DATA 301
Introduction to Data Analytics
Spreadsheets: Microsoft Excel

Dr. Ramon Lawrence
University of British Columbia Okanagan
ramon.lawrence@ubc.ca

Why Spreadsheets and Microsoft Excel?

Spreadsheets are the most common, general-purpose software for data analysis and reporting.

Microsoft Excel is the most popular spreadsheet program with hundreds of millions of installations.

- The spreadsheet concepts translate to other products.

Excel and spreadsheets are not always the best tool for data analysis, but they are great for quick analysis, reporting, and sharing.

Spreadsheet Overview

A ***spreadsheet*** organizes information into a two-dimensional array of cells (a *table*).

A ***cell*** has two components:

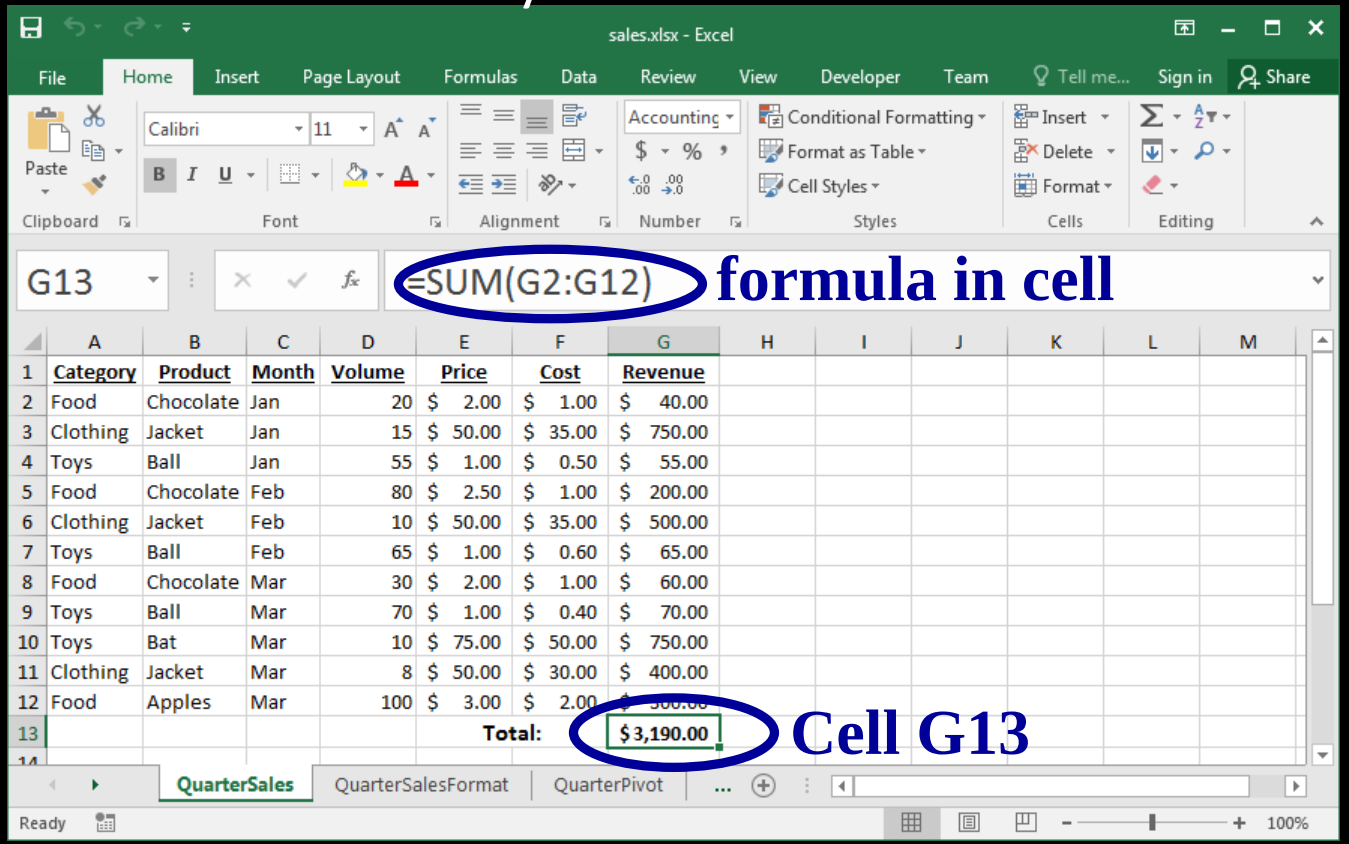
- an address - specified given a column letter and row number
- a location - that can store a number, text, or formula

The power of a spreadsheet is that we can write simple formulas (commands) to perform calculations and immediately see the results of those calculations.

Spreadsheets are very common in business and reporting applications.

Spreadsheet Addressing

A ***cell*** is identified by a column letter and row number.



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13						Total:	$ 3,190.00

Spreadsheet Addressing

The rows in a spreadsheet are numbered starting from 1.

The columns are represented by letters.

- A is column 1, B is column 2, …, Z is column 26, AA is column 27, …

A cell is identified by putting the column letter first then the row number.

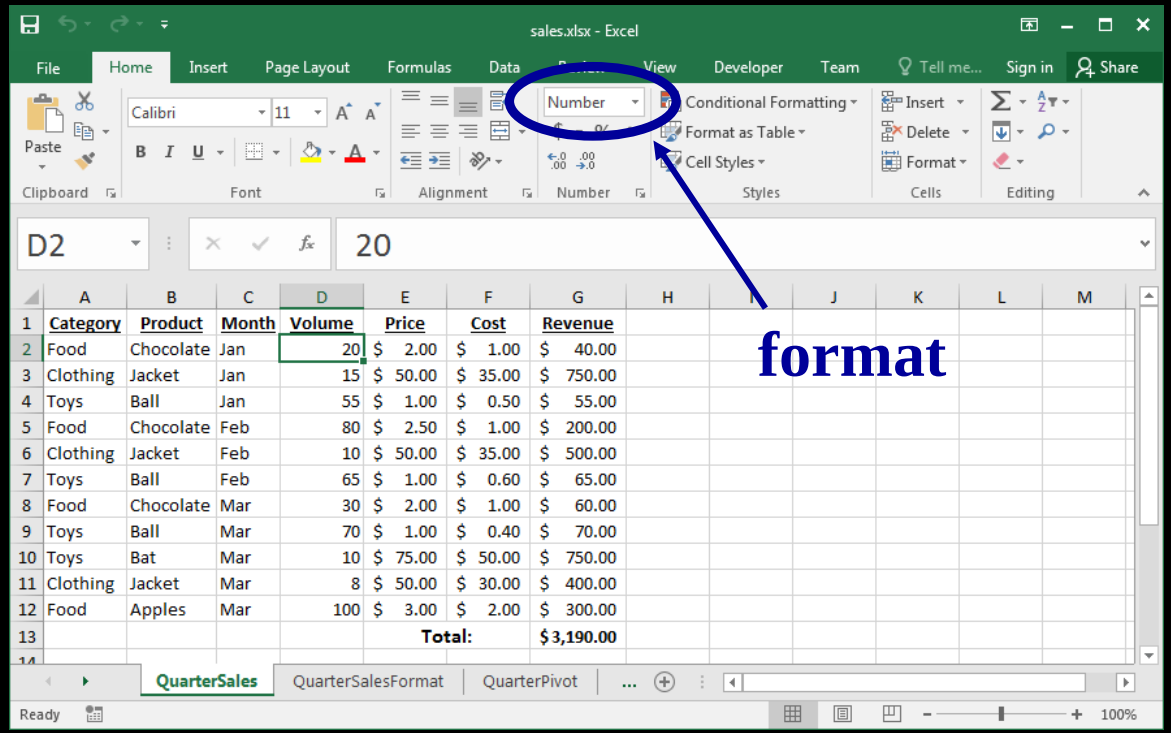
- e.g. B3 is the 2nd column and the 3rd row.

Question: What column number is AD? How about BAD?

Spreadsheet Data Entry

An entry is added to a cell by clicking on it and typing in the data.

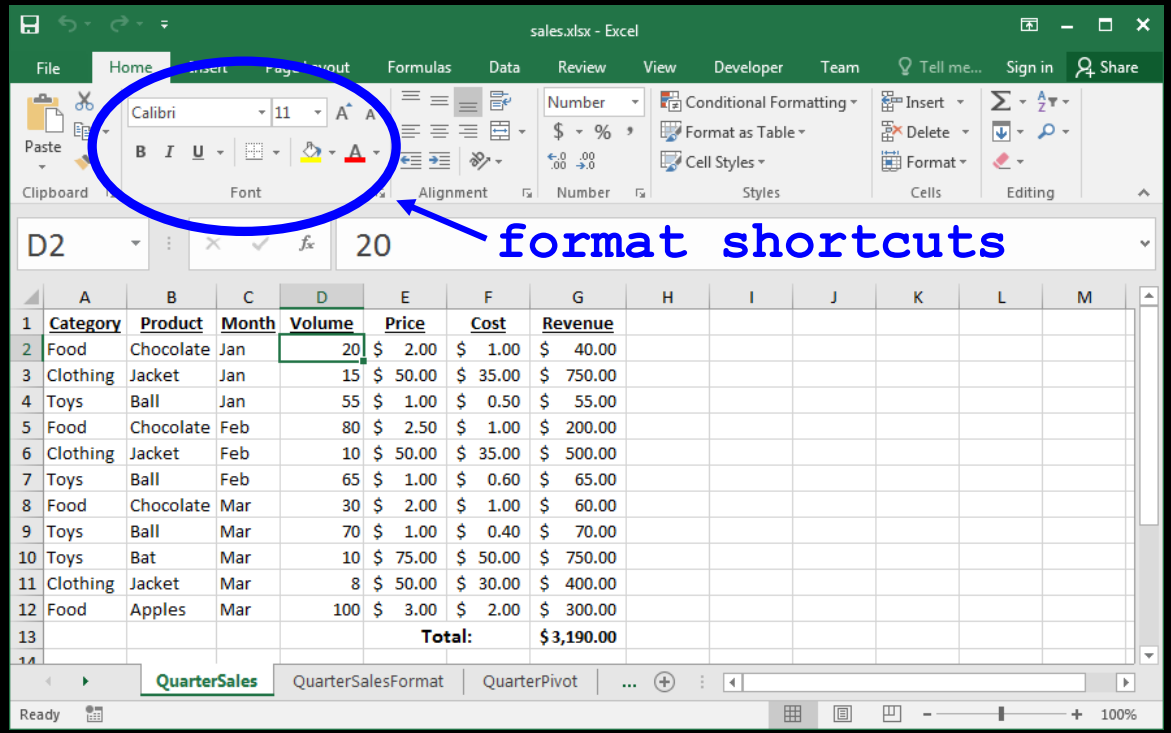
- The data may be a number, text, date, etc. Type and *format* are auto-detected.



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00

Spreadsheet Formatting

Formatting: bold, italics, underline, fonts, colors



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00

Spreadsheet Selecting Cells

Multiple ways of selecting cells:

1) With the mouse, (left) click and drag mouse to select a rectangle region of cells.

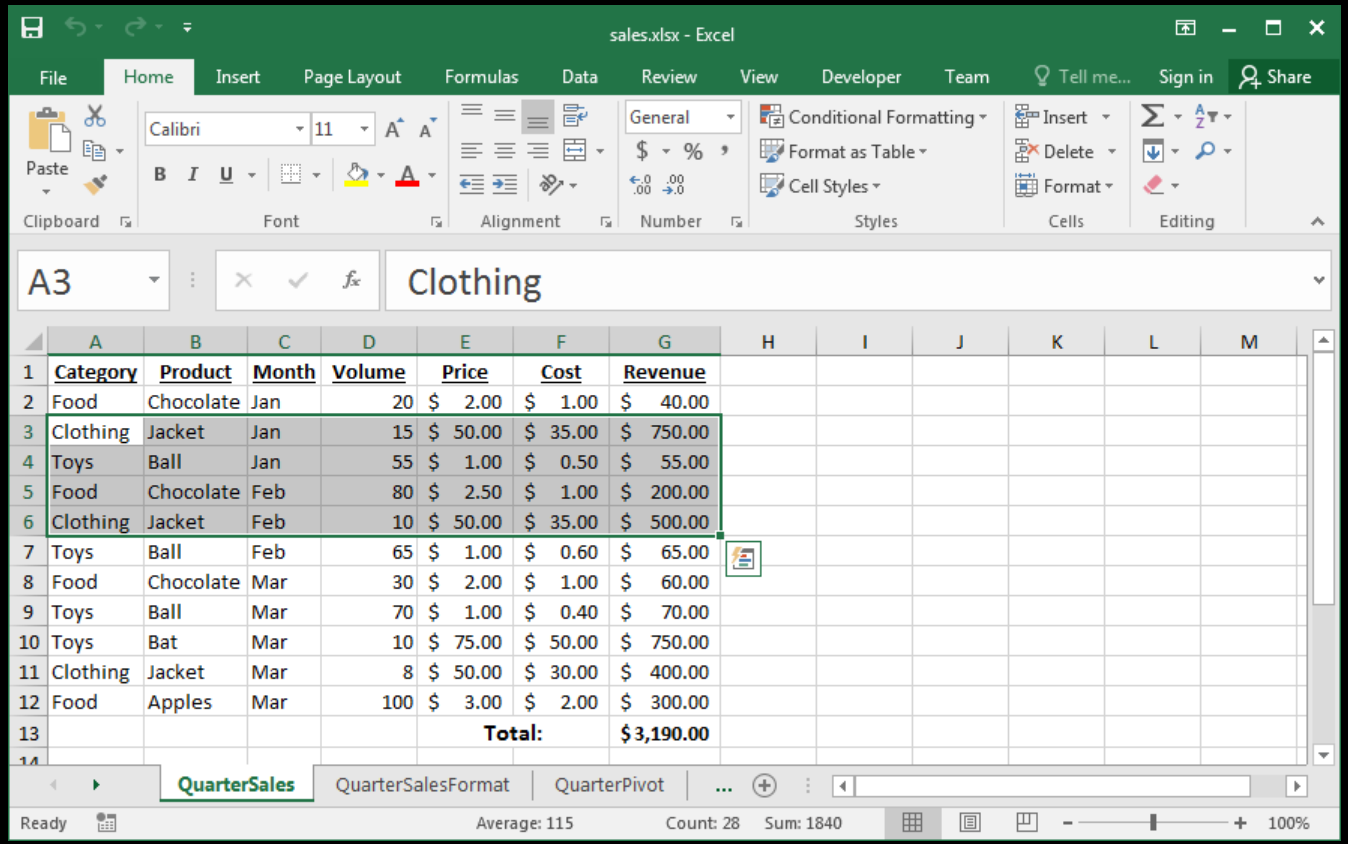
2) With keyboard, hold `SHIFT` key and use arrow keys to select a rectangle region of cells.

3) With mouse and keyboard, while holding `CTRL` key, (left) click on individual cells to select non-contiguous cells.

4) Click on a row number to select a whole row.

