ConvertoDayOfWeek
- Write this formula:
- =WEEKDAY(A2,1)



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 relaesed to other data), defines data (name for each field, data types and data type values).
- Do not assume anything on data.

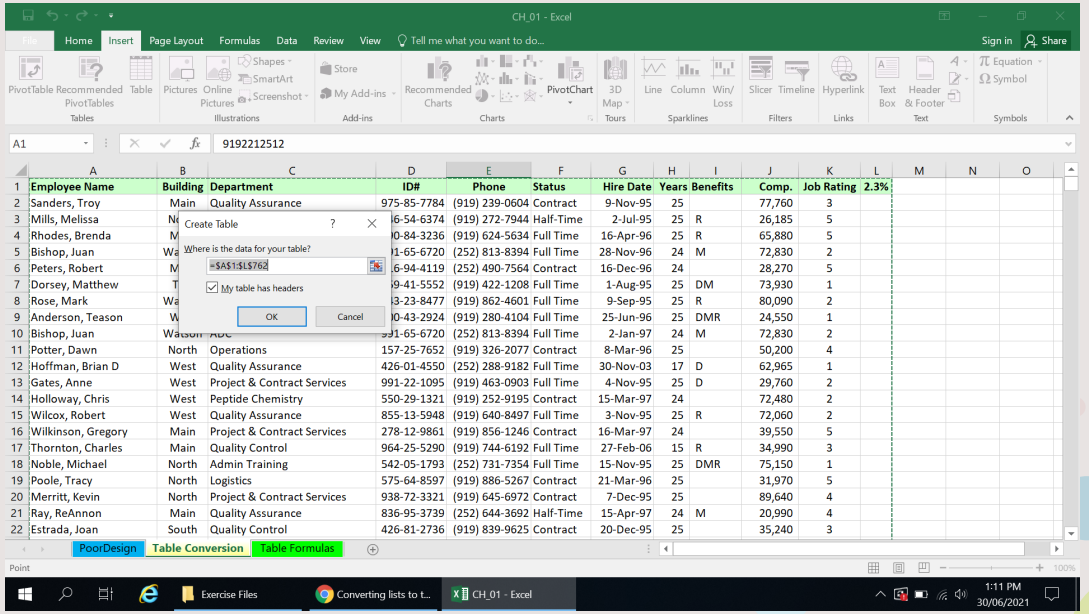
Important:

- 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 Cloumn 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.

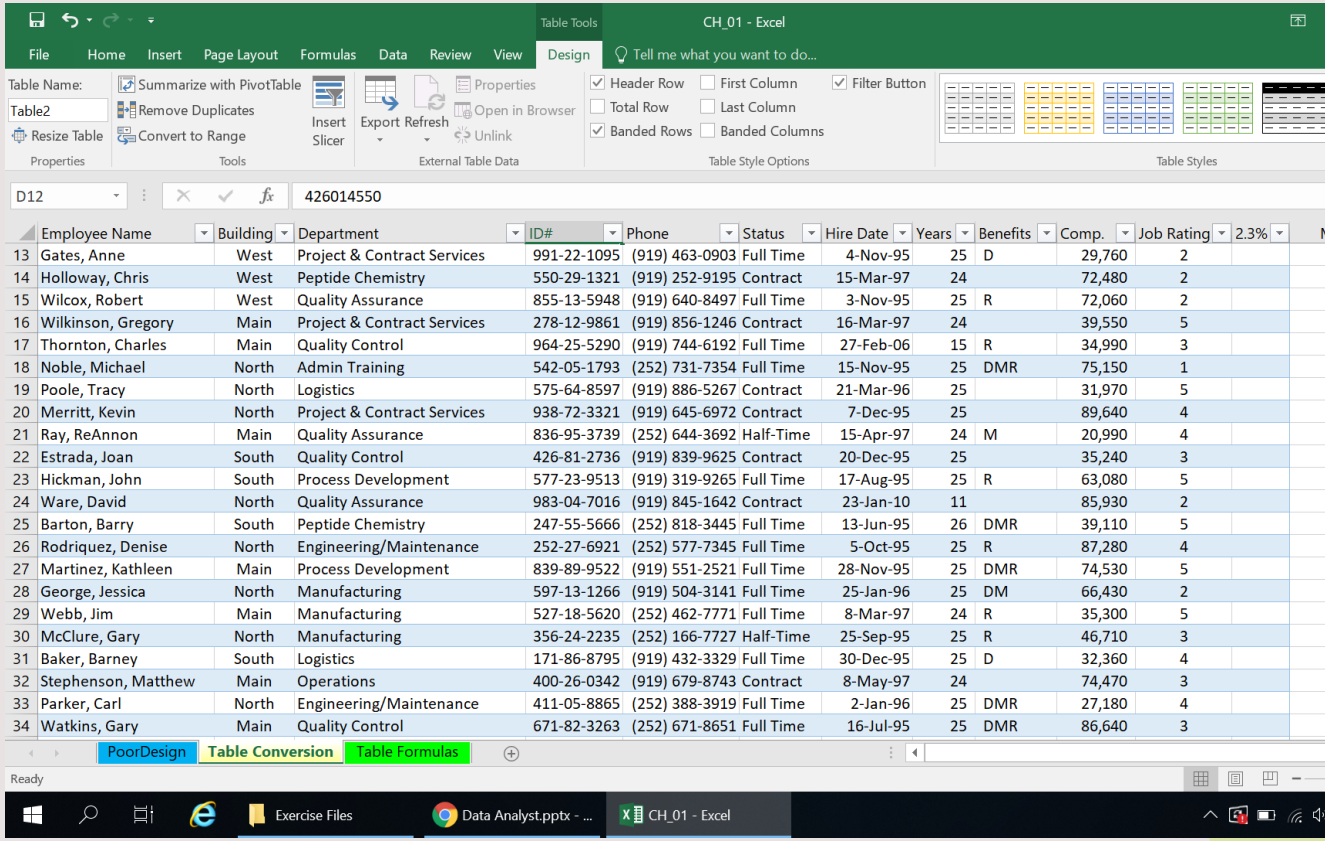


- Click on table / design tab / Filter button. You can check it or not according to your needs.
- When you scroll down the table you can see the header which is really handy.
- You can add new row or

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)



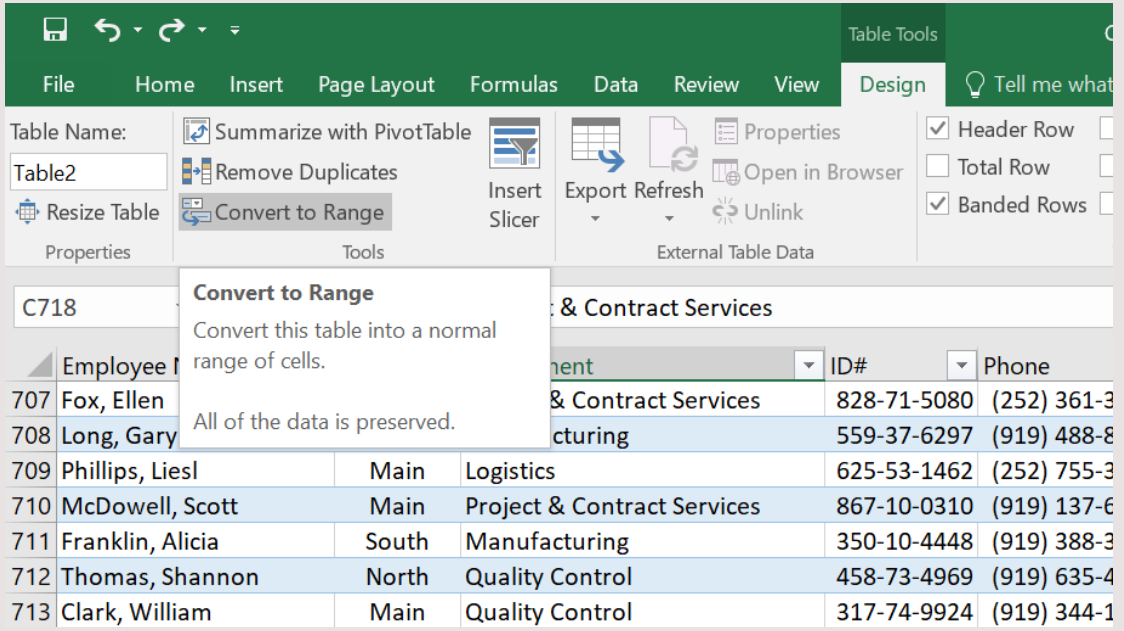
Activity

- 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.



