MS 365 Excel Basics #8

Data Analysis, Tables, PivotTables, Slicers, Excel Charts & More!

Table of Contents

Data Analysis

- Data Analysis = Convert raw data into useful information to gain insight and make decisions.
  - Synonyms: Data Analytics, Analytics, Business Intelligence, Data Science, Data Analysis
  - Goal: help make data-driven decisions, which tend to be more accurate & help to achieve goals more consistently
- Data Analysis High Level Definition:
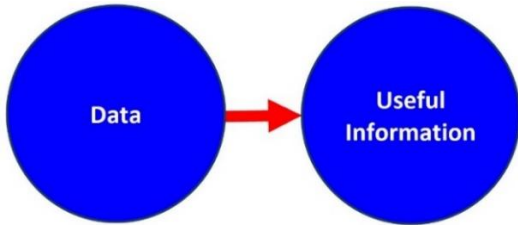
- Steps in data analysis process:
  1. Determine what questions need answers, what decisions need to be made, or what report, visual or other information you need.
  2. Get the data to help achieve the goals.
  3. Clean and shape the data (into proper tables). Build data model.
  4. Create useful information: reports, visuals and dashboards.
  5. Refresh when new data arrives.
  6. Change and update model as necessary.
- Examples:

1) Sales data in a table converted into a Monthly SalesRep Sales Report. What are top 3 sales amounts?
   Q: If a bonus is given for the 3 biggest monthly sales amounts, who gets bonus? A: Chantel, Jo, Tyrone.
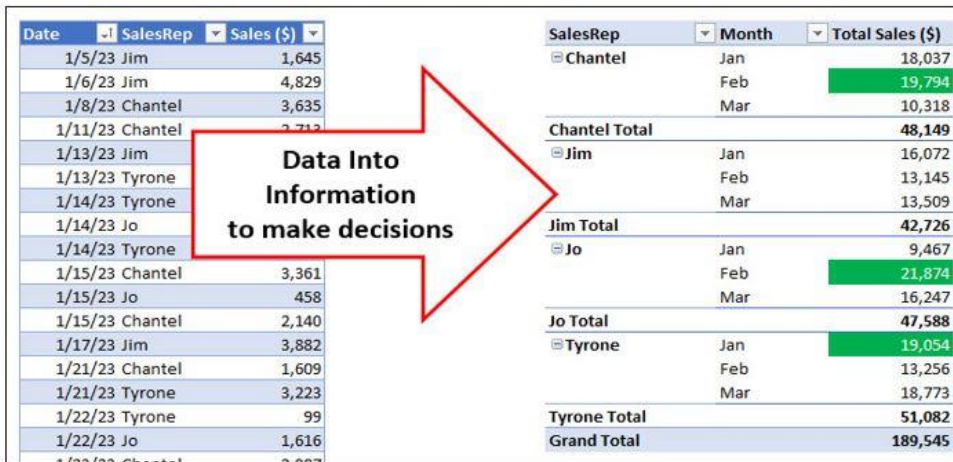
2) Sales data in a Table converted into a Monthly Sales Trends Chart. What is the trend?
   Q: What were sales trends over last six months? A: Mostly up.

Data

Data = Raw Data = data stored in its smallest form in a cell

Not Raw Data:

Sales Record
01/05/2023, Chantel, $1,645.01

Raw Data:

Date	Person	Sales
1/5/2023	Chantel	$1,645.01

Because each bit of data is stored separately, we can easily perform actions like sorting or summing a column of data.

Data Default Alignment

Default alignment indicates data type and can help track down errors

<= Text is aligned left	Numbers aligned right =>	Booleans are capitalized and centered	
Chantel	$1,645.01	TRUE	
Jim	1/5/2023	FALSE	
Excel	8:29 AM	TRUE	
$45..50	2/31/2025	TURE	<= Default alignment indicates error

Number Formatting

Number Formatting displays a number in a certain way without changing the underlying number

Number Format displays in certain way =>

Sales	Date	Time	
$51	2/28/2025	8:00 AM	
50.57	45716	0.333333	<= What is really in the cell