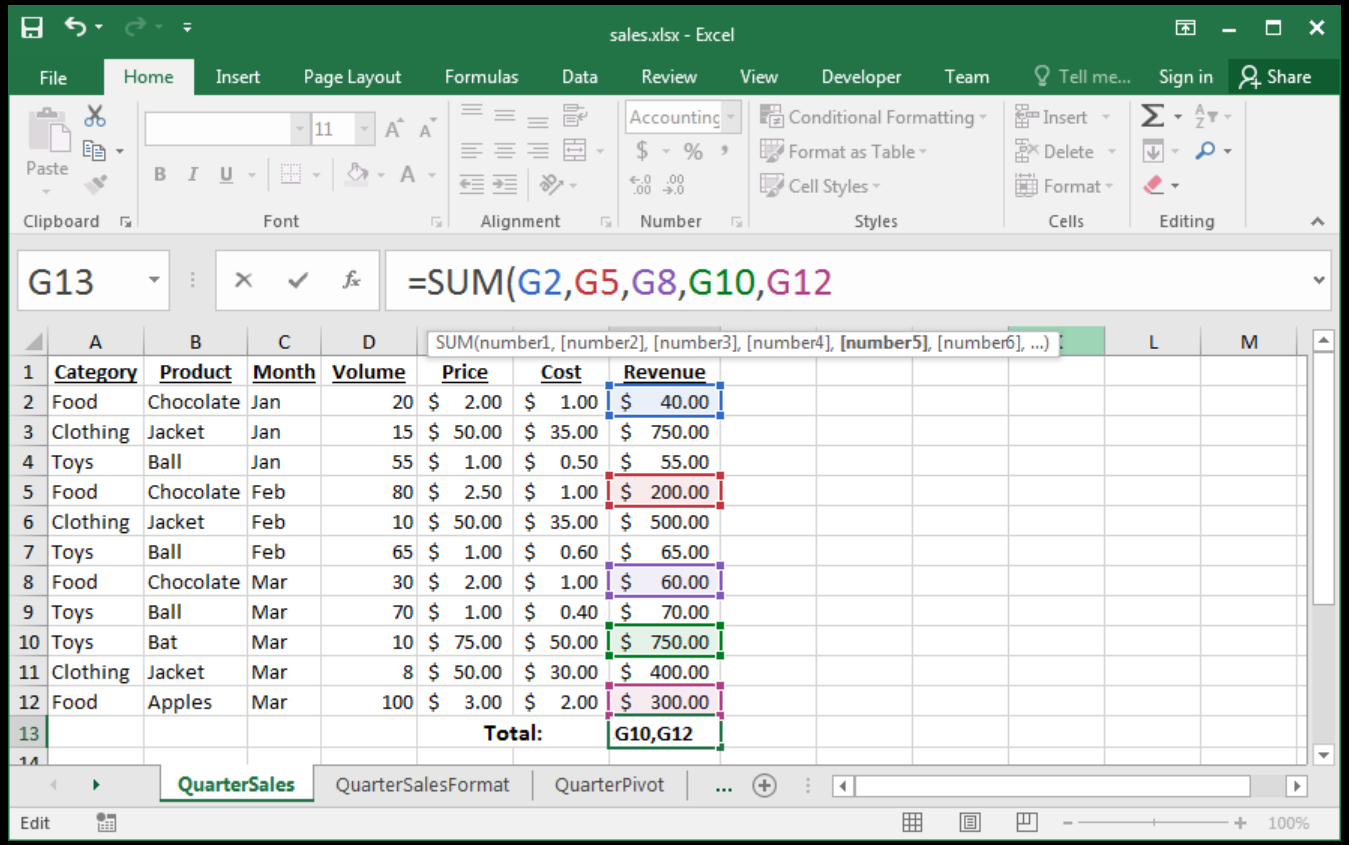
5) Click on a column header to select a whole column.

Range Selecting Cells Example



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00

Selecting Individual Cells Example



	A	B	C	D		E		F		G
1	Category	Product	Month	Volume	Price		Cost		Revenue	
2	Food	Chocolate	Jan	20	$	2.00	$	1.00	$	40.00
3	Clothing	Jacket	Jan	15	$	50.00	$	35.00	$	750.00
4	Toys	Ball	Jan	55	$	1.00	$	0.50	$	55.00
5	Food	Chocolate	Feb	80	$	2.50	$	1.00	$	200.00
6	Clothing	Jacket	Feb	10	$	50.00	$	35.00	$	500.00
7	Toys	Ball	Feb	65	$	1.00	$	0.60	$	65.00
8	Food	Chocolate	Mar	30	$	2.00	$	1.00	$	60.00
9	Toys	Ball	Mar	70	$	1.00	$	0.40	$	70.00
10	Toys	Bat	Mar	10	$	75.00	$	50.00	$	750.00
11	Clothing	Jacket	Mar	8	$	50.00	$	30.00	$	400.00
12	Food	Apples	Mar	100	$	3.00	$	2.00	$	300.00
13						Total:				G10,G12

Manipulating Cells

Once you have selected one or more cells, there are several common actions you can perform:

1) DELETE

- delete the contents of all cells by pressing delete key
- delete the contents and the cell locations (then shift remaining) by selecting `Edit` menu, `Delete...` or `Delete`… from pop-up menu (brought up by right click).

2) Cut, Copy, Paste

- cut - copies selected cells to clipboard and removes from document
- copy - copies selected cells to clipboard
- paste - copies cells in clipboard to sheet starting at currently selected cell

3) Add selected cells to a formula (requires that you were previously constructing a formula before selecting the cells).

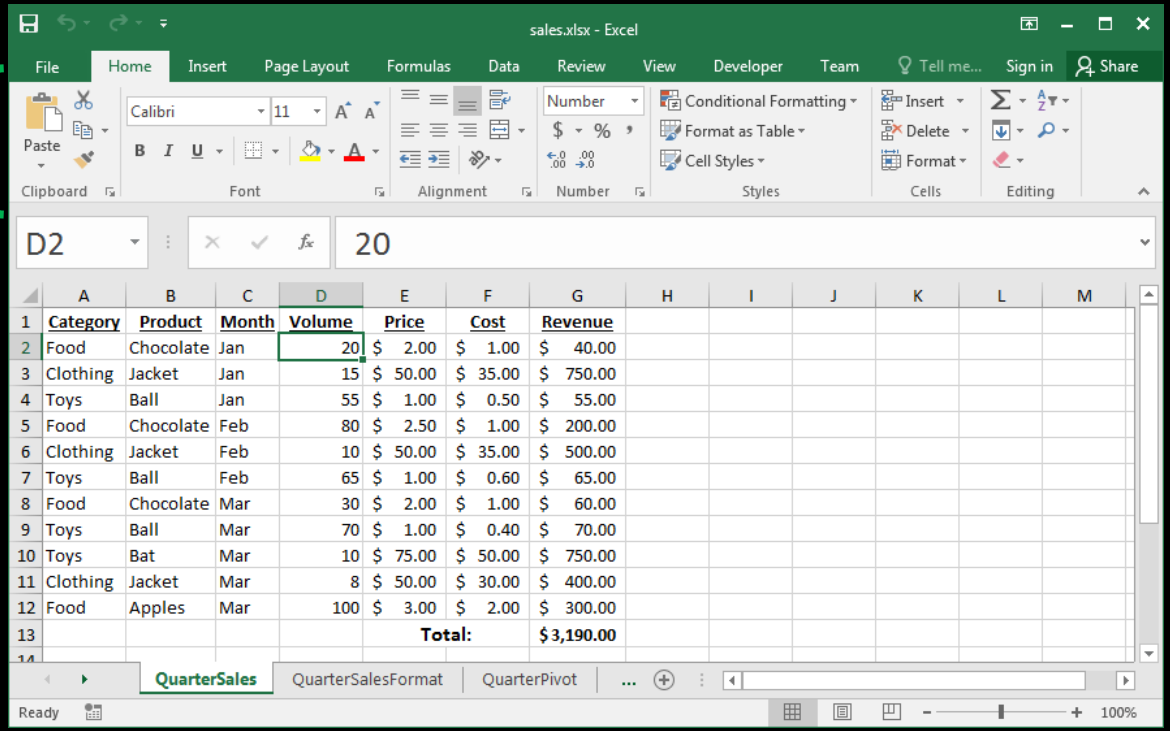
Manipulating Cells - Filling

Filling combines copy and paste.

There is a small box or tab beyond the cell's lower right corner (fill handle). Grab it with the cursor and pull to other cells.

Cut, Copy, Paste

cut,
copy,
paste

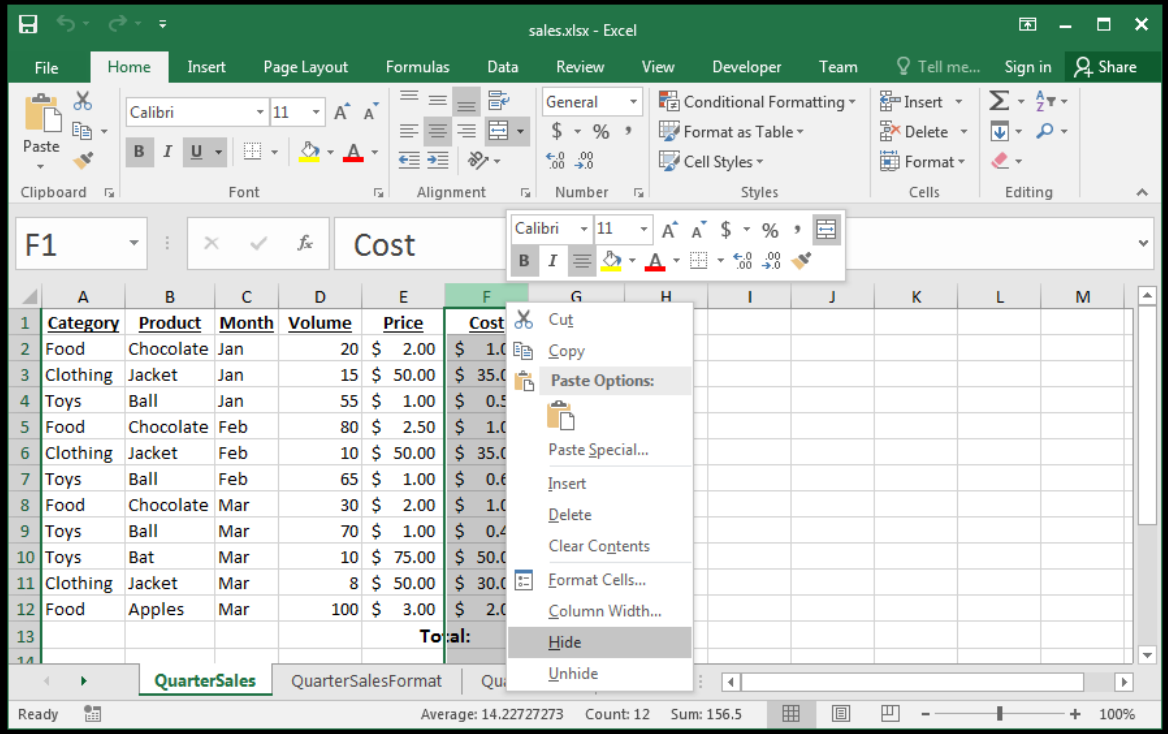


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	Product	Month	Volume	Price	Cost	Revenue						
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00						
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00						
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
13					Total:		$ 3,190.00						

Hiding Columns and Rows

Right-clicking on the column or row header and selecting **Hide**.

- The column/row still exists but will not be displayed or printed unless unhidden.



Selecting Cells Question

Question: Which method allows you to select non-contiguous cells in a spreadsheet?

A) hold `SHIFT` key and use arrow keys

B) With the mouse left click on a cell and drag mouse

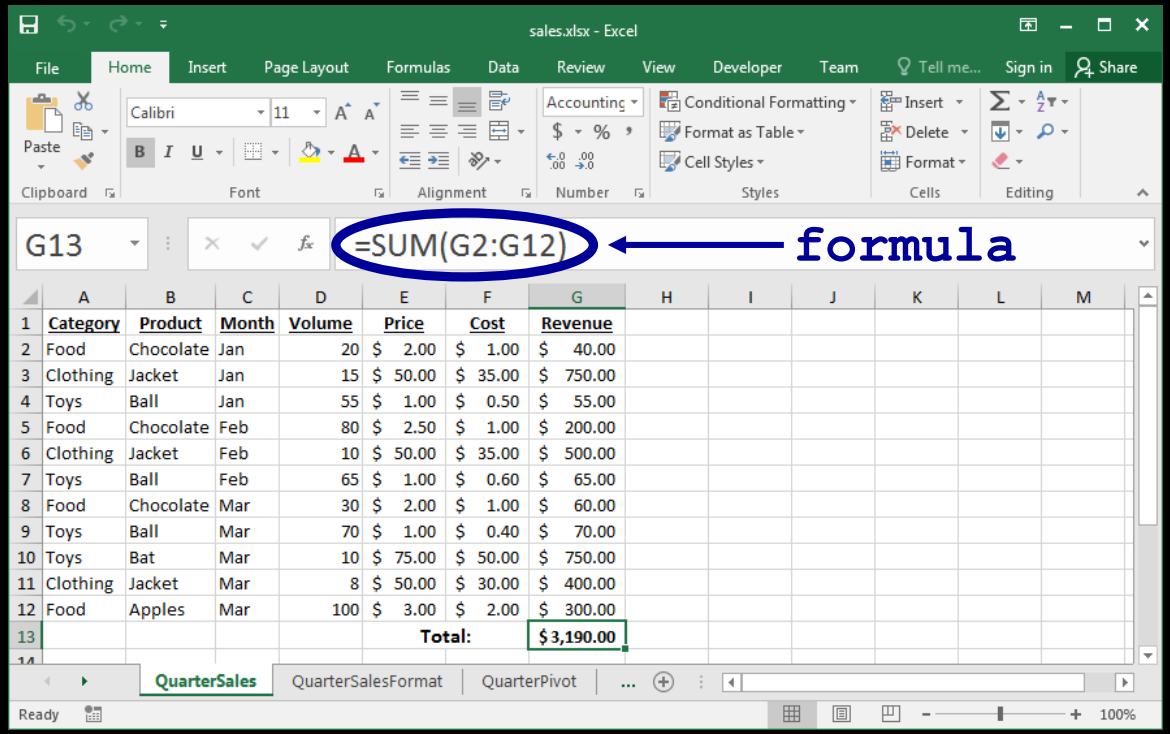
C) hold `CTRL` key and use arrow keys

D) hold `CTRL` key and left click on cells

Entering Formulas

A ***formula*** is any expression that begins with an equal sign ("=").

- The equal sign means that a calculation must be done to compute the cell value.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	Product	Month	Volume	Price	Cost	Revenue						
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00						
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00						
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
13						Total:	$ 3,190.00						

Formula Expressions

A ***formula*** expression can consist of literals (numbers, text strings), operators, functions, and cell references.

Simple mathematical expressions:

- `= 1 + 5`
- `= 1.5 * 3.14 + 42`

Common functions:

- `= ROUND(PI(),2)` // Result is 3.14
- `= CONCATENATE("Hello", " World")` // Hello World
- Other common functions for trigonometry, dates, and financial.

Formula Expressions

The power of formulas comes from using cell references (similar to variable names in programming).

Cell reference examples:

- `= A1 + A2`
- `= B1 + A3 - A4`

Formulas Question

Question: A cell contains the following: **=2+4*3** What is the value of the cell?