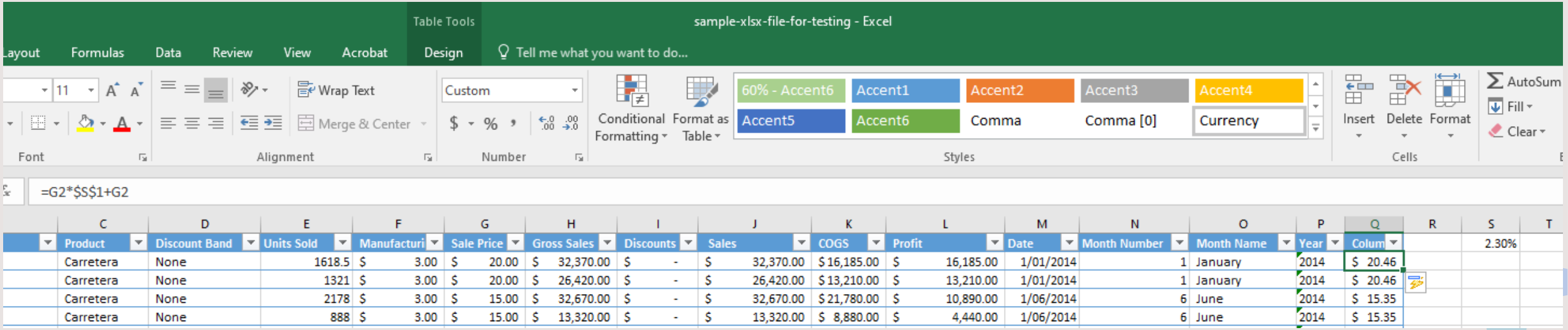
Using formulas in tables

- 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

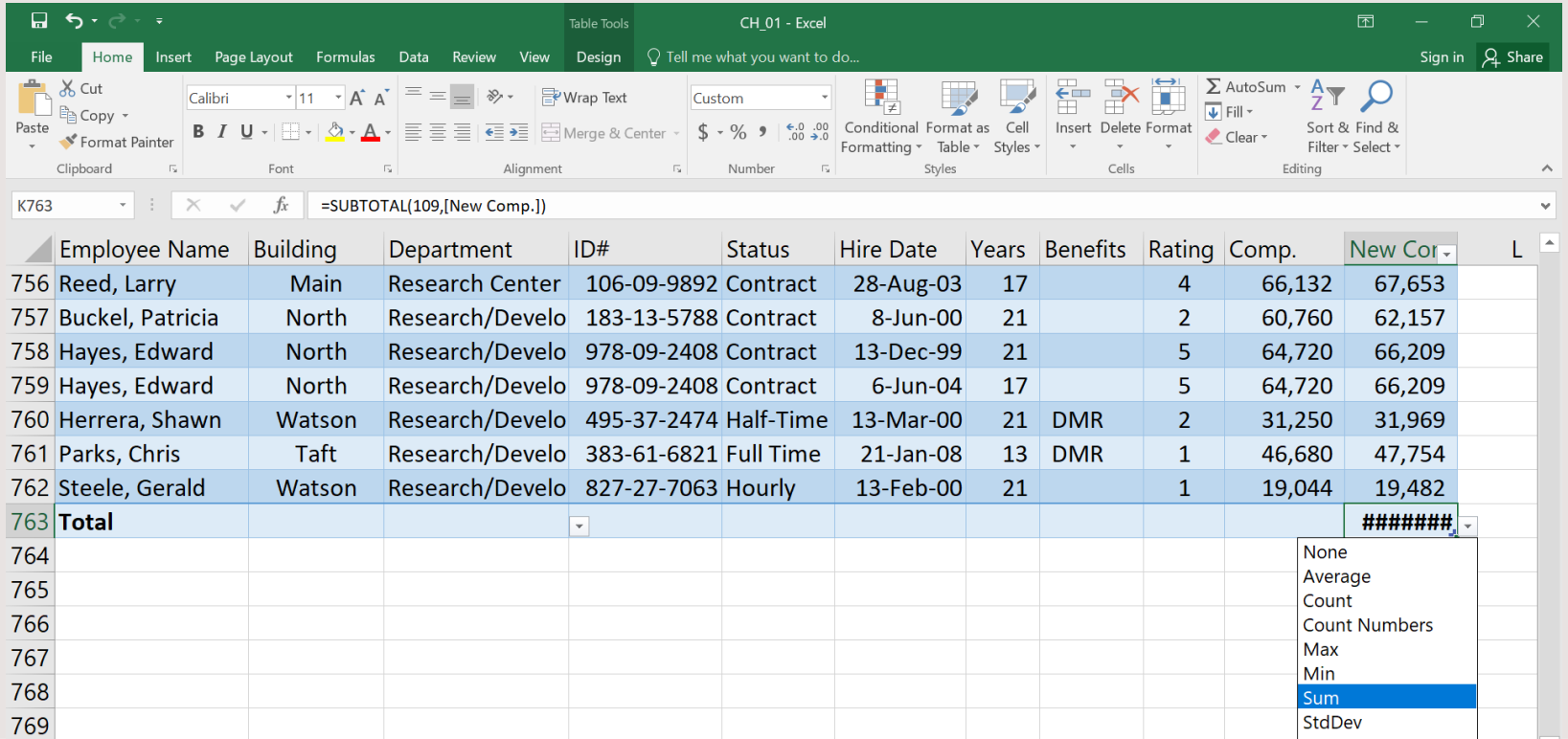


	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
	Product	Discount Band	Units Sold	Manufacturi	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Number	Month Name	Year	Colum		2.30%	
	Carretera	None	1618.5	$ 3.00	$ 20.00	$ 32,370.00	$ -	$ 32,370.00	$ 16,185.00	$ 16,185.00	1/01/2014	1	January	2014	$ 20.46			
	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:

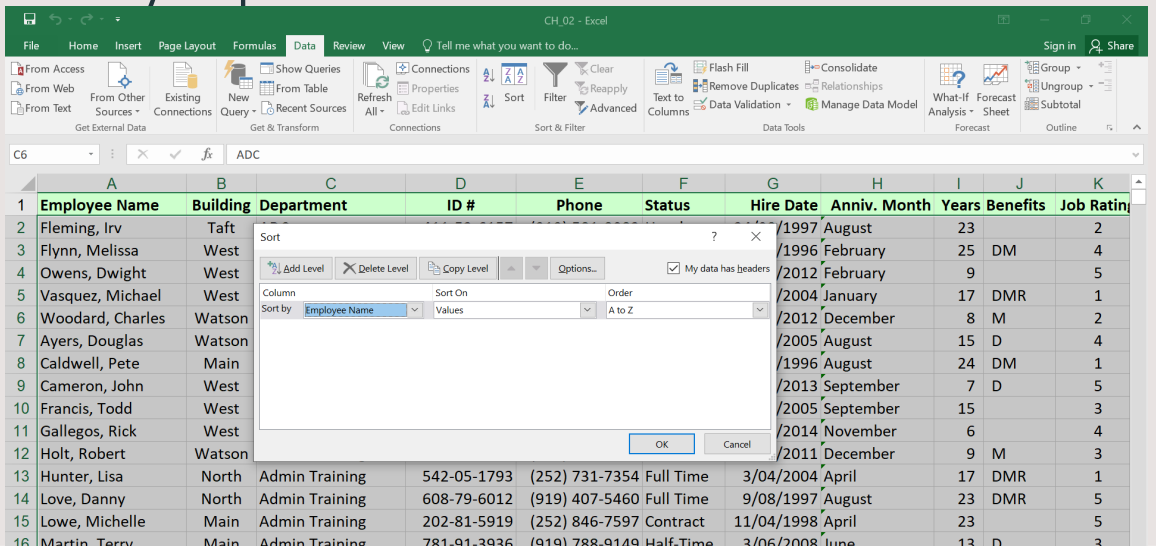
- How many people in this list (sampledatawinterathletes.xlsx) do not have weight?
  - Answer: click on table / Design tab / change the table name to : Winter_Table
  - Write this in empty cell : =COUNTBLANK(Winter_Table[Wt kg]) / press Enter

- 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.