Dates are the number of days since Dec. 31, 1899. Examples: 1/1/1900 = 1, 1/2/1900 = 2, 2/28/2025 = 45,716
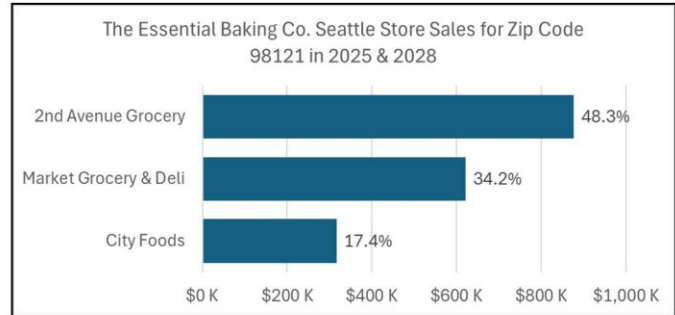Time is proportion of 24 hour day. Examples: 8 AM => 8/24 = 1/3 = 0.33333, 12 PM => 12/24 = 0.5

Information

Information is NOT data.
Information is created from data.
Information helps people make decisions and gain insight.

28	Date	Store	Zip	UPC	Units	Sales($)
29	1/1/2025	City Foods	98121	8-13305-01659-7	65	492.05
30	1/1/2025	City Foods	98121	8-13305-01647-4	23	133.4
39	1/1/2025	Metropolitan Market	98199	8-13305-01647-4	31	179.8
40	1/1/2025	Melrose Market	98122	8-13305-01265-0	30	213.3
60741	12/31/2028	Seattle Super Market	98108	8-13305-01265-0	65	477.1
60742	12/31/2028	Seattle Super Market	98108	8-13305-01647-4	34	207.4

The Essential Baking Co. Seattle Store Sales for Zip Code 98121 in 2025 & 2028

- 2nd Avenue Grocery — 48.3%
- Market Grocery & Deli — 34.2%
- City Foods — 17.4%

Tables

- Table = Proper Data Set = Data Set
    - A table is made up of:
        - Fields = columns in table. Synonyms for field are column and header.
            - Each field has a name that describes what data goes into the column, such as the field name Date which indicates that dates must be put into the column, or the field name Sales($) which indicates that sales amounts with the unit $ must be put into the column.
        - Record = one row in table.
            - Records contain raw data for each field such as date, sales or product data.
    - The structure of a table is as follows:
        - Field names in the first row (formatted differently than the data in the table)
        - Records of related data in subsequent rows
        - Empty cells or Excel Row/Column Headers, all the way around table
- Example of two tables:

Date	Person	Sales
1/5/2023	Chantel	1,645
1/6/2023	Jim	4,829
1/8/2023	Chantel	3,635
1/11/2023	Chantel	2,713
1/13/2023	Jim	679
1/13/2023	Tyrone	1,527
1/14/2023	Tyrone	3,152
1/14/2023	Jo	3.554

← Fields
← Records

Survey Data
Yes
Yes
No
Yes
No
No
No
Somewhat

← Fields
← Records

Why are tables important to data analysis?

- When data is not stored in a table, data analysis and simple tasks such as summing with a condition can be very difficult.
    - Data in table = Data analysis is easy.
    - Data NOT in table = Data analysis is difficult.
- Data analysis tools such as the PivotTable, Power Query and the Filter feature **require** that data is stored in a table.

What is NOT a table?

Not a Table = Collections of data that are not structured as tables.

Problem 1:
Records are in 1 column.   Proper Table

Data
01/05/2023
Chantel
$1,645
01/06/2023
Jim
$4,829
01/08/2023
Chantel
$3,635
01/11/2023
Chantel
$2,713

→

Date	Person	Sales
1/5/2023	Chantel	1,645
1/6/2023	Jim	4,829
1/8/2023	Chantel	3,635
1/11/2023	Chantel	2,713

^ ^ ^
With this structure, it is hard to add numbers or count names

Problem 2:
Each row contains data for 4 records.

Units Sold:

Week/SalesRep	Chantel	Jim	Jo	Tyrone
Weel 1	776	545	465	814
Week 2	697	706	671	555
Week 3	753	607	410	703

^ ^ ^
With this structure, it is hard to sort or filter

Proper Table ↓

Week	SalesRep	Units
Weel 1	Chantel	776
Weel 1	Jim	545
Weel 1	Jo	465
Weel 1	Chantel	814
Week 2	Jim	697
Week 2	Jo	706
Week 2	Chantel	671
Week 2	Jim	555
Week 3	Jo	753
Week 3	Chantel	607
Week 3	Jim	410
Week 3	Jo	703