A) 14

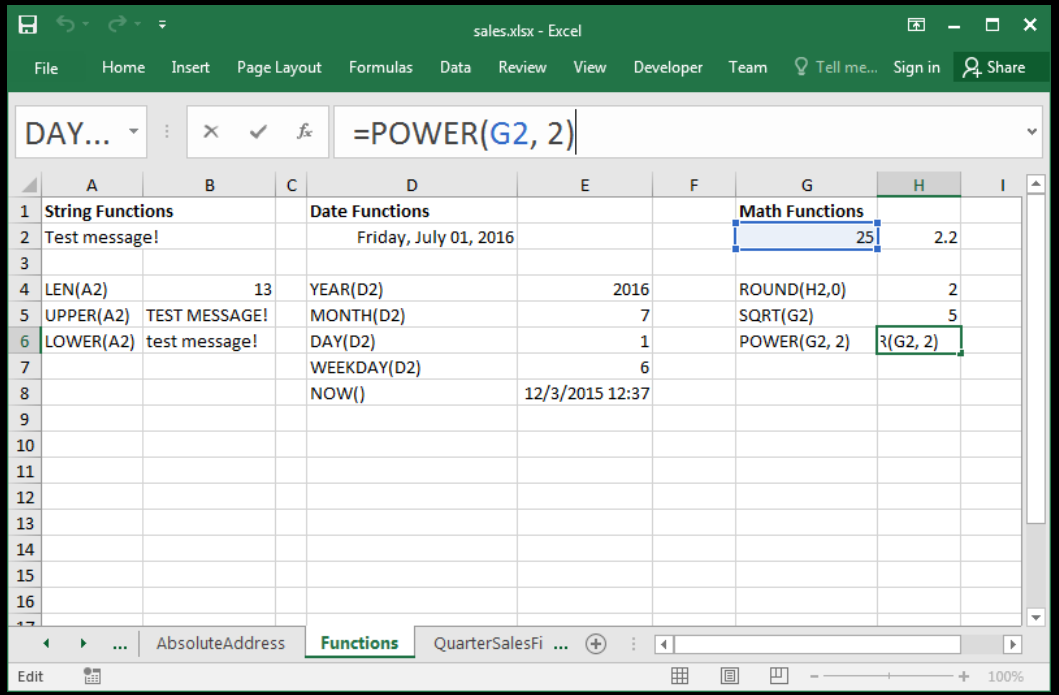
B) 18

C) =2+4*3

Using Excel Functions

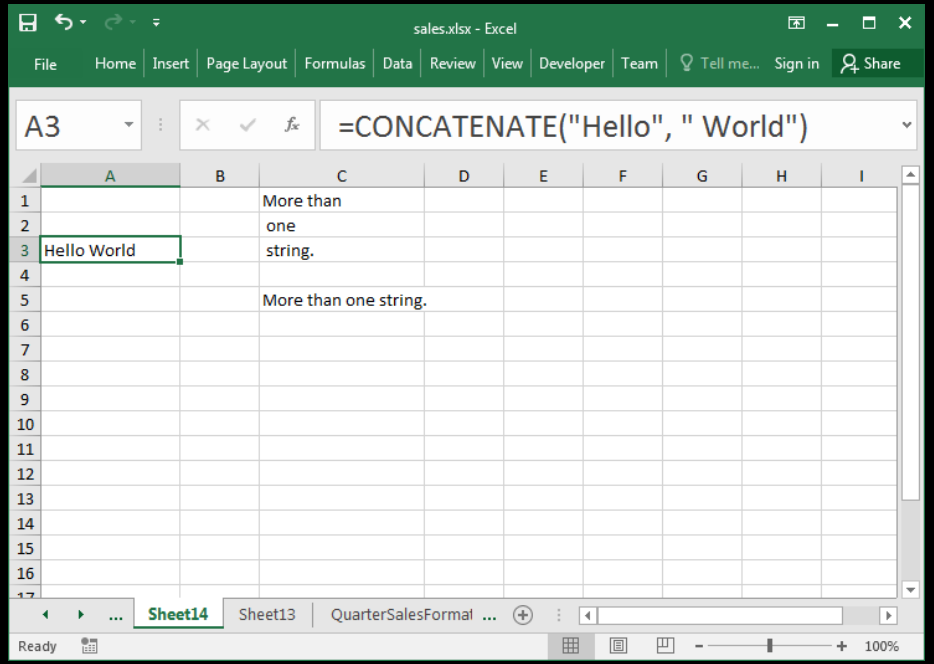
Excel has a large number of built-in functions to use.

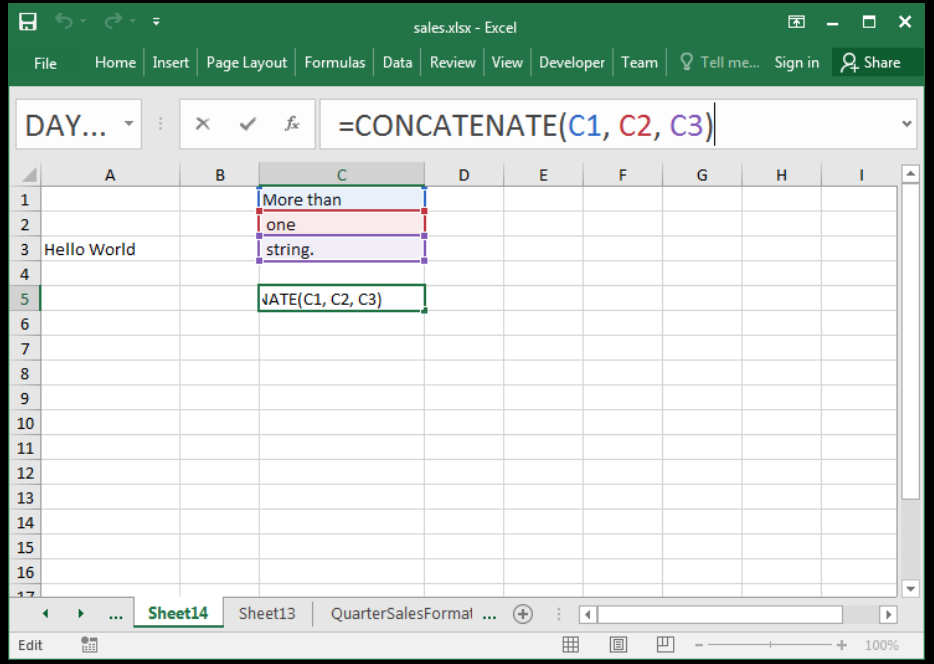
A ***function*** takes arguments as input and produces an output.



Concatenation

String concatenation is when two or more strings are combined by appending them in order. Function in Excel is `CONCATENATE()` or `&`.

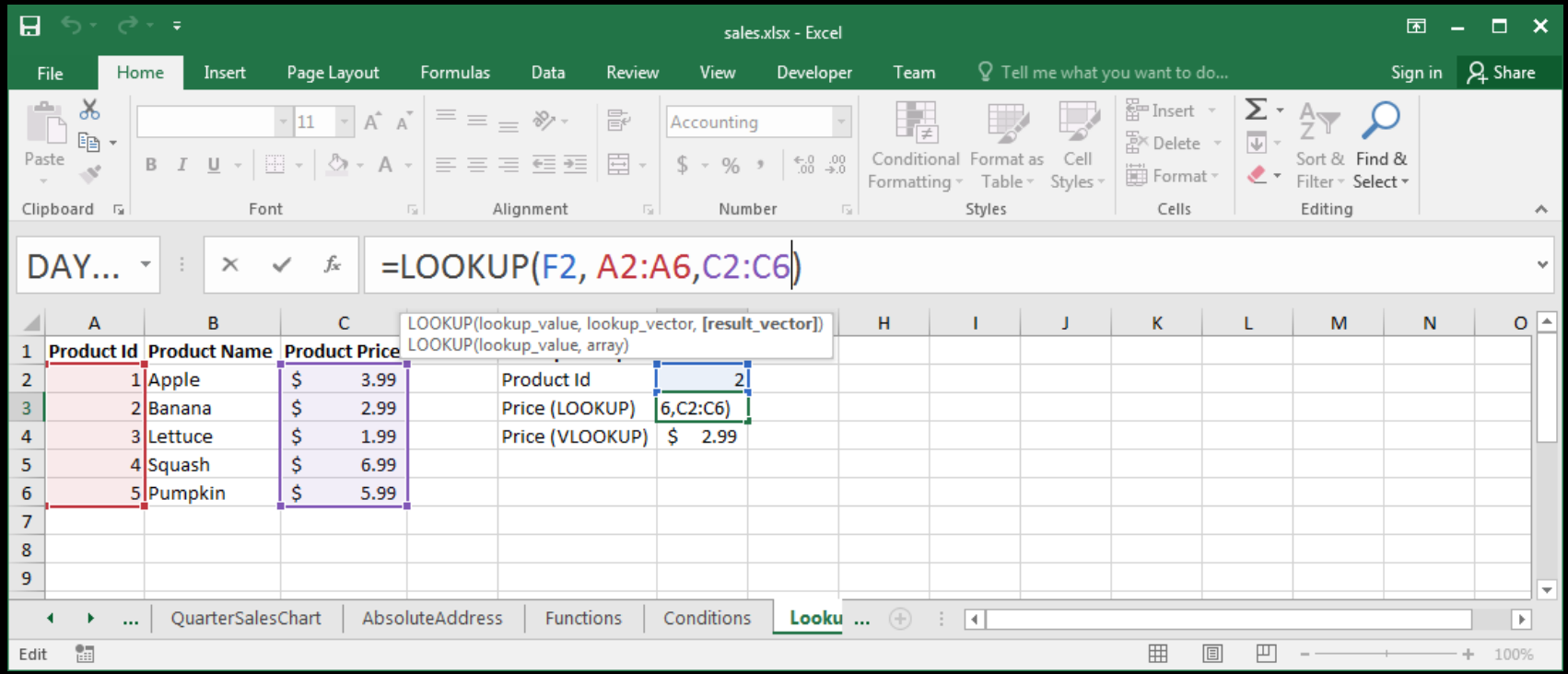




LOOKUP Function

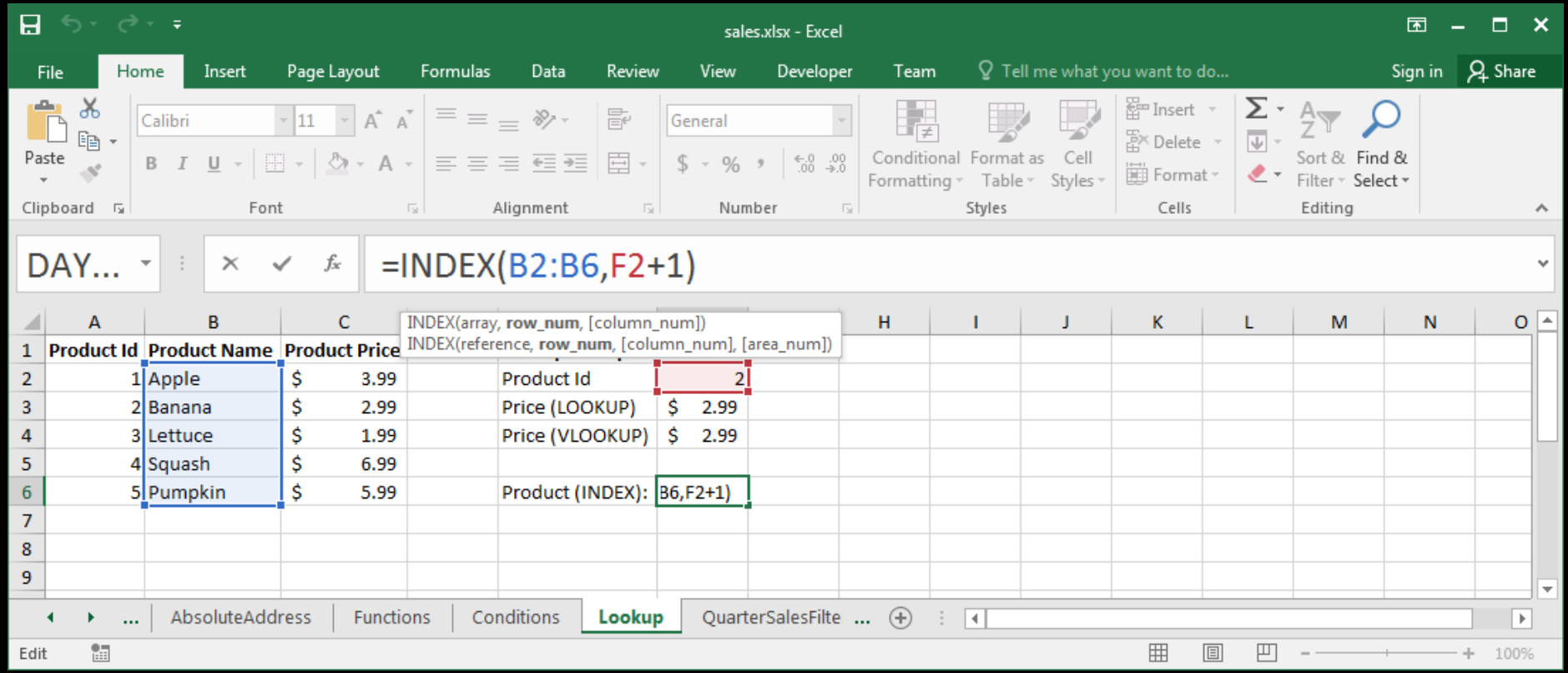
The LOOKUP function searches for a value in a column.

- VLOOKUP searches a column in a table ; HLOOKUP searches a row in a table.



INDEX Function

INDEX() returns the value in the array of cells at the given index.



Formulas Question

Question: A cell contains the following: **'ABC'+'DEF'**. What is the value of the cell?

A) error

B) ABCDEF

C) 'ABC'+'DEF'

Formulas Question

Question: How many of the following statements are **TRUE**?

1) `CONCATENATE` function can take 3 arguments.

2) There is an Excel function that has 0 arguments.