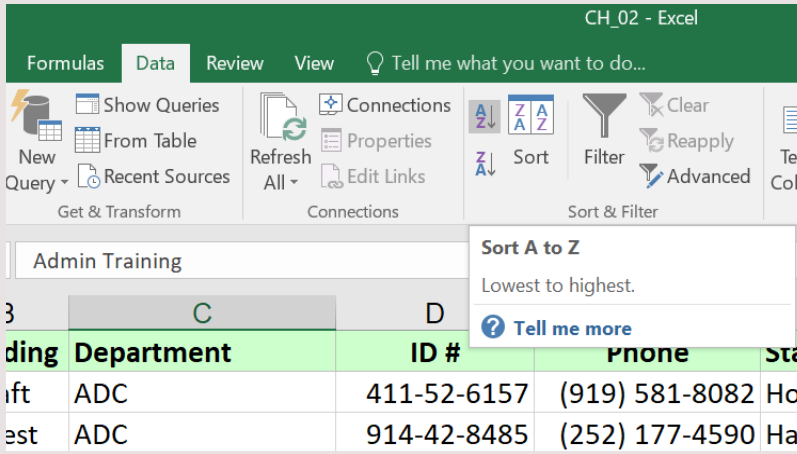
Sorting concepts and sort menu options

- 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



	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

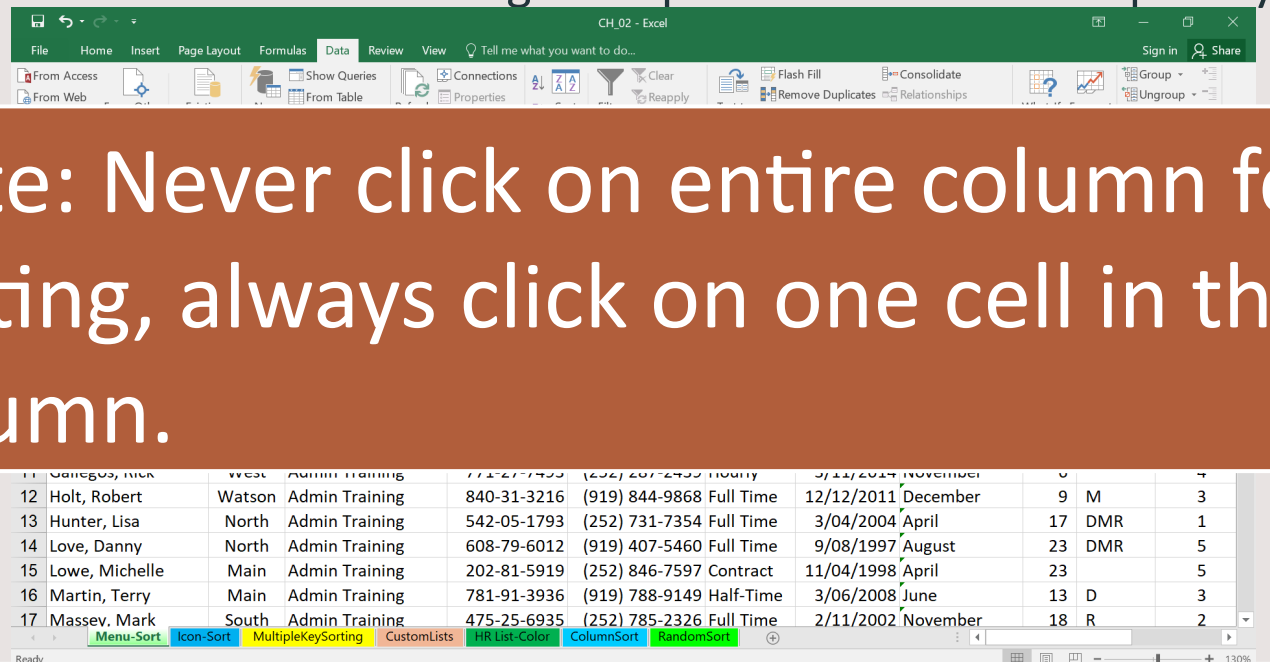
Sorting from menu icons



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

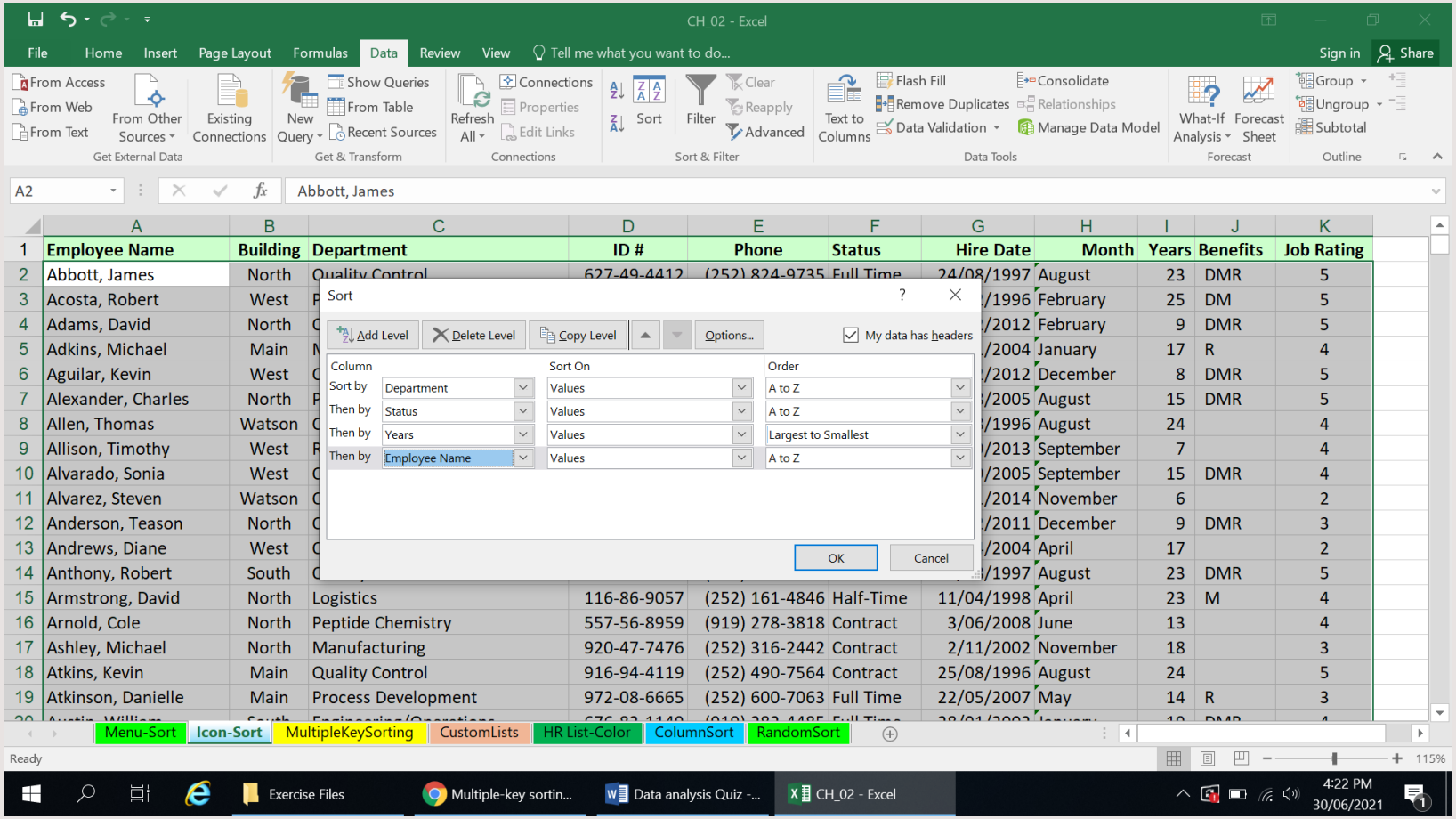
then click on sort A...Z. after that click on one of the cells in Department and then click on sort A...Z.