Types of Tables

- Fact Tables
  - These are the tables that have the data to summarize, such as sales tables, survey result tables, student grade data tables, internet usage tables, baseball statistics tables and other tables that have data to summarize or measure.
- Examples:

Date	Person	Product ID	Units Sold
1/5/2023	Chantel	Q4369	45
1/6/2023	Jim	S5189	43
1/8/2023	Chantel	A3451	75
1/11/2023	Chantel	Q4369	61
1/13/2023	Jim	S5189	89

← Each row represents a sale or transaction

Date	Close GOOG
11/26/2024	$20.43
11/25/2024	$20.43
11/22/2024	$19.74
11/21/2024	$18.79
11/20/2024	$18.65

← Each row represents the stock price at the end of the day

Student ID	Student	Class Taken
A2-088-05-0499	Chike	Acct& 455
A2-088-05-0933	Jamila	Acct& 202
A2-088-06-7551	Mizuki	Acct& 455
A2-088-06-7551	Mizuki	Acct& 202
A2-088-01-4811	Akosua	Acct& 203

← Each row represents a class that the student attended

- Lookup Tables / Dimension Tables
  - These are tables that have a unique list in the first column (called a primary key), such as product ID, customer ID, invoice number, or a unique list of dates, and have other columns that have lookup data such as product price, customer name, student phone, or have report attribute data such as month name, quarter name, product name, customer name or other attribute data related to the primary key.
- Examples:

Unique Identifier   Items to lookup

Product ID	Product	Price
Q4369	Quad	43.95
S5189	Sunshine	24.95
A3451	Aspen	28.95

Unique Identifier   Report attribute data

Date	Year	Month	Fiscal Q
12/31/2024	2024	Dec	3
1/31/2025	2025	Jan	4
2/28/2025	2025	Feb	4

Unique Identifier   Items to lookup

CustomerID	Name	Zip
34-00987-M	56th Market	98103
65-00514-S	Sosio's Store	98101
80-00781-M	Blue Market	98106

The Grain of a Table

- The word "grain" means the size of the detail in the table. For example, in a sales table with a transactional or product grain, each row has the sales for a single product, or a single transaction; in this case, the grain or size of the table would be transactional. As a second example, in a sales table with an invoice grain, each row has the total sales for one invoice, where the total invoice sales can be the sum of multiple product or transactional amounts; in this case the grain or size of the table would invoice level. The invoice level grain is larger than the transactional grain. As a third example, in a sales table with a monthly grain, each row has the total sales for a month; whereas in a sales table with a yearly grain, each row has the total sales for a year, where the total yearly sales is the sum of 12 months of sales. The yearly grain is larger than the monthly grain. If we compare the grain of these four tables, the year grain is bigger than the month grain, the monthly grain is bigger than the invoice level grain, and the invoice level grain is bigger than the transactional grain. It is helpful to think of grain as going from small, like sand, to big, like a huge rock or boulder. The reason that grain is important is because if you are trying to create a report or a visual from tables with different grains, you must take special care to make sure that all the numbers used possess the same grain or size. Just think of trying to compare transactional sales amounts to monthly budgeted amounts: you cannot do it until you summarize the transactional amounts at the monthly grain level.
- Examples of tables with different grains:

The word **"grain"** means the size of the detail in the table.

Date	Invoice No.	Product Cod	Sales($)
1/1/2025	50256	Q4369	492.05
1/1/2025	50256	S5189	133.4
1/1/2025	50257	A3451	179.8
1/1/2025	50258	Q4369	213.3
12/31/2028	50258	S5189	477.1
12/31/2028	50258	A3451	207.4

The size of the detail is
is at the sales or transactional level
Month Grain > Invoice Grain > **Transactional Grain**

Invoice No.	DeliveryFee	Discount ($)
50256	7.7	71.93
50257	4.81	35.99
50258	6.74	44.4

The size of the detail is
is at the invoice level.
Month Grain > **Invoice Grain** > Transactional Grain

Month	Ad Cost ($)	Sales ($)
Jan-25	20,749	192,642
Feb-25	20,613	196,425
Sep-26	15,832	211,775
Oct-26	21,954	205,801
Nov-26	31,034	687,783
Dec-26	26,273	568,151

The size of the detail is
is at the Month level.
Month Grain > Invoice Grain > Transactional Grain