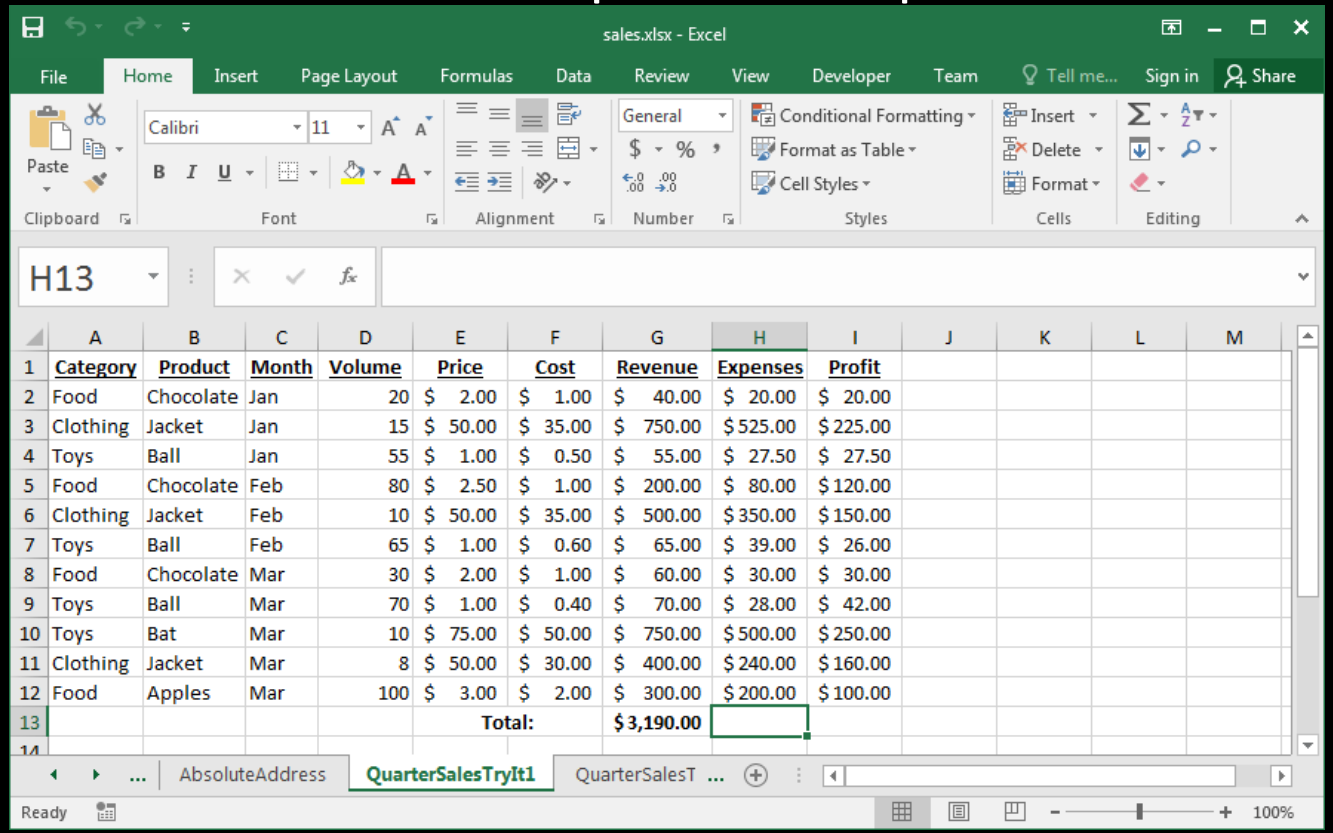
3) `=INDEX({1,3,5},2)` returns `5`.

4) `=LOOKUP(5,{1,3,5},{"a","b","c"})` returns `"c"`.

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Try it: Entering Formulas

Question: Add a column for expenses and profit as below:



	A	B	C	D	E	F	G	H	I
1	Category	Product	Month	Volume	Price	Cost	Revenue	Expenses	Profit
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00	$ 20.00	$ 20.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00	$ 525.00	$ 225.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00	$ 27.50	$ 27.50
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00	$ 80.00	$ 120.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00	$ 350.00	$ 150.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00	$ 39.00	$ 26.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00	$ 30.00	$ 30.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00	$ 28.00	$ 42.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00	$ 500.00	$ 250.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00	$ 240.00	$ 160.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00	$ 200.00	$ 100.00
13					Total:		$ 3,190.00		

Advanced Spreadsheet Addressing

The dollar sign "**$**" is a symbol that indicates an ***absolute address***.

- By default, addresses are "relative" in the sense that if they are in a formula that is copied to another cell, they will be changed relative to where they were copied from their origin.

Example:

- Cell `A1` has the formula `=A2+B1`
- Copy contents of cell `A1` to cell `C4`.
- Formula changes to `=C5+D4` because moved down three rows and over two columns.
- If cell `A1` had the formula `=$A$2+$B$1`, then the same formula would be in cell `C4`.
- Question: What if formula was `=$A2+B$1`?

Formulas and References Question

Question: Cell **A1** contains the following: **=$B2+D$4**. What is the formula if the cell is copied to cell **D3**?

A) error

B) =$B2+D$4

C) =$B4+F$4

D) =$B4+G$4

Aggregate Functions

An ***aggregate function*** computes a summary function over a range of cells. The values can either be data values or cell locations.

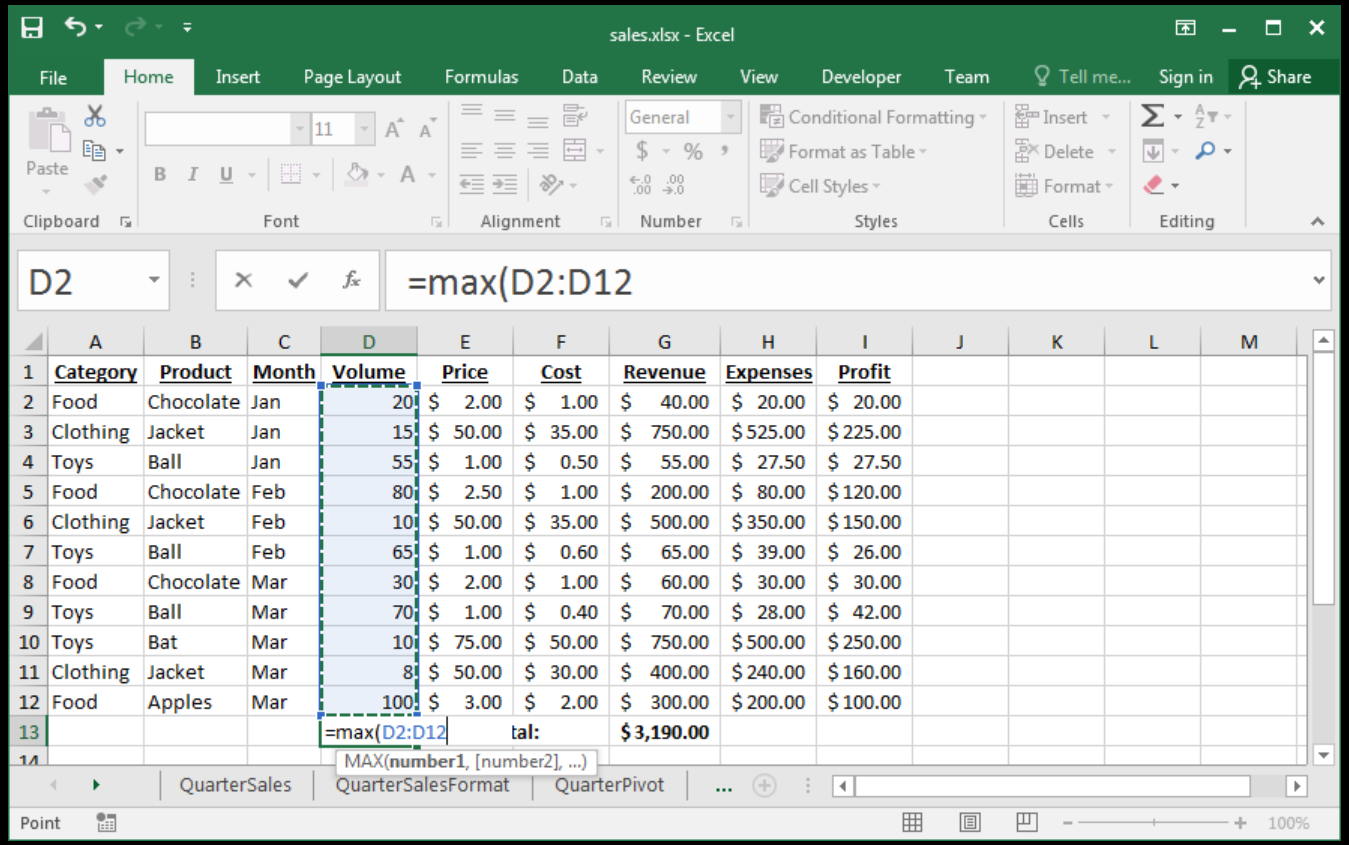
Common functions are:

- `MIN(<value list>)` - returns minimum value in list
- `MAX(<value list>)` - returns maximum value in list
- `SUM(<value list>)` - returns sum of all values in list
- `AVERAGE(<value list>)` - returns average of values in list
- `COUNT(<value list>)` - returns count of values in list
- `MEDIAN(<value list>)` - returns median value of list

If specifying a cell rectangle, give the upper left and lower right corners, separated by a colon.

- e.g. `=AVERAGE(A3:E6)` - rectangle of 4 rows and 5 columns

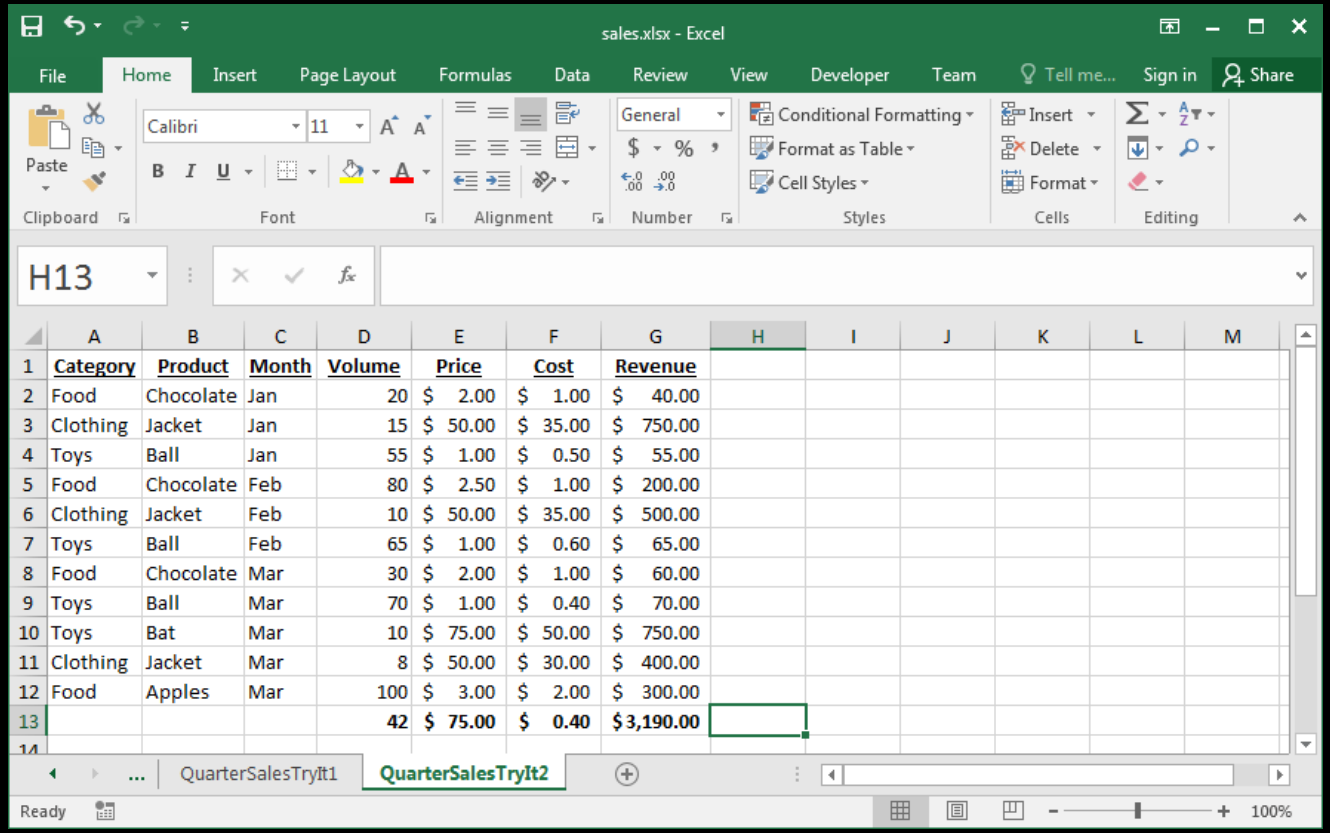
Aggregate Functions Example



	A	B	C	D	E	F	G	H	I
1	Category	Product	Month	Volume	Price	Cost	Revenue	Expenses	Profit
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00	$ 20.00	$ 20.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00	$ 525.00	$ 225.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00	$ 27.50	$ 27.50
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00	$ 80.00	$ 120.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00	$ 350.00	$ 150.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00	$ 39.00	$ 26.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00	$ 30.00	$ 30.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00	$ 28.00	$ 42.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00	$ 500.00	$ 250.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00	$ 240.00	$ 160.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00	$ 200.00	$ 100.00
13				=max(D2:D12		tal:	$ 3,190.00		
14									

Try it: Aggregate Functions

Question: Create aggregate functions to match below:



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Category**	**Product**	**Month**	**Volume**	**Price**	**Cost**	**Revenue**						
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00						
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00						
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
13				**42**	**$ 75.00**	**$ 0.40**	**$ 3,190.00**						

Aggregate Functions Question

Question: Assume the cells in the range `A1:C4` each contain a number that is equal to their row number (e.g. `B3` contains `3`). How many of the following statements are **TRUE**?

1) The number of cells in the range is `12`.
2) The value of `SUM(A1:C4)` is `20`.
3) The value of `COUNTIF(A1:B4,">2")` is `4`.
4) `AVERAGE(A1:C4) > MAX(C2:C3)`

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Aggregate Functions Question

Question: Assume the three cells in the range `A1:C1` contain numbers. Which of these formula output results is **ALWAYS** the largest?

A) `MAX(A1:C1)`

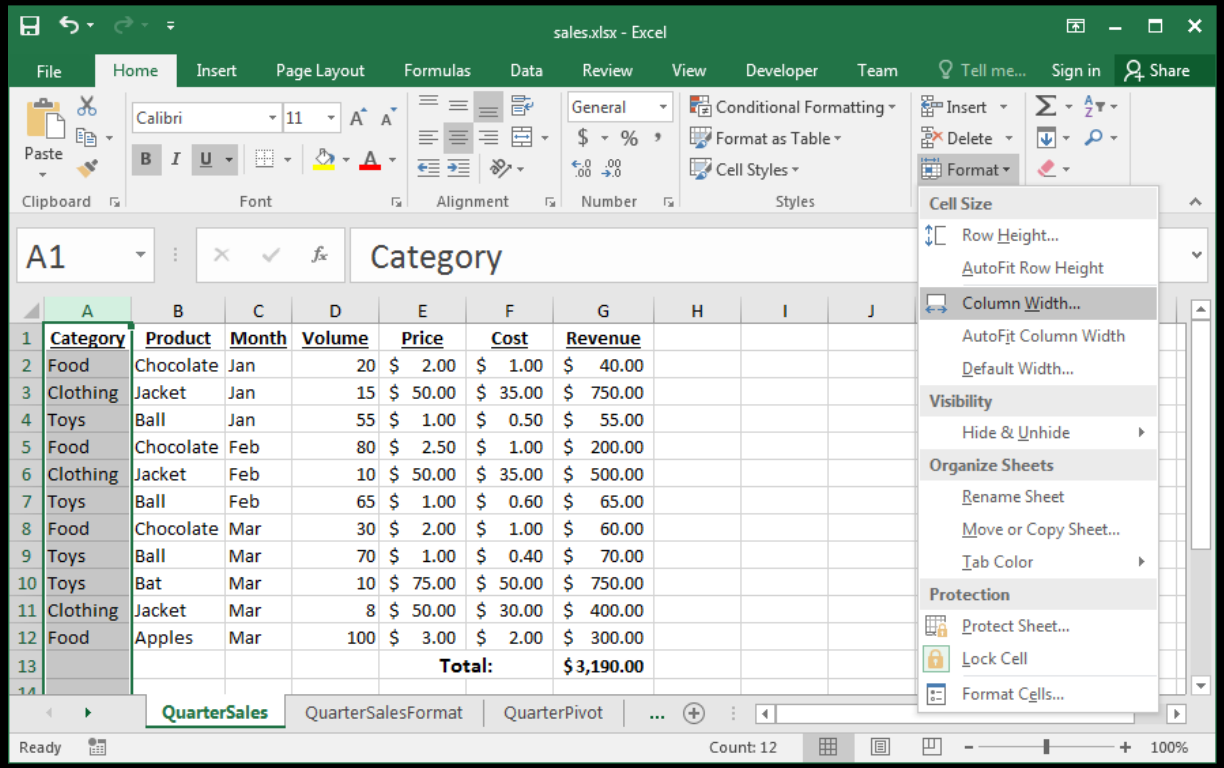
B) `MIN(A1:C1)`

C) `COUNT(A1:C1)`

D) `SUM(A1:C1)`

E) none of the above are always guaranteed to be the largest

Other Formatting: Column Width



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00

Resizing columns/rows:

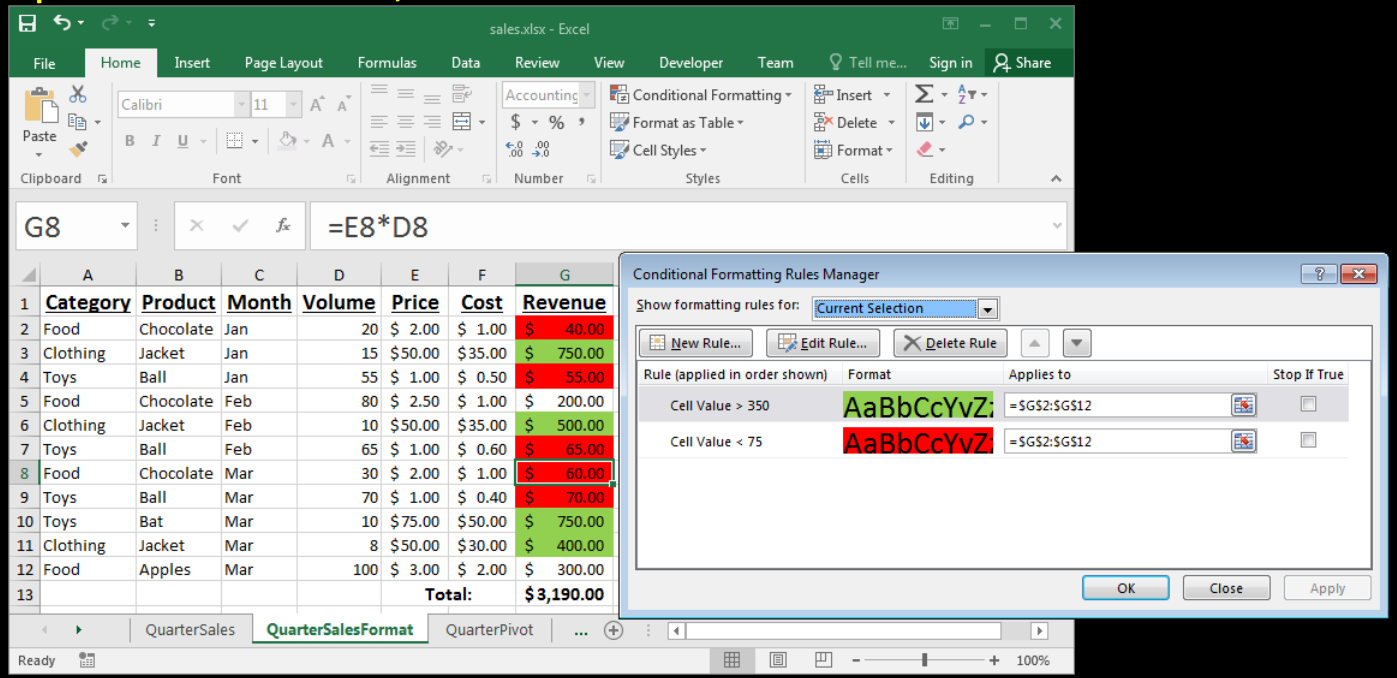
Auto-resize by double clicking on border between columns or using the `Format` option.

Drag row/column border for manual resize.

Conditional Formatting

Conditional formatting allows you to change the cell format based on data values. This is accessible under **`Styles`**.

- Other options: data bars, color scales

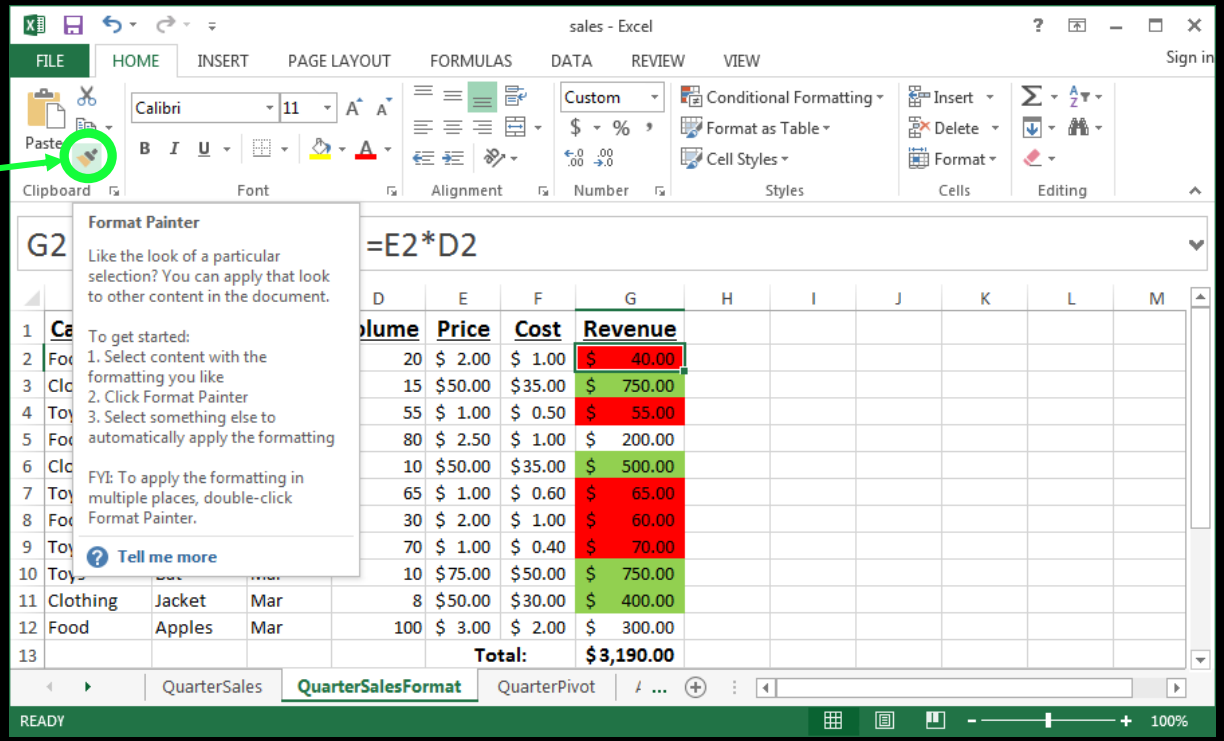


	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$75.00	$50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$50.00	$30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13						Total:	$3,190.00

Conditional Formatting Result

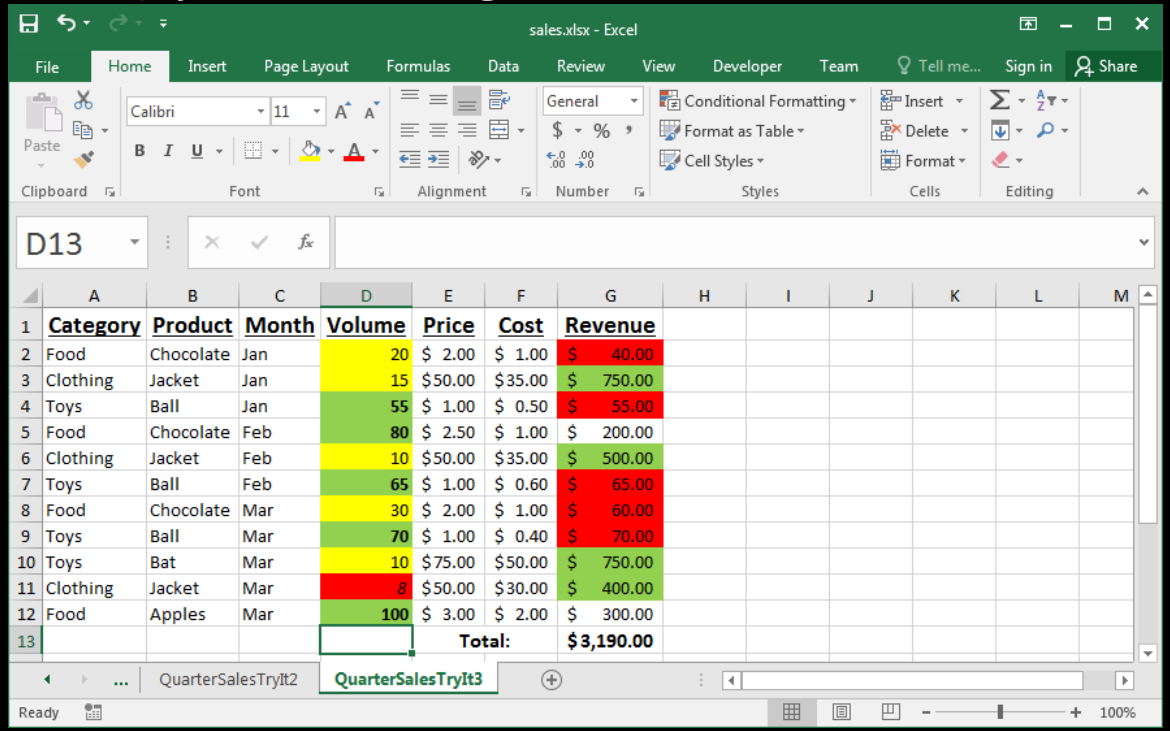
The format painter button allows you to copy formatting to many cells. Select the cell, click paint button, then highlight cells to have identical formatting.

format painter button



Try it: Conditional Formatting

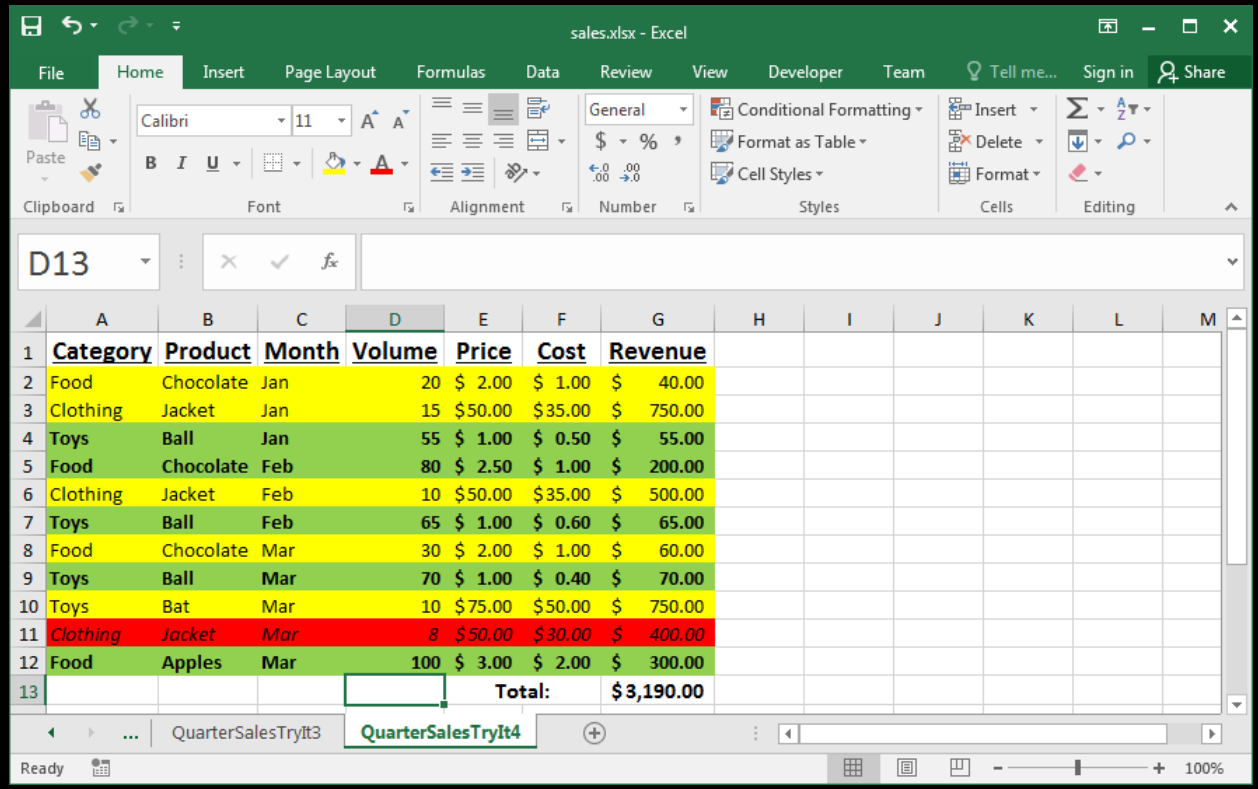
Question: Format rows so: 1) bold/green if volume > 50, 2) italics/red if volume < 10, 3) yellow background otherwise as below:



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Category**	**Product**	**Month**	**Volume**	**Price**	**Cost**	**Revenue**						
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$ 750.00						
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$ 500.00						
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00						
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00						
10	Toys	Bat	Mar	10	$75.00	$50.00	$ 750.00						
11	Clothing	Jacket	Mar	*8*	$50.00	$30.00	$ 400.00						
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
13						**Total:**	**$3,190.00**						

Try it: Conditional Formatting Challenge

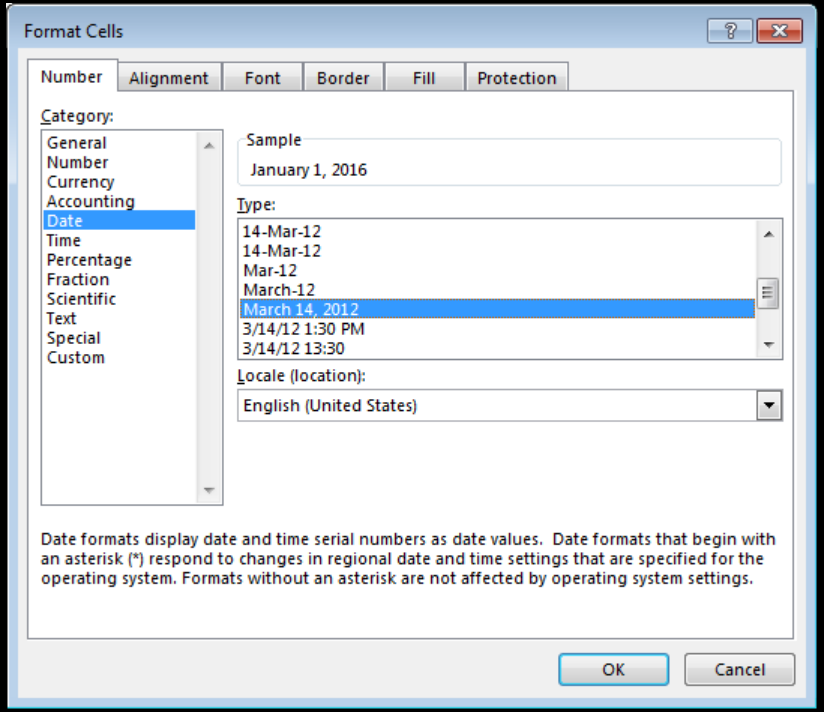
Question: Take the previous formatting and apply it to whole row:



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$50.00	$35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$50.00	$35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$75.00	$50.00	$ 750.00
11	*Clothing*	*Jacket*	*Mar*	*8*	*$50.00*	*$30.00*	*$ 400.00*
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$3,190.00

Date and Type Formats

Formatting data helps users read and understand data and is especially important for numbers and dates. Use built-in or custom formats.