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



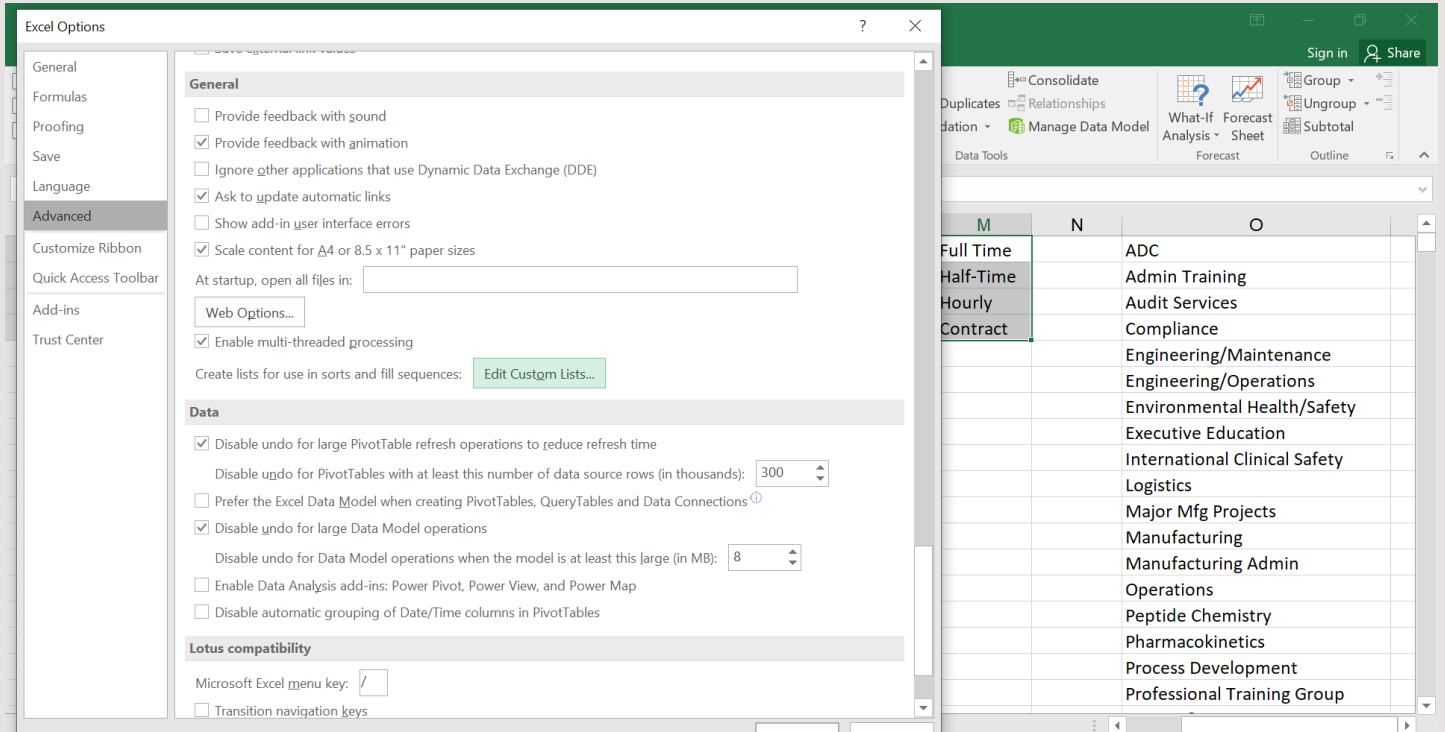
Note: Never click on entire column for sorting, always click on one cell in the column.