Main Data Analysis tools in Excel

- Excel Tables
  - Excel Table feature = Convert proper data sets to a Table Object that can expand and contract and auto column fill formulas. This tool is used so that you can have a dynamic data source and objects such as PivotTables, Charts and formulas can update when new data is added to the Excel Table.
- Sort feature
  - Sort data A-Z (Ascending) or Z-A (Descending). This tool is used when you want to organize data from A-Z or Z-A.
- Filter feature
  - Dropdown arrows allow you to show or extract records based on conditions, criteria and logical tests.
- Excel Charts
  - Used to visualize data such as a Line Chart to show trends over time, Column or Bar charts to show differences across categories, or X-Y Scatter Charts to visualize a relationship between two quantitative variables.
- PivotTables
  - This is a drag and drop summary report tool that makes calculations based on conditions, criteria and logical tests. It allows to make aggregate calculations with the Summarize Values By option or other various calculations with the Show Values As option.
- Conditional Formatting
  - Allows you to highlight data based on conditions, criteria and logical tests. There are built-in options and the ability to create logical formulas to apply highlighting based on any type of logical test.
- Worksheet formulas
  - Used to create helper columns in tables, like with XLOOKUP function) or make reports based on conditional calculations, like with SUMIFS, COUNTIFS and other functions.
- Dynamic Spilled Array Functions
  - Functions such as: GROUPBY, PIVOTBY, SORT, FILTER, and many more, that allow you to create PivotTable like reports, sort data or filter data. These are used when you need the model to update instantly when source data changes.
- Flash Fill
  - One-time data cleaning tool that allows you to provide an example of how you want the data cleaned. This tool is only used when the cleaning pattern is simple, and it is a one-time cleaning task. If it is not simple, or you need the solution to update when source data changes, then you use Power Query or worksheet functions.
- Power Query
  - Tool to import, clean, transform data and then load to one of the following locations: Worksheet, PivotTable Cashe, Power Pivot Data Model. This is the most advanced data cleaning and transforming tool in Excel. This tool is in Power BI Desktop also.
- Functions and Features to Clean & Transform Data
  - Functions such as: TEXTBEFORE, TEXTSPLIT, LEFT, RIGHT, MID, TOCOL, WRAPCOLS, and many more. Use functions to clean data when you need the solution to update instantly when source data changes, like with What-if analysis.
- Relationships
  - Creates relationships between tables and can be a substitute for XLOOKUP. It is an option in the Data Ribbon tab that automatically adds tables to the Power Pivot Data Model, creates Relationships and then allows you to create Data Model PivotTables.
- Power Pivot
  - Tool to handle big data and make Data Model PivotTables. This tool is used when you need Excel worksheet reports and visualizations and you have large data (about 50,000 rows or more), calculations that are hard to do with a PivotTable, or you have related tables of data.
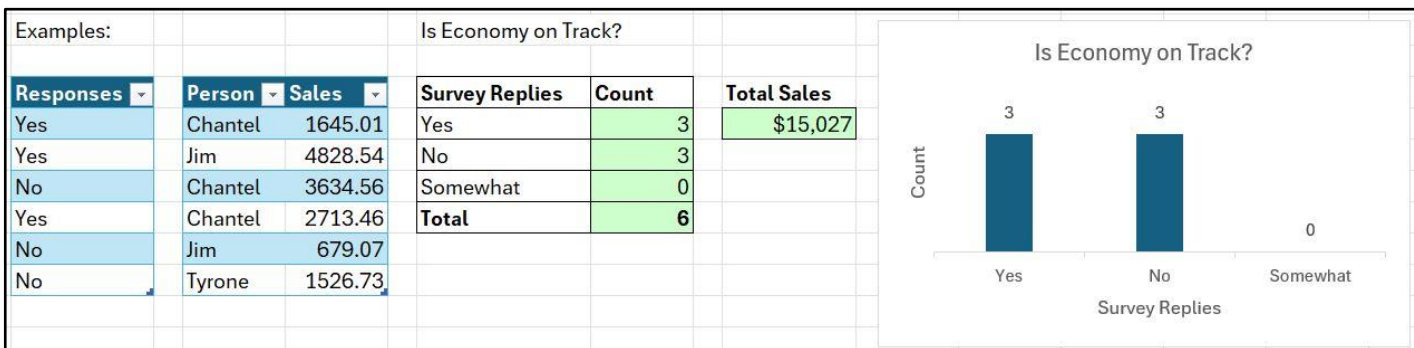
Excel Tables