Spreadsheets for Data Management

A spreadsheet is often used as a "database". A database is an organized representation of information.

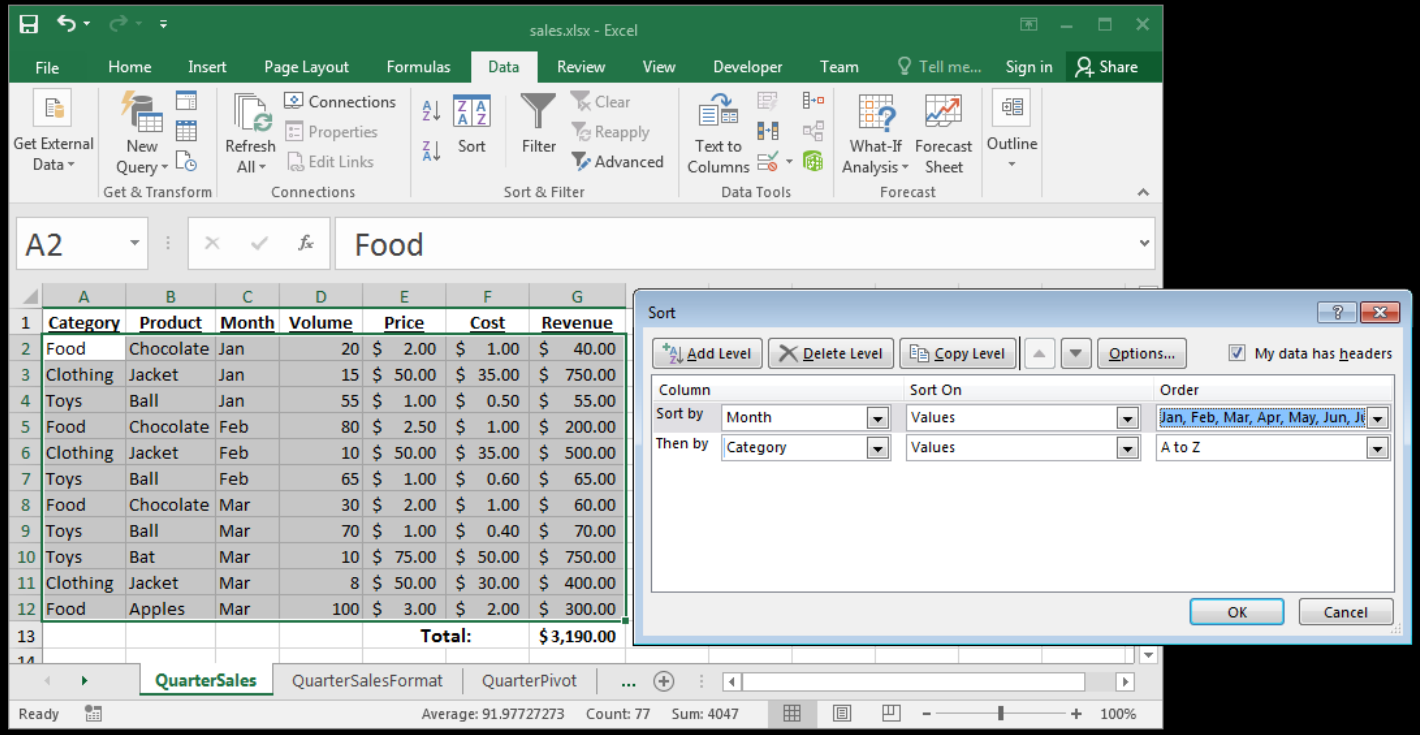
- Examples: schedules and calendars, timesheets, expenses and finances, records, notes, and recipes, data research/analysis

We can use a spreadsheet as a database by:

- Using a row to store all the information about something we want to represent.
- Giving each column a meaningful name. A column represents a property or feature of the object stored in the row.
- Using the formulas to calculate new facts from the data.
- Using sorting to organize the data by key features.
- Using simple filtering (querying) to only show the most important data or data of interest.

Sorting Data

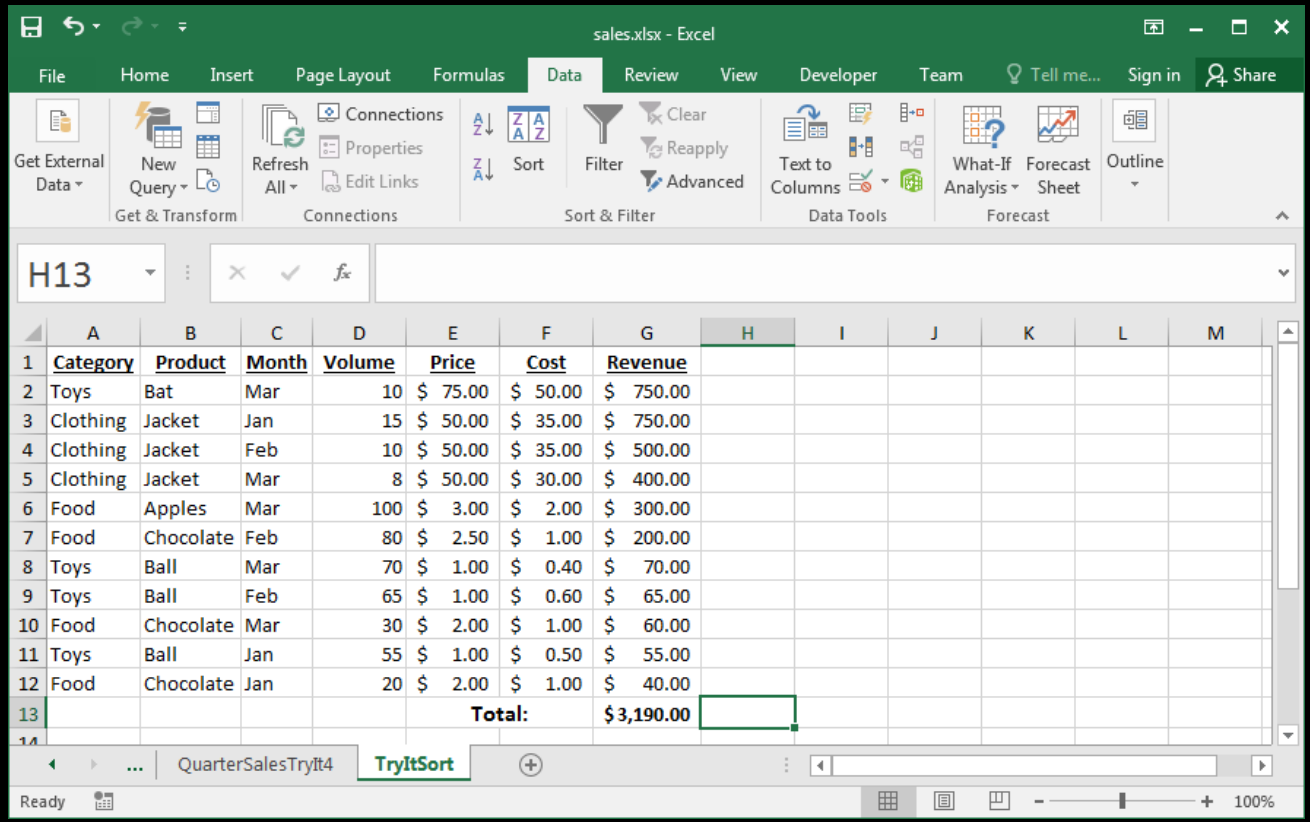
Data can be sorted by selecting the **Sort** option under the **Data** menu. Select the column(s) to sort on and order to sort by.



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$3,190.00

Try it: Sort

Question: Sort the data by revenue (desc) then product (asc).



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	Product	Month	Volume	Price	Cost	Revenue						
2	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
4	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
5	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
6	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
7	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
8	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00						
9	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00						
10	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
11	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
12	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
13						Total:	$ 3,190.00						

Filtering

A ***filter*** shows a subset of the rows in the spreadsheet that pass a given condition (test).

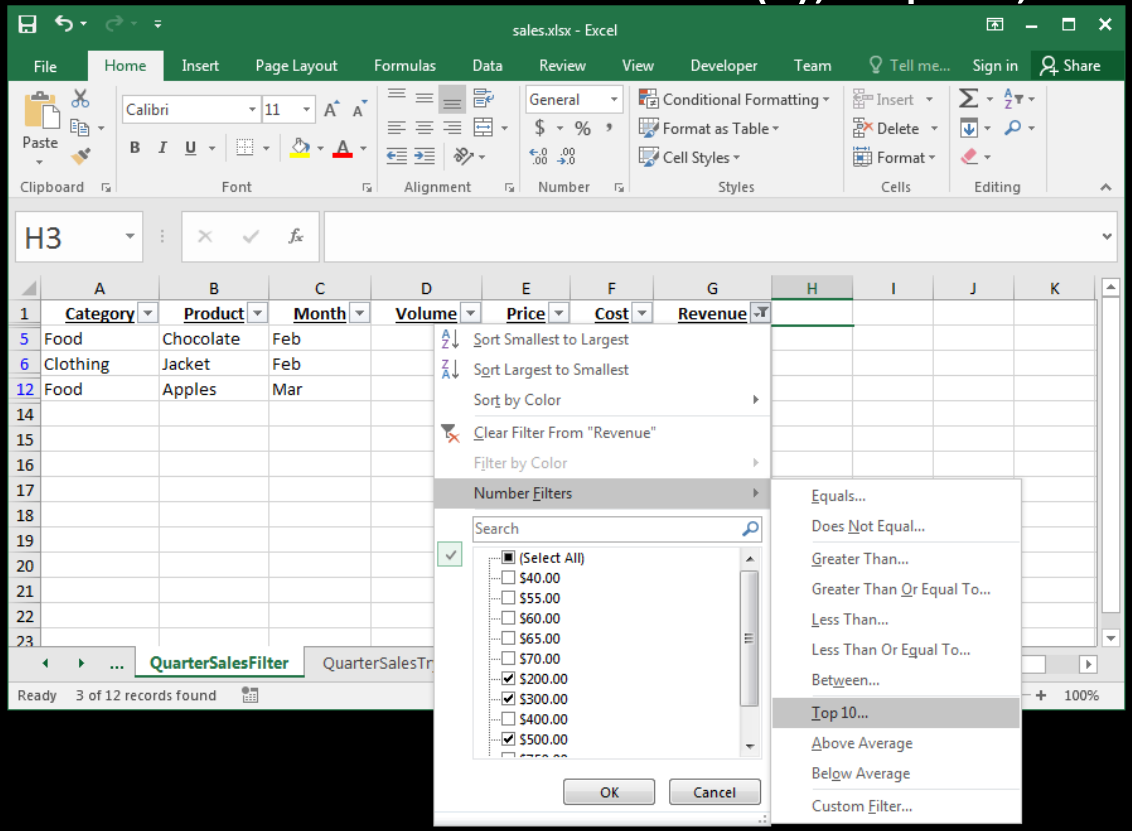
Select **`Auto Filter`** under the **`Data`** then **`Filter`** menu.

Once you select `Auto Filter`, each column heading has a drop-down list. By selecting a filtering criteria from the list, you can limit the rows that are displayed.

It is possible to filter on more than one column at the same time.

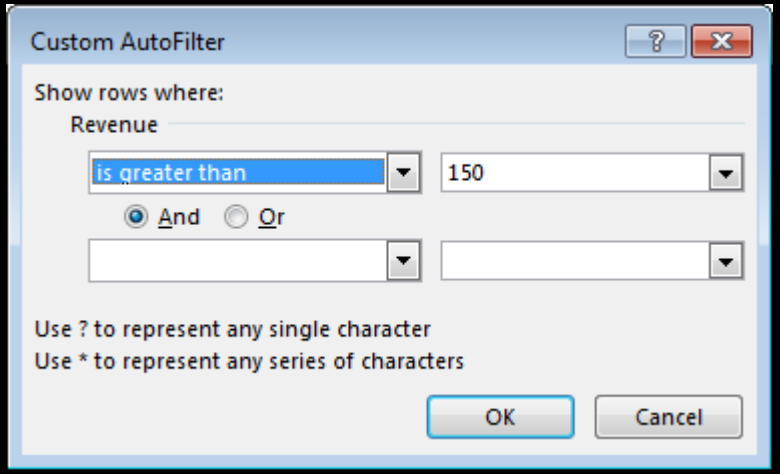
Filter Example

Filter on `Revenue` column: Select value(s), Top 10, or custom filter.