Multiple key sorting

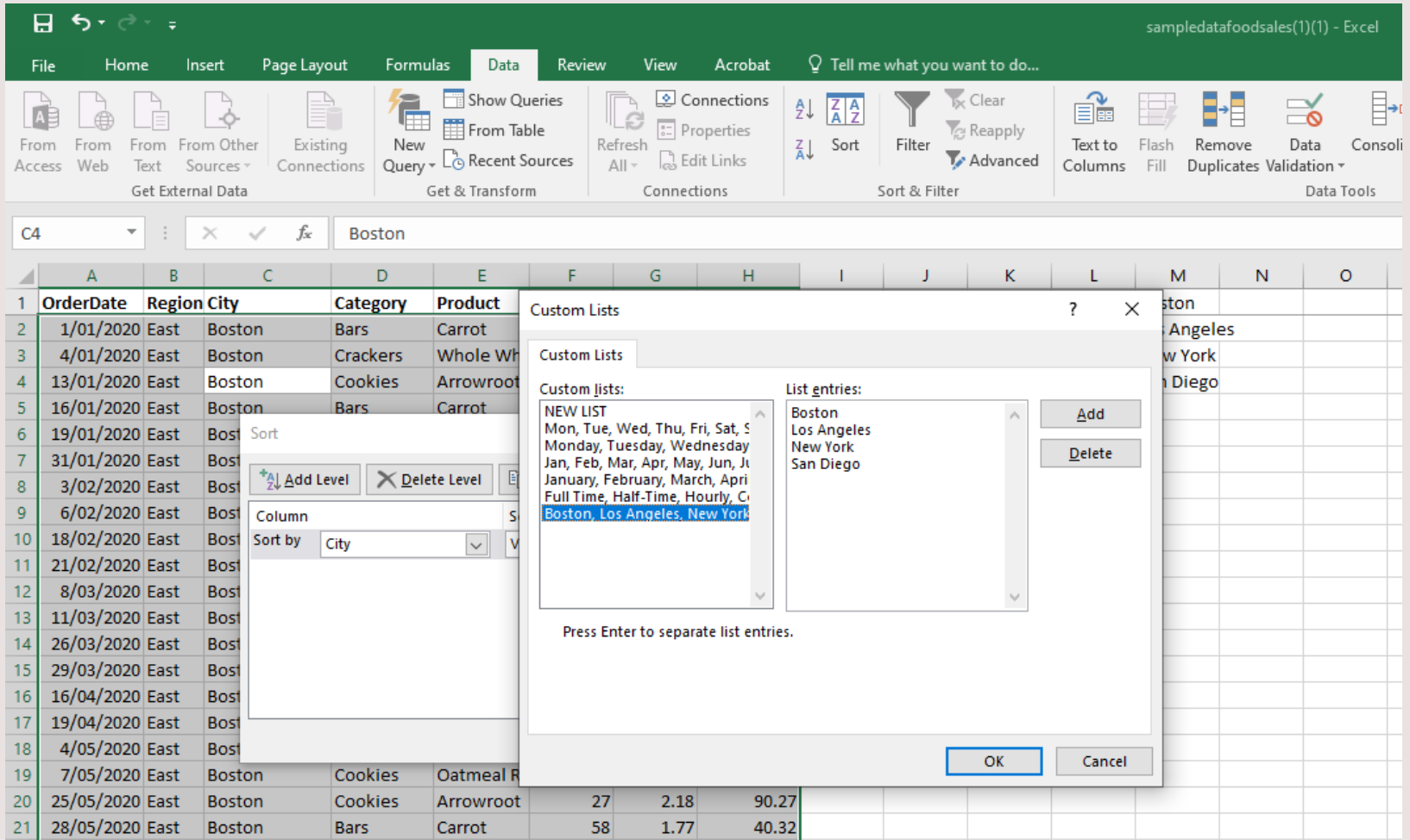


Sorting based on the order of data in custom lists

- 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



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



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

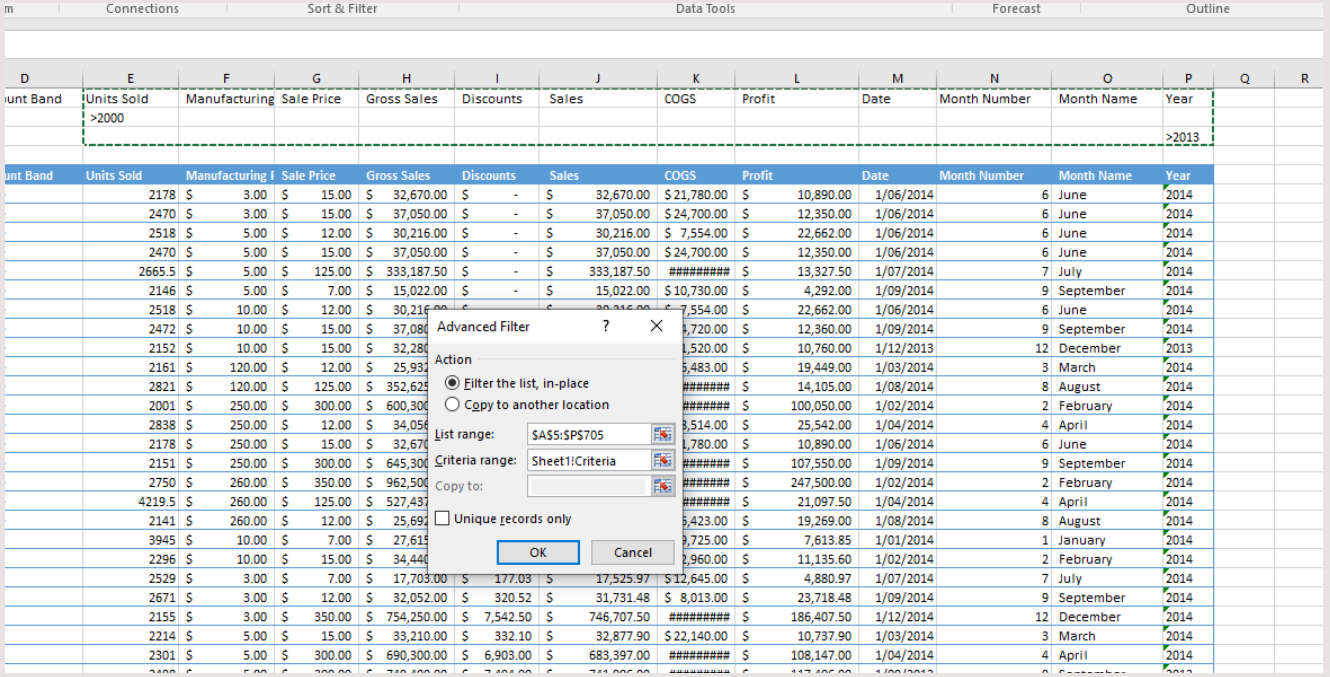
Filtering interactively using slicers with tables

From insert tab / table / insert slicer /

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

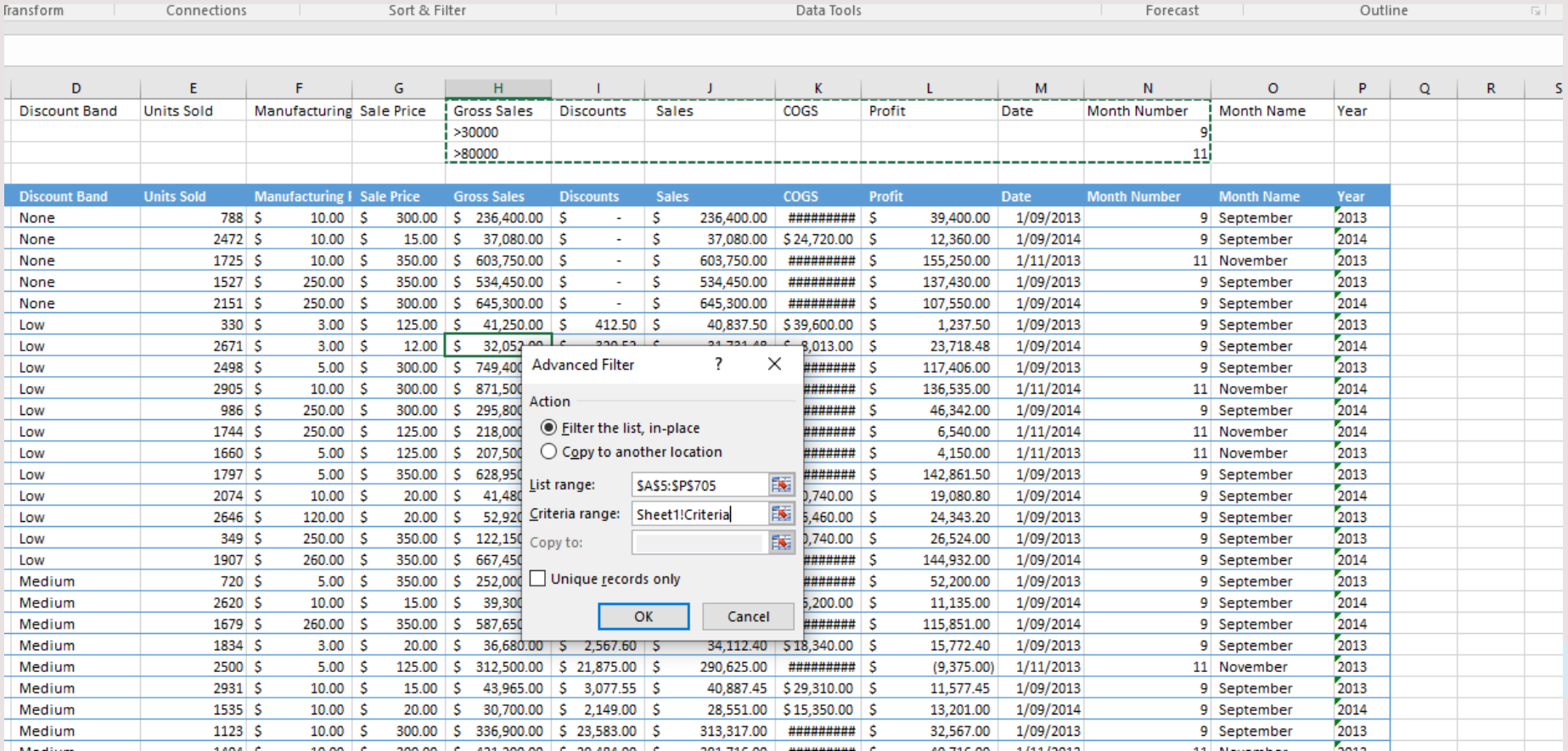
Using Advanced Filter for complex "OR" criteria

- Open sample-xlsx-file-for-testing.xslx
- 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.