- Excel Table Feature
  - o If you have a proper data set, you can convert your proper data set to an Excel Table by selecting one cell in the proper data set, click on the Table button in Table group in Insert Ribbon Tab, or just use **keyboard: Ctrl + T.**
  - o ALWAYS name your Excel Table. To name an Excel Table:
    - Select one cell in Excel Table
    - Click the Table Design Ribbon tab
    - In the Properties group click in Table Name text box, type name (no spaces), then press Enter.
  - o You can add new records to Excel Table by typing or pasting new data in the first row below the Excel Table.
  - o Filter and Sort dropdown arrows at top of each field allow you to sort or filter the records in the Excel Table.
- If new rows and columns are not added to Excel Table, you will need to change option settings for Excel Tables: File menu, Options button, Proofing tab on left, "AutoCorrect Options" button, check "Include new rows and columns in table".

Advantages of Excel Tables

- The advantage of using the Excel Table feature is that when new rows or columns are added to the Excel Table, all objects, such as PivotTable, Charts, Formulas, or other features can be refreshed so that the new data will be incorporated into the object.

Example BEFORE adding new data to the Excel Tables:

Example AFTER adding new data to the Excel Tables:
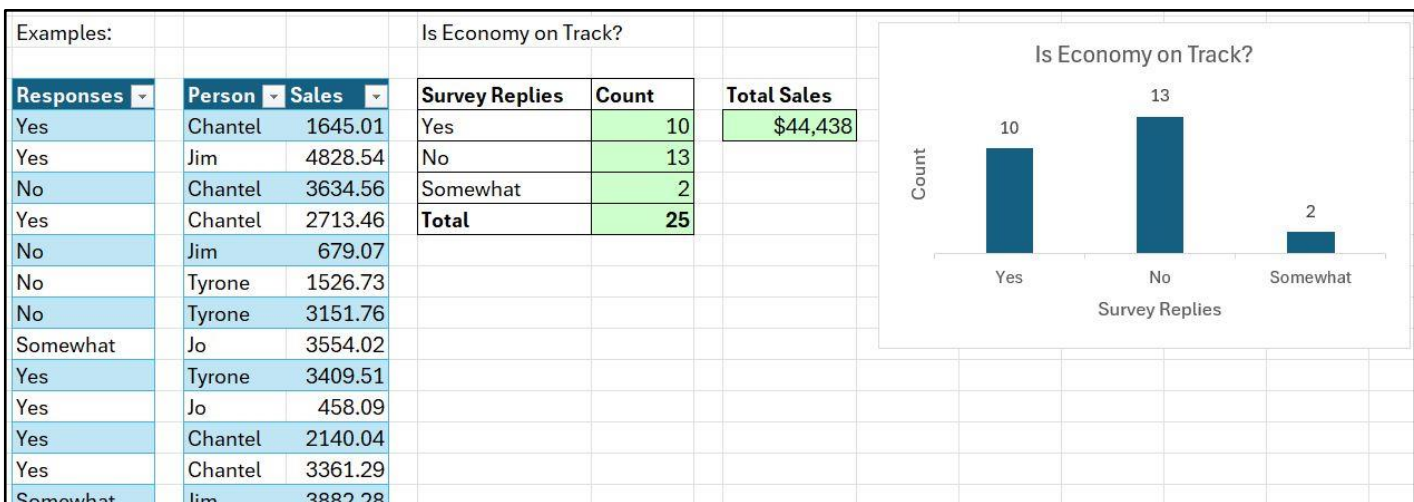
PivotTables & Slicers

- What PivotTables do:
  - Create Summary Reports that contain calculations with Conditions or Criteria.