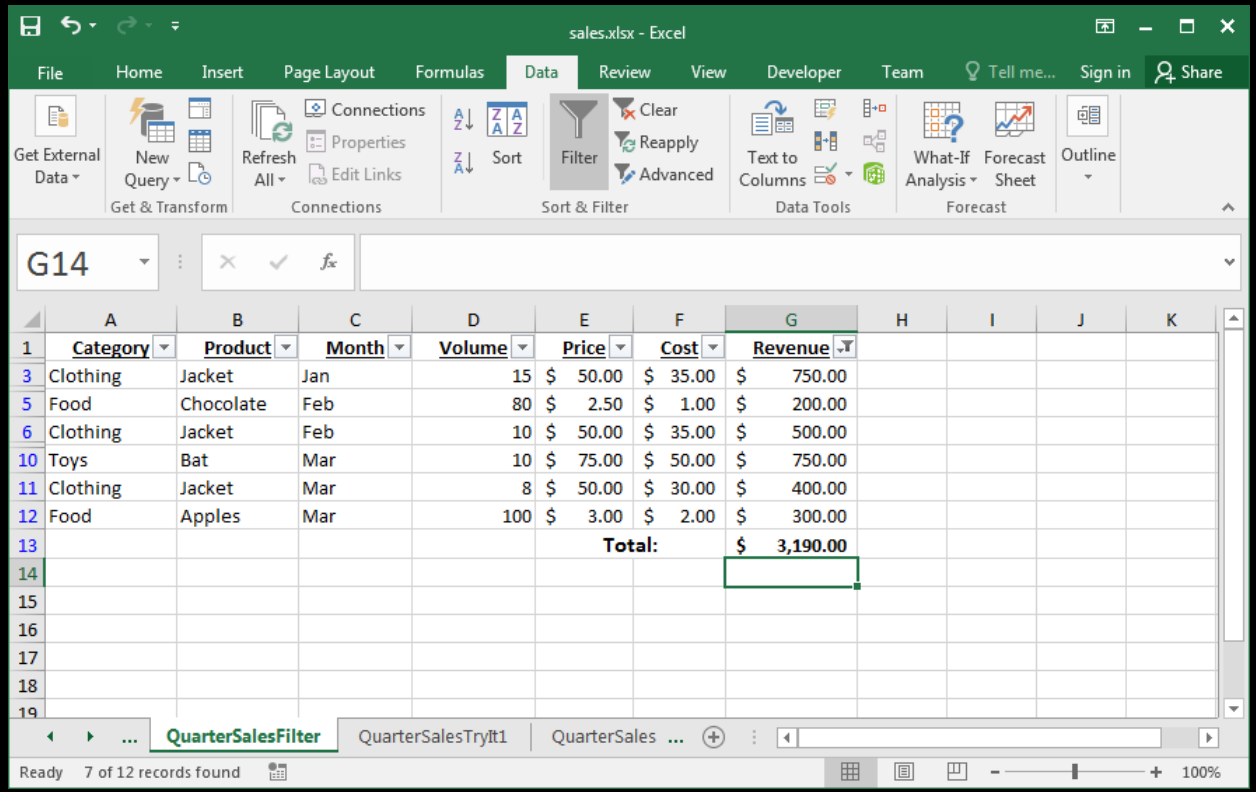
Custom Filter Example

Filter on `Revenue` column: Custom filter with **`Revenue > 150`**



Custom Filter Result

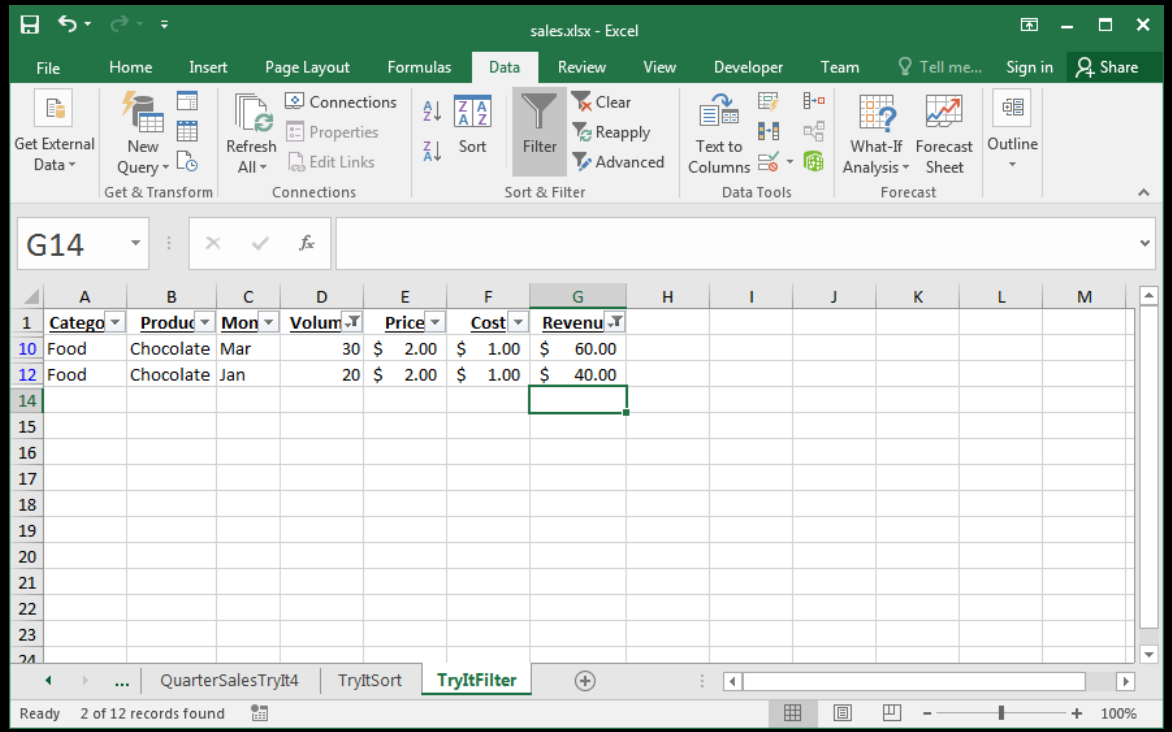
Filter on `Revenue`: Custom filter result with **Revenue > 150**



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00
14							
15							
16							
17							
18							
19							

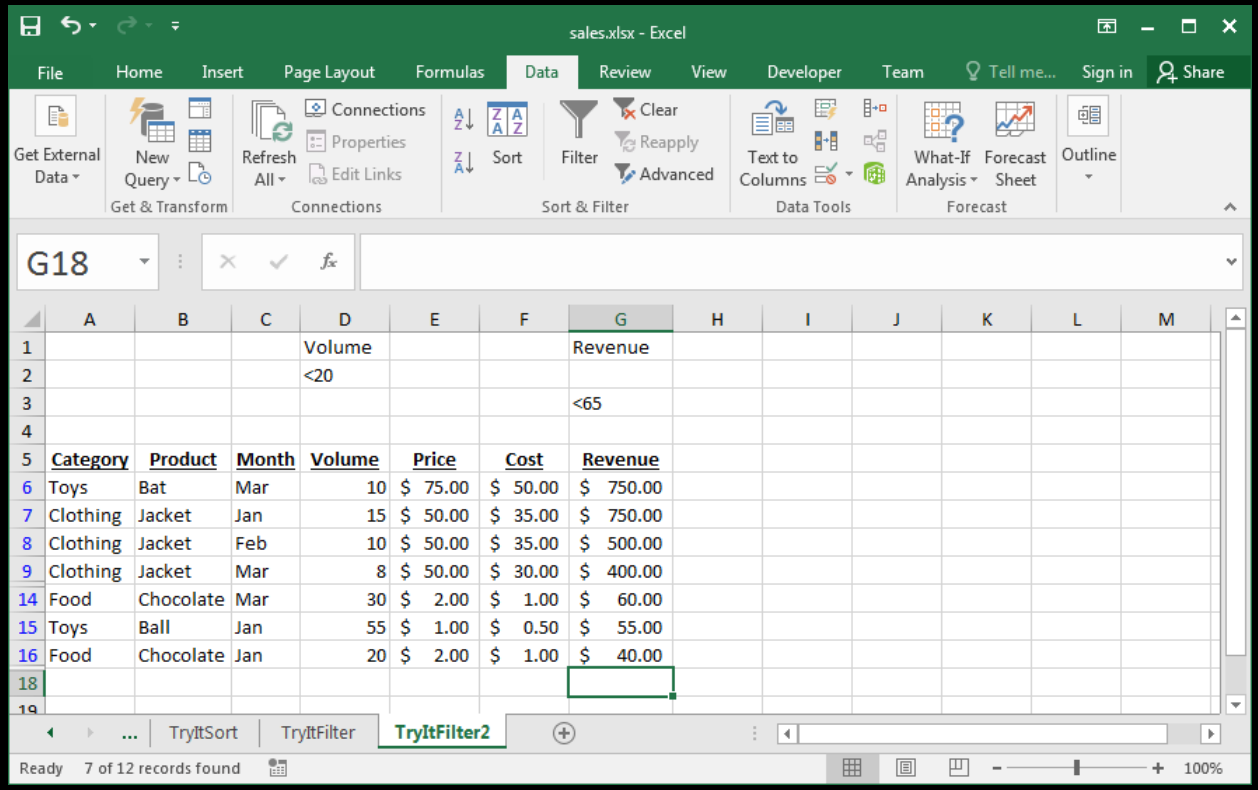
Try it: Filter

Question: Filter the data so only products with volume < 50 and revenue < $100 are shown.



Try it: Filter Challenge

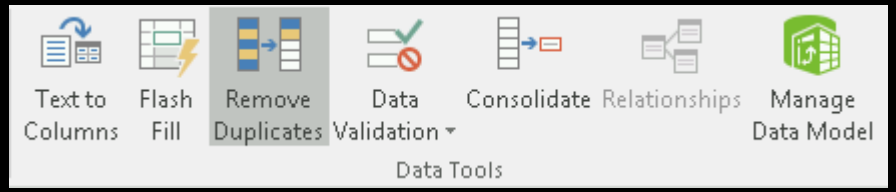
Question: Filter the data so only products with volume < 20 ***or*** revenue < $65 are shown.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1				Volume			Revenue						
2				<20									
3							<65						
4													
5	**Category**	**Product**	**Month**	**Volume**	**Price**	**Cost**	**Revenue**						
6	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
7	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
8	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
9	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
14	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
15	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00						
16	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
18													
19													

Removing Duplicates

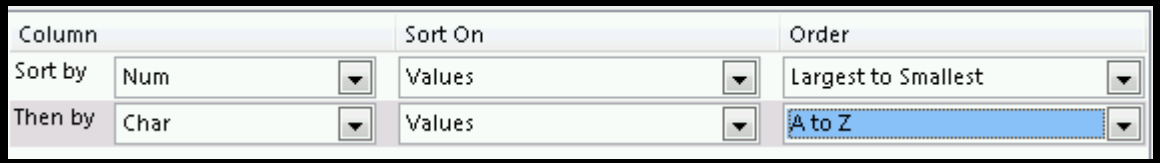
To remove duplicates, select your Data then `Remove Duplicates`.



Sorting Question

Question: Given this spreadsheet and sort order, what is the output?

Num	Char
1	A
1	a
1	B
1	b
2	A
2	b
3	a
3	B



A)

Num	Char
3	B
3	a
2	A
2	b
1	A
1	B
1	a
1	b

B)

Num	Char
3	a
3	B
2	A
2	b
1	a
1	A
1	b
1	B

C)

Num	Char
3	a
3	B
2	A
2	b
1	A
1	a
1	B
1	b

Filtering Question

Question: Given this spreadsheet, how many of these statements are **TRUE**?

	A	B
1	Number	Letter
2	1	a
3	2	b
4	3	c
5	4	d
6	5	e
7		

1) The data is sorted ascending by `Number`.

2) Filter `Number > 3` shows 3 rows.

3) Filter `Letter >= "c"` shows 3 rows.

4) Filter `Number < 3 OR Letter > "b"` shows 5 rows.

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Charts

A ***chart*** is a graphical representation of spreadsheet data.

A chart is of a particular type (line, bar, etc.) and requires the user to supply the data that will be displayed in the chart.

Chart: Select Data and Type

Select **Insert**, then click **Chart** Icon, and pick the chart type.

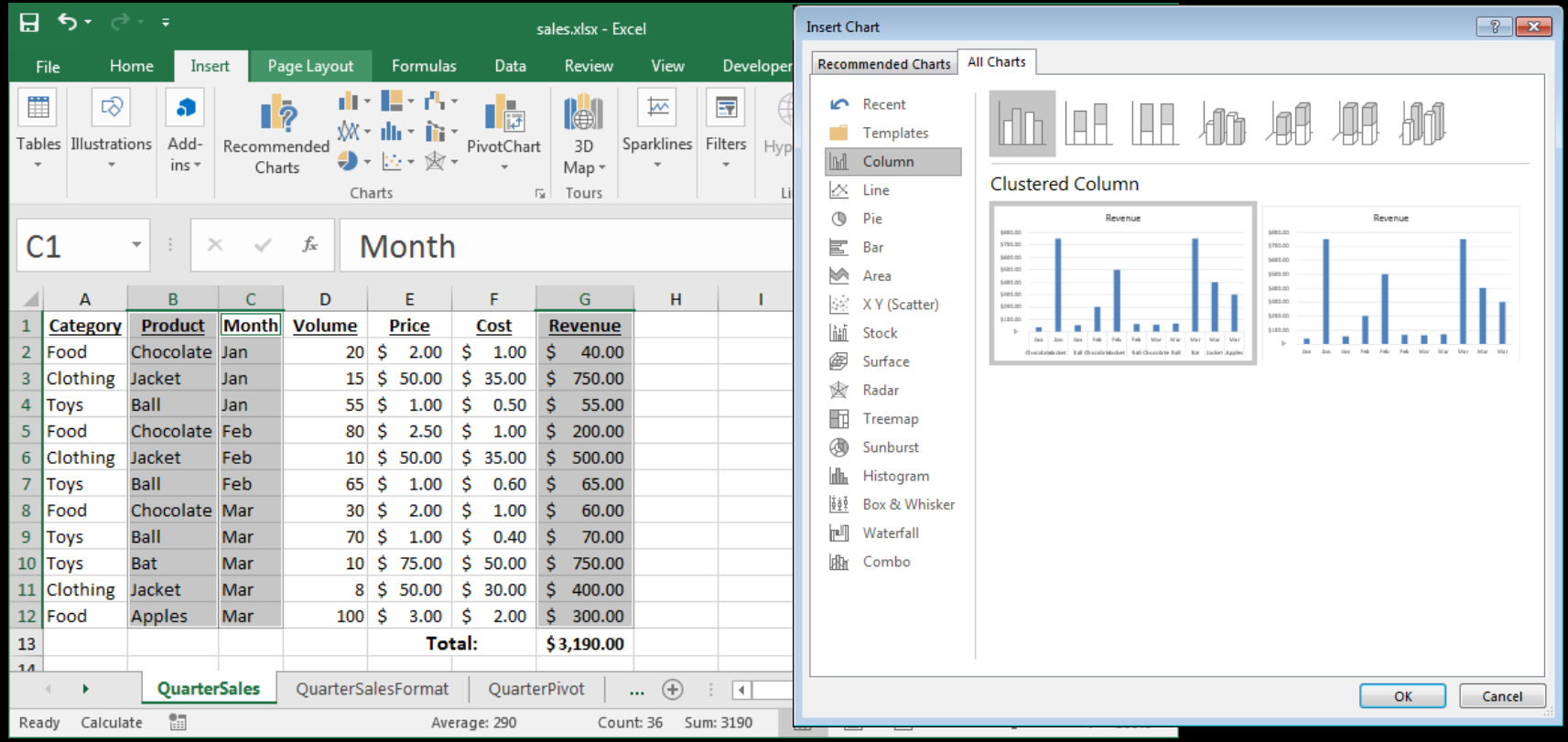
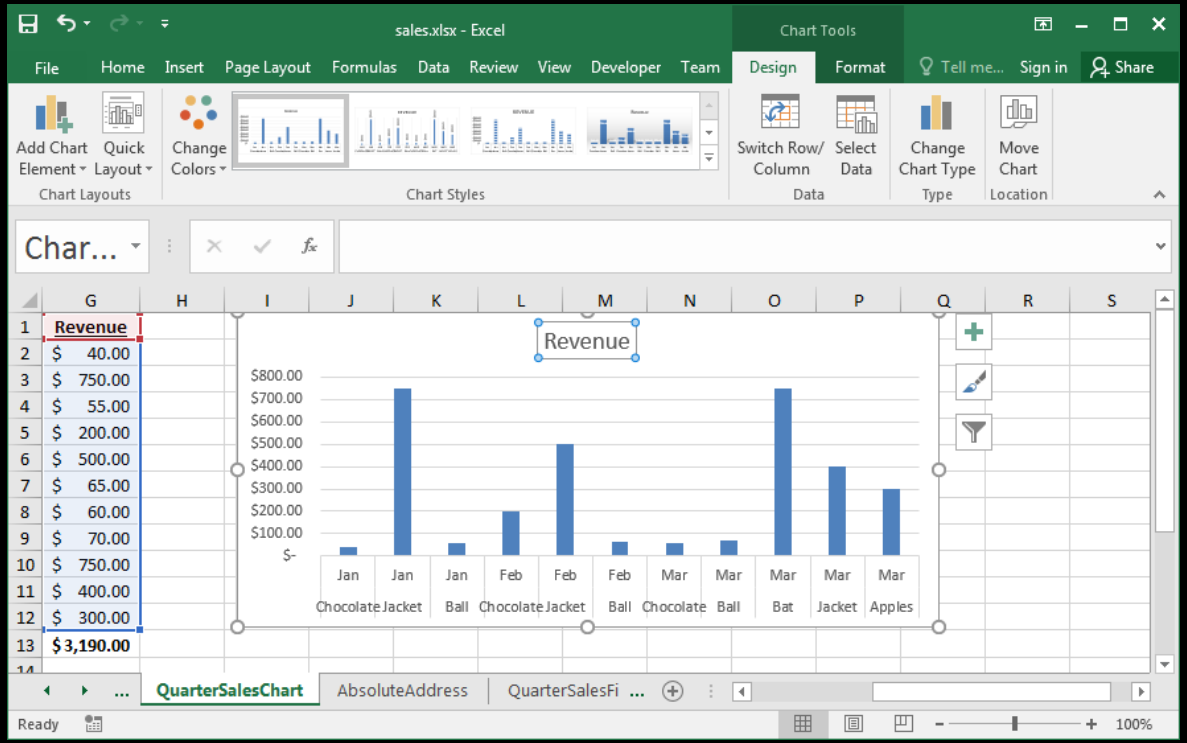


Chart Options

Chart Tools allows you to modify the data in the chart, change the chart type, and move the chart in the Worksheet.