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
	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	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	1/11/2013	11	November	2013
None	1527	$ 250.00	$ 350.00	$ 534,450.00	$ -	$ 534,450.00	#########	$ 137,430.00	1/09/2013	9	September	2013
None	2151	$ 250.00	$ 300.00	$ 645,300.00	$ -	$ 645,300.00	#########	$ 107,550.00	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	$	$		$ 23,718.48	1/09/2014	9	September	2014
Low	2498	$ 5.00	$ 300.00	$ 749,400				$ 117,406.00	1/09/2013	9	September	2013
Low	2905	$ 10.00	$ 300.00	$ 871,500				$ 136,535.00	1/11/2014	11	November	2014
Low	986	$ 250.00	$ 300.00	$ 295,800				$ 46,342.00	1/09/2014	9	September	2014
Low	1744	$ 250.00	$ 125.00	$ 218,000				$ 6,540.00	1/11/2014	11	November	2014
Low	1660	$ 5.00	$ 125.00	$ 207,500				$ 4,150.00	1/11/2013	11	November	2013
Low	1797	$ 5.00	$ 350.00	$ 628,950				$ 142,861.50	1/09/2013	9	September	2013
Low	2074	$ 10.00	$ 20.00	$ 41,480				$ 19,080.80	1/09/2014	9	September	2014
Low	2646	$ 120.00	$ 20.00	$ 52,920				$ 24,343.20	1/09/2013	9	September	2013
Low	349	$ 250.00	$ 350.00	$ 122,150				$ 26,524.00	1/09/2013	9	September	2013
Low	1907	$ 260.00	$ 350.00	$ 667,450				$ 144,932.00	1/09/2014	9	September	2014
Medium	720	$ 5.00	$ 350.00	$ 252,000				$ 52,200.00	1/09/2013	9	September	2013
Medium	2620	$ 10.00	$ 15.00	$ 39,300				$ 11,135.00	1/09/2014	9	September	2014
Medium	1679	$ 260.00	$ 350.00	$ 587,650				$ 115,851.00	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	#########	$ (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	1/09/2013	9	September	2013

Advanced Filter

Action

- (•) Filter the list, in-place
- ( ) Copy to another location

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

Criteria range: Sheet1!Criteria

Copy to:

[ ] Unique records only

OK | Cancel

Activity

- Show me three more queries.

Using the Remove Duplicates command

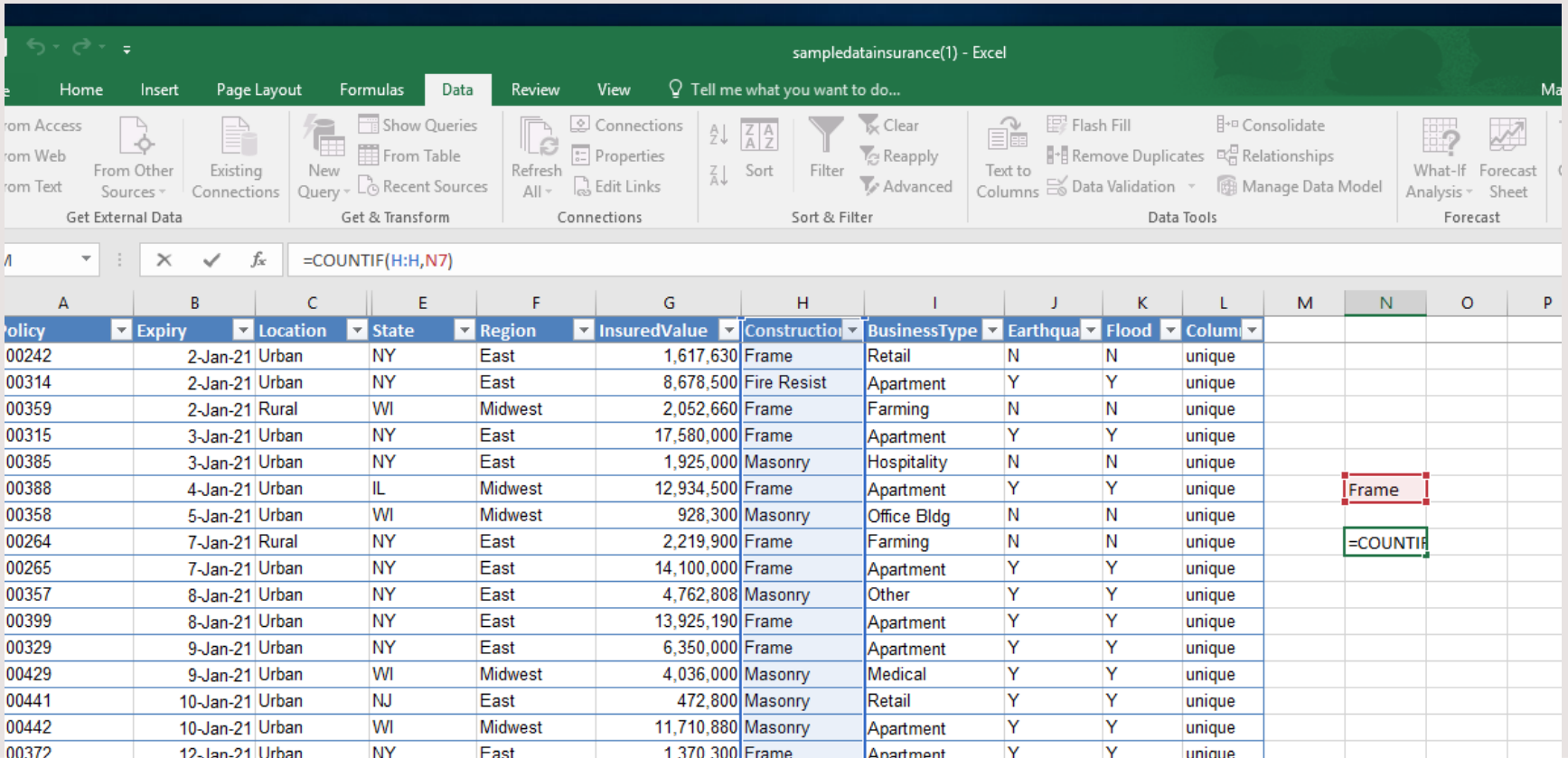
- Go to Data tab / Data tools section / remove duplicates

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

- Open sampledatainsurance.xslx
- 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)

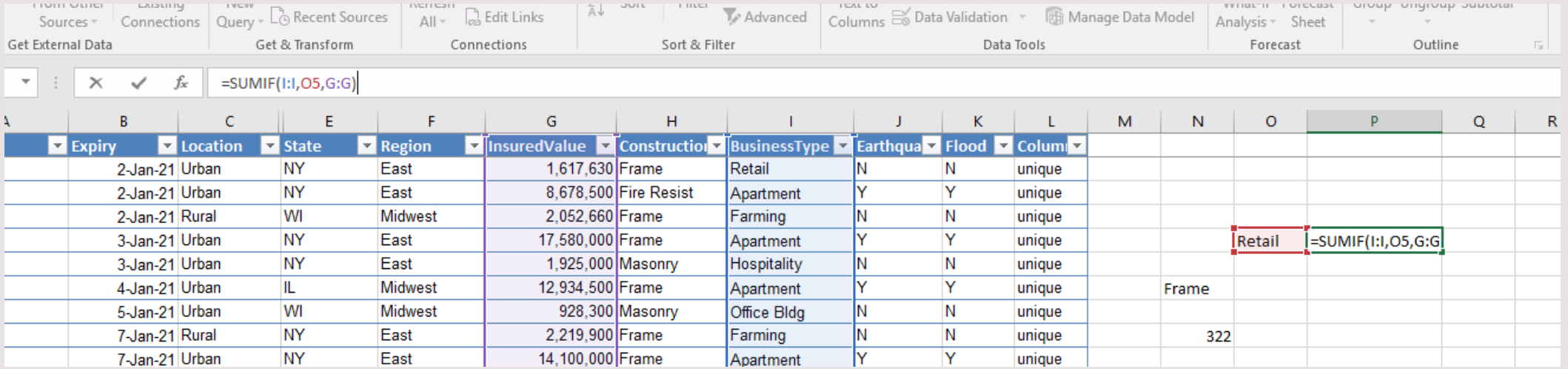
Using SUMIF, COUNTIF, and related functions for quick data analysis

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



Policy	Expiry	Location	State	Region	InsuredValue	Construction	BusinessType	Earthqua	Flood	Colum
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