- Summary of how to create PivotTable:
  - Click in one cell in Proper Data Set
  - Insert Ribbon Tab, Tables group, PivotTable button.
    - Keyboard = Alt, N, V, T.
  - From Field List, drag field name to Rows area or Columns area or Filter area. These are the conditions/criteria for the calculation in the Values area of the PivotTable.
    - Fields in the Rows area add a condition/criterion to the row
    - Fields in the Columns area add a condition/criterion to the column
    - Fields in the Filter area add a condition/criterion to the entire PivotTable.
  - From Field List drag the field you would like to make a calculation on to values area.
    - Number fields default to a SUM calculation (add numbers)
    - Text fields default to a COUNTA calculation (count non-empty cells)
  - With a cell selected in the PivotTable, click on PivotTable Tools Design Ribbon Tab, go to the Layout group, click drop- down for Report Layout and then click on "Show in Tabular Form" or "Show in Outline Form". These Layouts shows the Field Names in your Report.
  - To add Number Formatting to the Values area of the PivotTable, click in one cell in the Values area of the PivotTable, Right-click the cell and click on "Number Format…", then in the Number Formatting dialog box select the Number Formatting that you would like and then click OK.
  - If you want to change the summary calculation in the Values area, right-click the Values area in the PivotTable Report, point to one of these two options:
    - "Summarize Values By" and then select an aggregate calculation such as "Average", or "Max", or "More Options".
    - "Show Values As" and then select a calculation such as "% of Column Total", "Running Total", "Difference From" or other calculations.

- Default PivotTable Report Layout
  - To set the layout default for all PivotTable, click on File menu, Options button, in the Excel Options dialog box, click the Data tab on the left, click the Edit Default Layout button, then in Report Layout check box, select "Show in Tabular Form" or "Show in Outline Form".

- Slicers can be added to the PivotTable to add a condition/criterion to the entire PivotTable, similar to a field in the Filter area. To insert a Slicer into a PivotTable, click in one cell in the PivotTable Report, then go to the PivotTable Tools Analyze Ribbon Tab, then in the Filter Group, click the Insert Slicer button

- Standard PivotTable vs. Data Model PivotTable.
  - A Standard PivotTable stores its data in the PivotTable cache. A Standard PivotTable is best when data is stored in one small table (less than about 50,000 rows of data) and you want simple calculations such as sum, averages, % of column totals or % of grand total.
  - A Data Model PivotTable stores its data in the Power Pivot Data Model Columnar Database. A Data Model PivotTable is best when you have a large amount of data, you have multiple related tables, or you want to make calculations that a Standard PivotTable cannot easily make.

- Standard PivotTable Cached Data
  - When you create a Standard PivotTable, Excel creates a copy of the source data and stores it in the Pivot Cache.

- o The Pivot Cache is stored in Excel's memory.
- o This is why the PivotTable does not update when source data changes.
- o If source data changes, you can right-click the PivotTable and click Refresh, or use the Refresh button in the Data Ribbon Tab.
- o You can't see Cached PivotTable Data, but that's the data the PivotTable references when you build your
- o PivotTable or change a Slicer selection or move rows/columns around.
- o When we group dates in a PivotTable, the Pivot Cache is where this action takes place. The Pivot Cache save the action of grouping by dates so that when we use the date column in other places, the date column will remain grouped.
- o When we use a Slicer or change the conditions and criteria in the PivotTable, these actions are interacting with the Pivot Cache of data, not the original source data.

Excel Charts

- What do Charts do?
  - o Visually portray quantitative data (number data).
  - o Give a quick impression of the number data.
  - o Create a picture that can communicate more quickly than just the numbers alone.
  - o Charts allow you to see patterns or trends that you may not be able to see if you are looking at just the number data.
  - o Allows you to make relative comparisons more quickly than if you are using a table
- Types of Charts
  - o **Column:** Use to compare differences across categories. Height of column conveys number.
  - o **Bar:** Use to compare differences across categories. Length of bar conveys number.
  - o **Stacked Column/Bar:** Good for displaying crosstabulation, emphasis on horizontal axis categories.
  - o **Clustered Column/Bar:** Good for displaying crosstabulation, emphasis on legend categories.
  - o **Line:** Use to show trend for a number variable over a category such as time.
  - o **X-Y Scatter:** Used to show relationship between two number variables (x and y variables).
- Rules for Charts:
  - o No Chart Junk.
  - o Eliminate all chart elements that do not help to communicate the message.
- Formatting Charts:
  - o You can add and remove chart elements by using the **Green +** on the right side of chart.
  - o To format a chart element, select element and use the keyboard Ctrl + 1 to open the Format Chart Element task pane.
- Link Labels to Cells
  - o Click on Chart Title, type an equal sign, click on cell with title, hit Enter.

Built-in Conditional Formatting

- Built-in Conditional Formatting
  - o Select cells and apply a conditional format from the Styles group in the Home Ribbon
  - o Each cell is evaluated to TRUE or FALSE.
    - ▪ TRUE = Formatting applied.
    - ▪ FALSE = Formatting not applied.
  - o To edit rule, go to Manage Rules