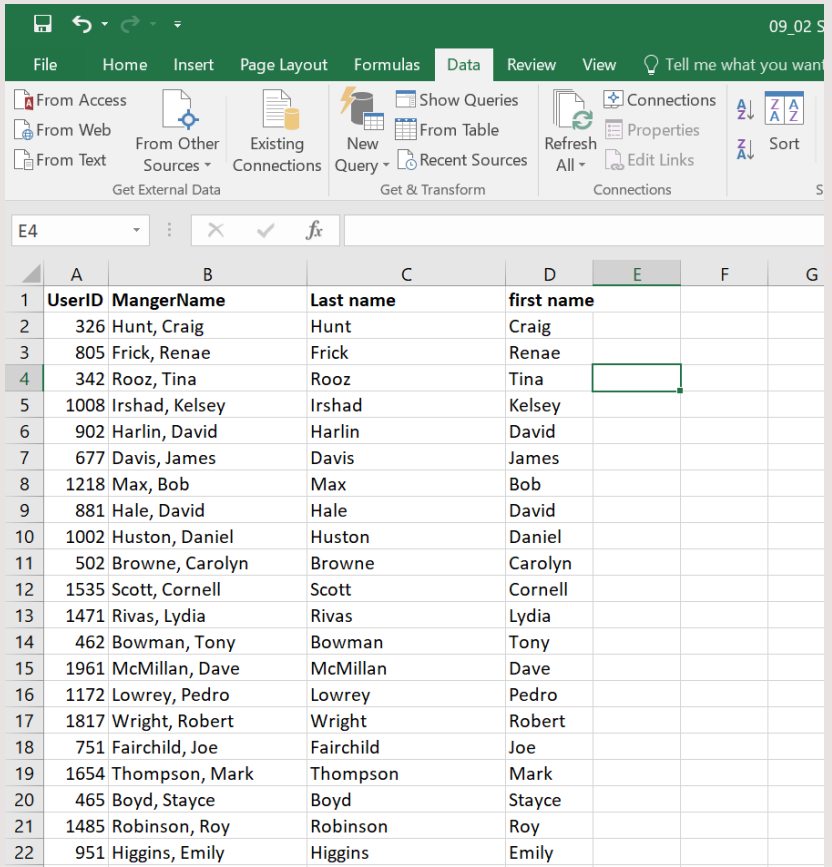
- How much is insured value for retail type?



- 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			

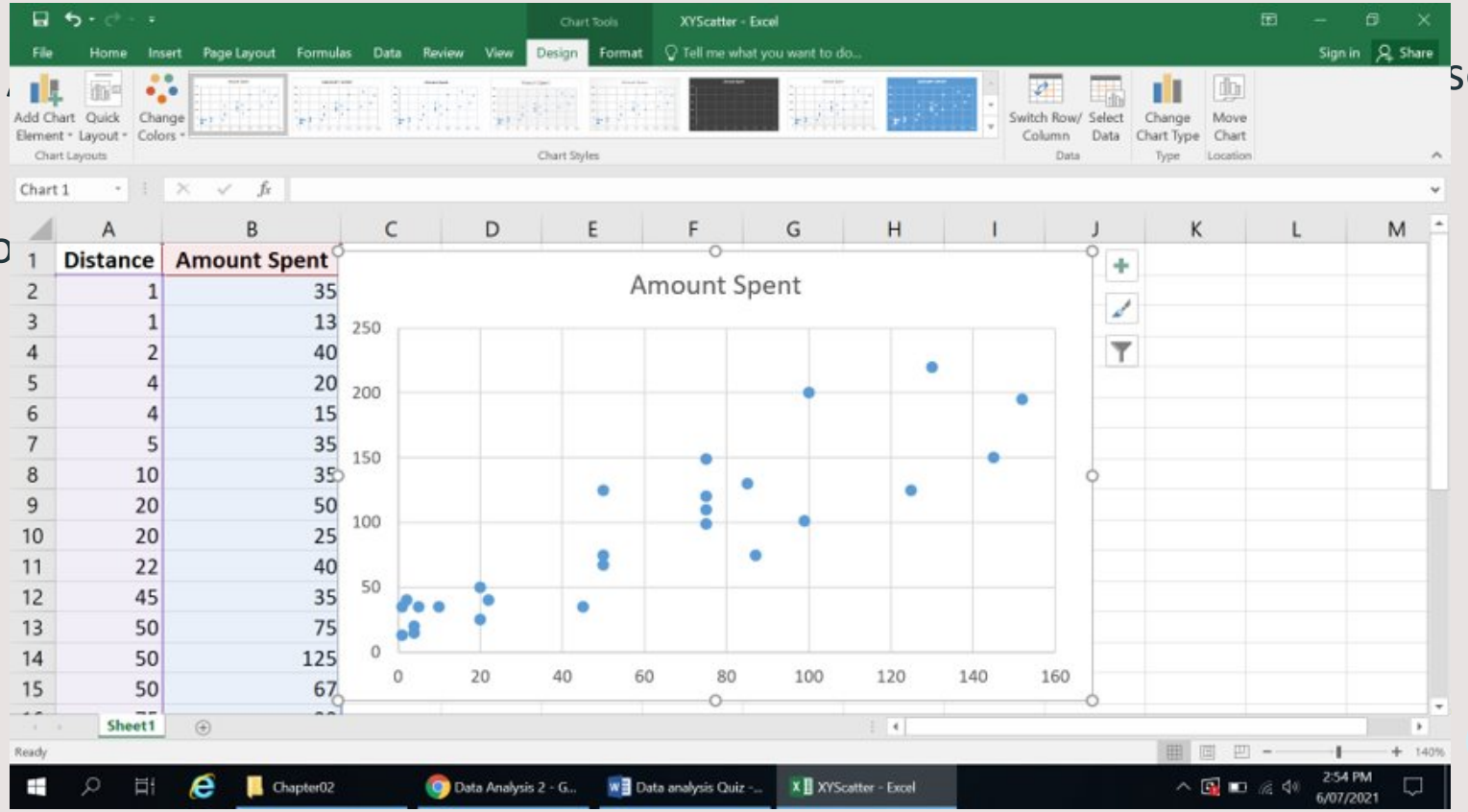
Calculate mean and median values

- 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.xslx
- 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)	

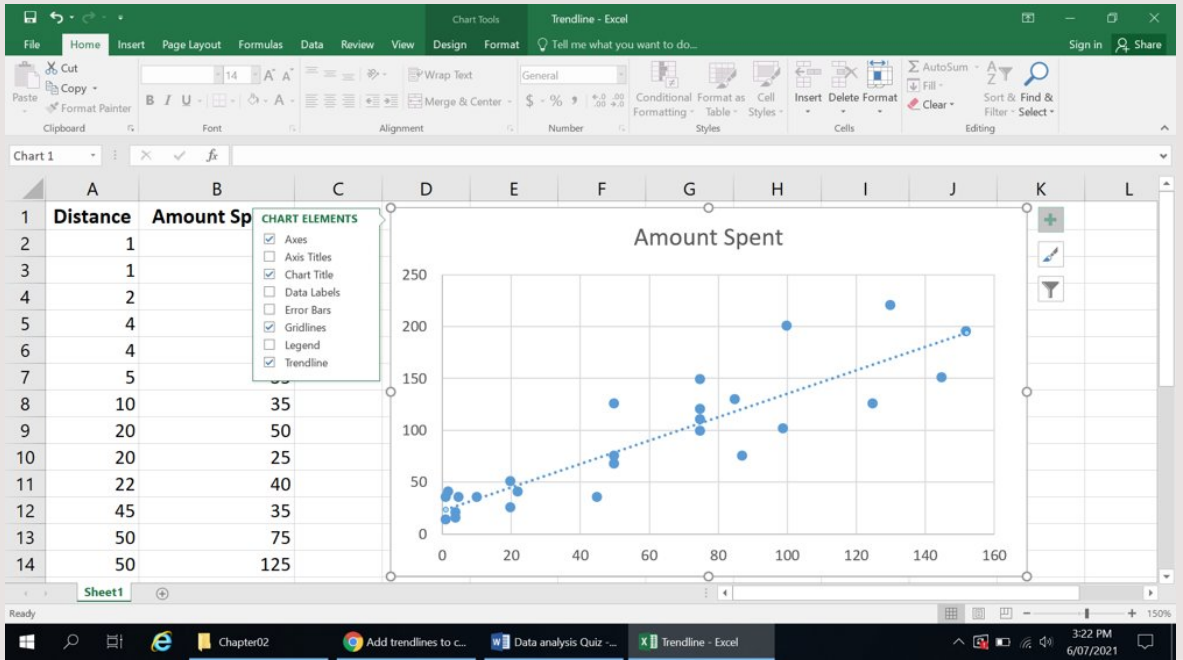
Identify relationships using XY scatter charts

- Open sample-xlsx-file-for-testing / XY Scatter tab
- Analysis to Discover relationship between values.
- Click on any cells , select any chart you want.
- Note: You have to



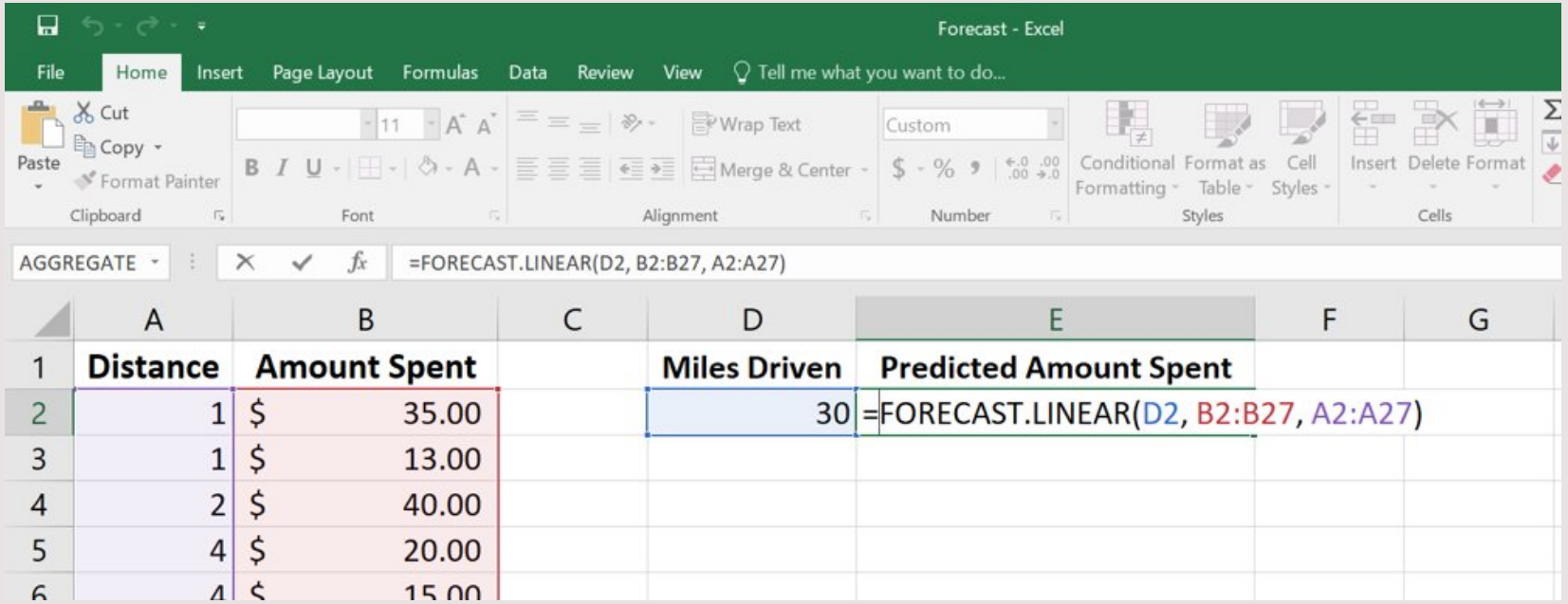
Add trendlines to charts

- How to examine your data visually by adding Trendline into a chart.
- Click on chart / click on + button / select Trendline on chart elemets
- 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.

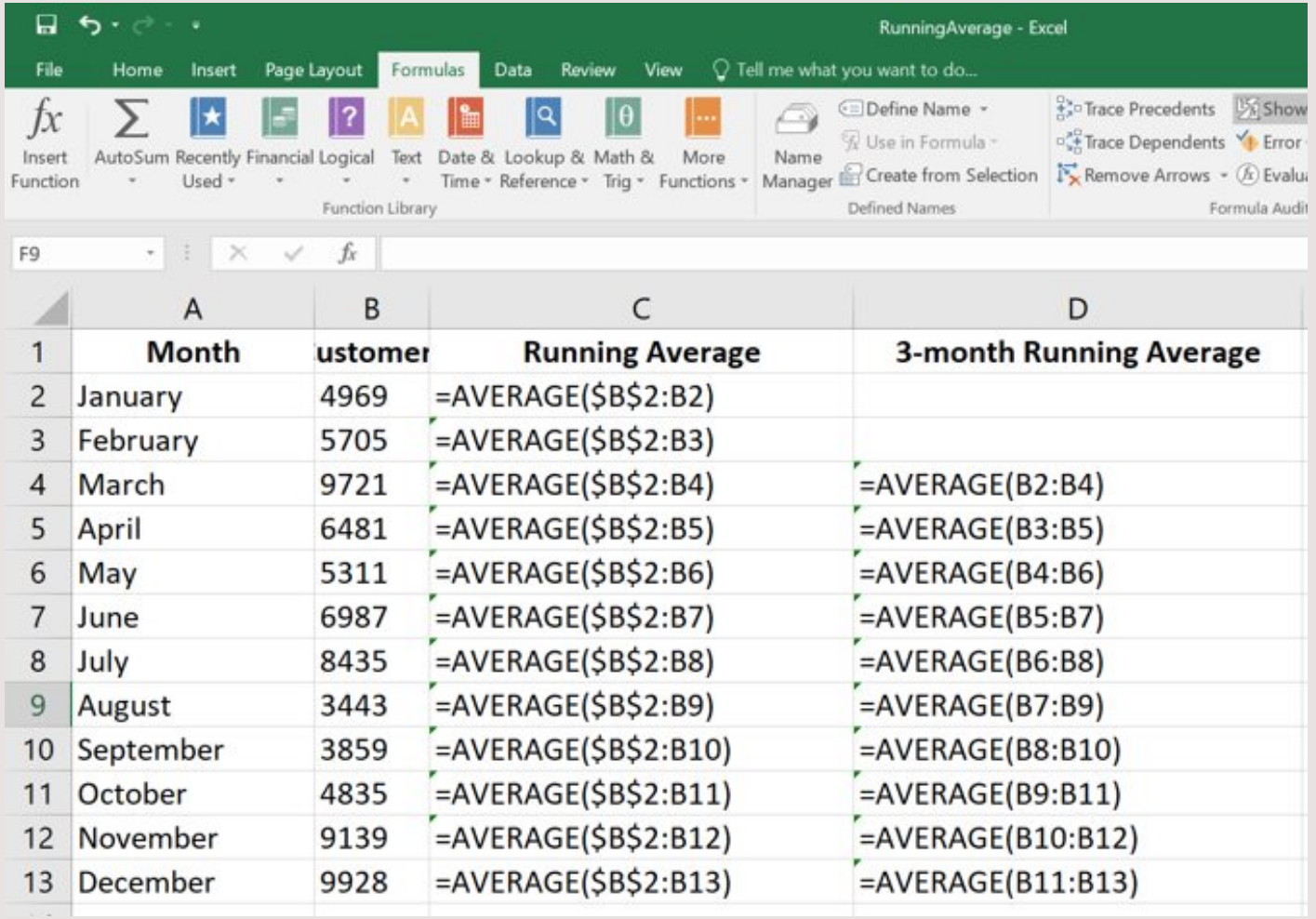


Forecast future results

- 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.



Activity: Calculate running averages



	A	B	C	D
1	**Month**	**ustomer**	**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.xslx 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.