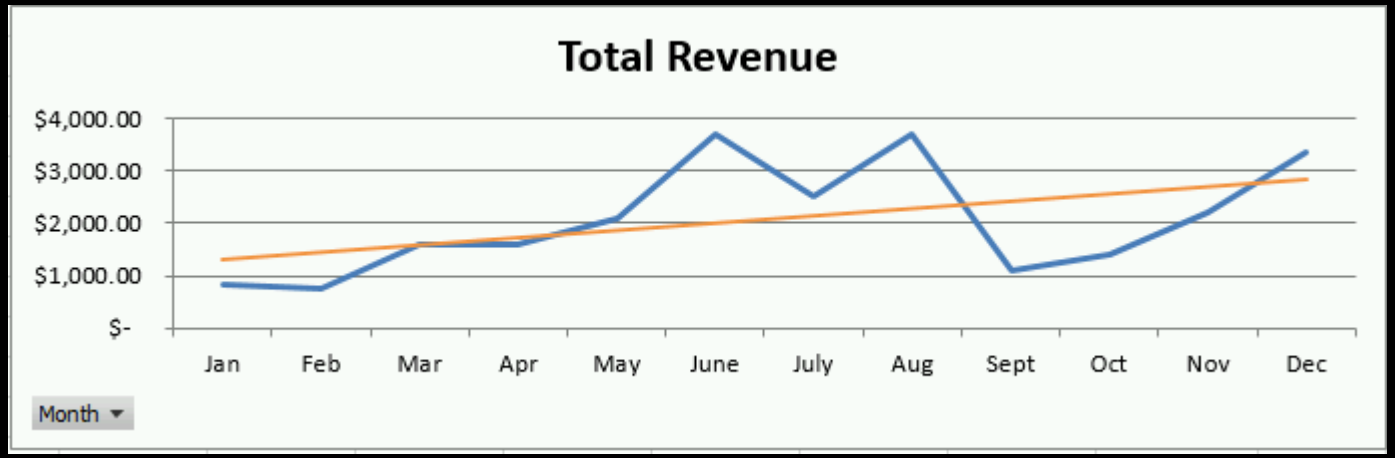


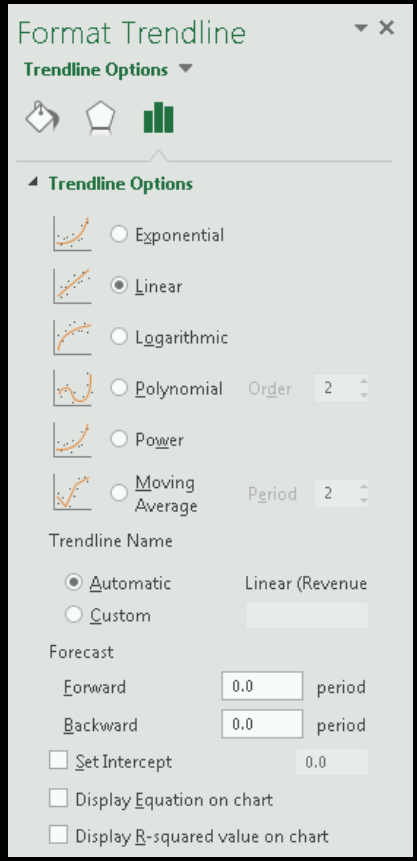
	Revenue
1	Revenue
2	$ 40.00
3	$ 750.00
4	$ 55.00
5	$ 200.00
6	$ 500.00
7	$ 65.00
8	$ 60.00
9	$ 70.00
10	$ 750.00
11	$ 400.00
12	$ 300.00
13	$ 3,190.00

Trendlines

`Trendlines` can be easily added to any chart.

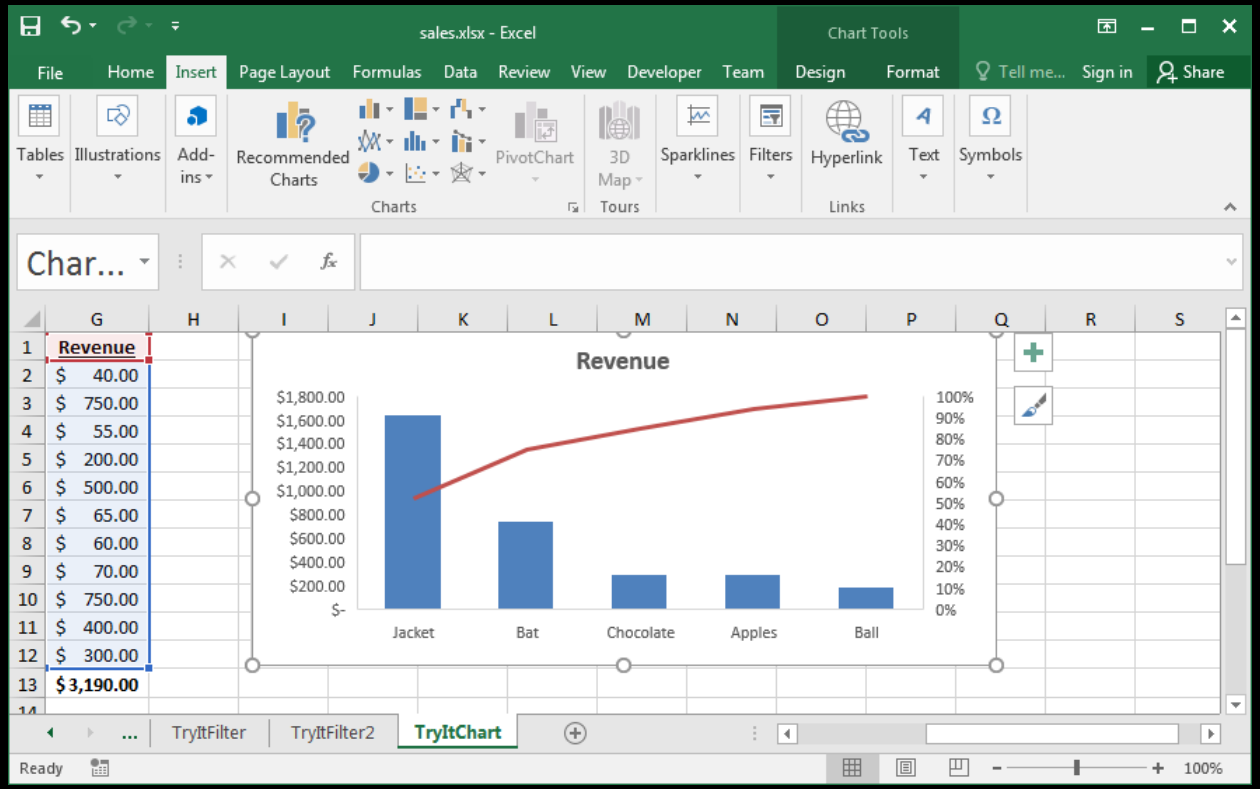
- Linear treadline for monthly revenue. Good choice?





Try it: Chart

Question: Create a chart that makes it easy to see the best selling product.

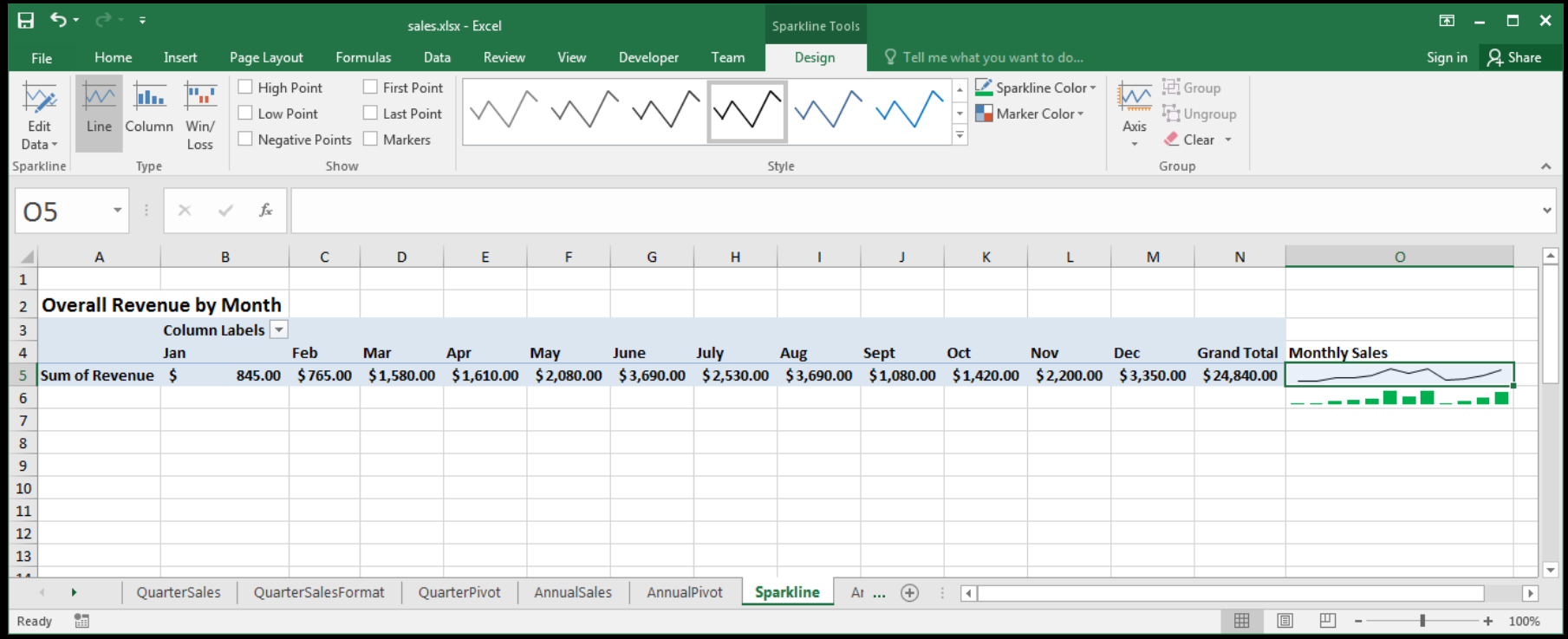


	G
1	Revenue
2	$ 40.00
3	$ 750.00
4	$ 55.00
5	$ 200.00
6	$ 500.00
7	$ 65.00
8	$ 60.00
9	$ 70.00
10	$ 750.00
11	$ 400.00
12	$ 300.00
13	$3,190.00

Sparklines

A ***sparkline*** is a tiny chart in a worksheet cell for a quick data overview.

- `Insert` then select a Sparkline (line, column, win/loss). May put text in sparkline cell.



What-If

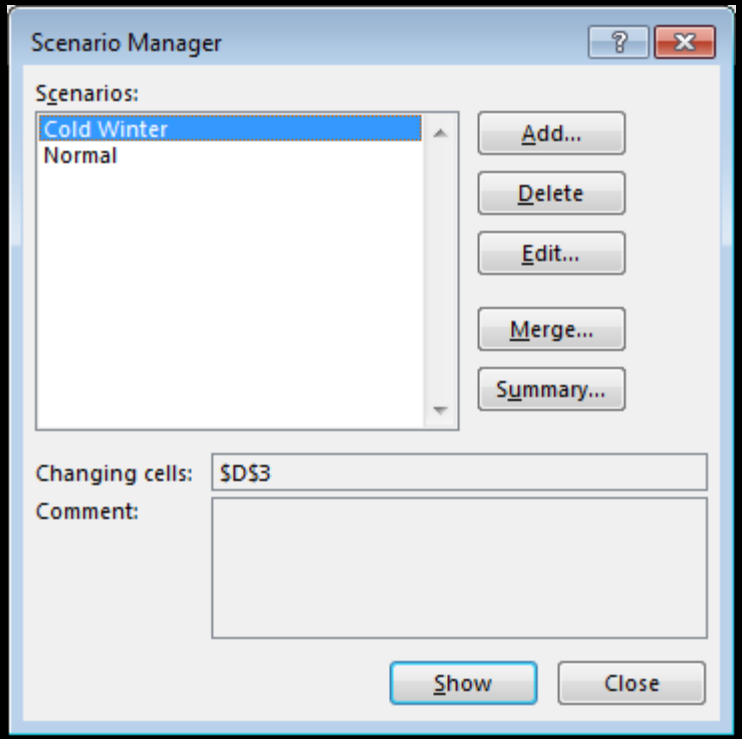
What-If scenarios help understand different possibilities.

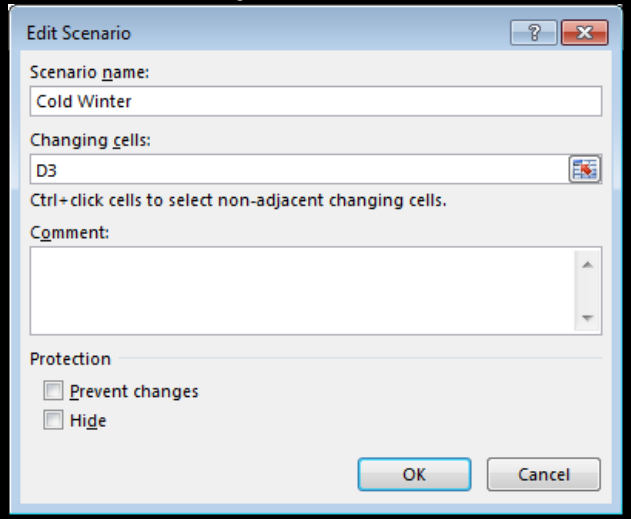
A what-if scenario is created under `Data` then `What-If Analysis` then `Scenario Manager`.

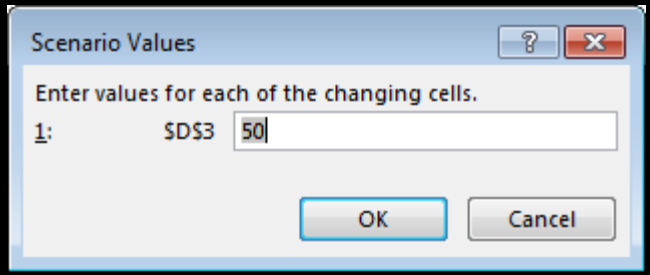
To define a scenario, give it a name and list the cells that will change with this scenario.

What-If Scenarios Example

Consider what happens with a cold winter and we predict to sell **50** jackets instead of the normal **15**.

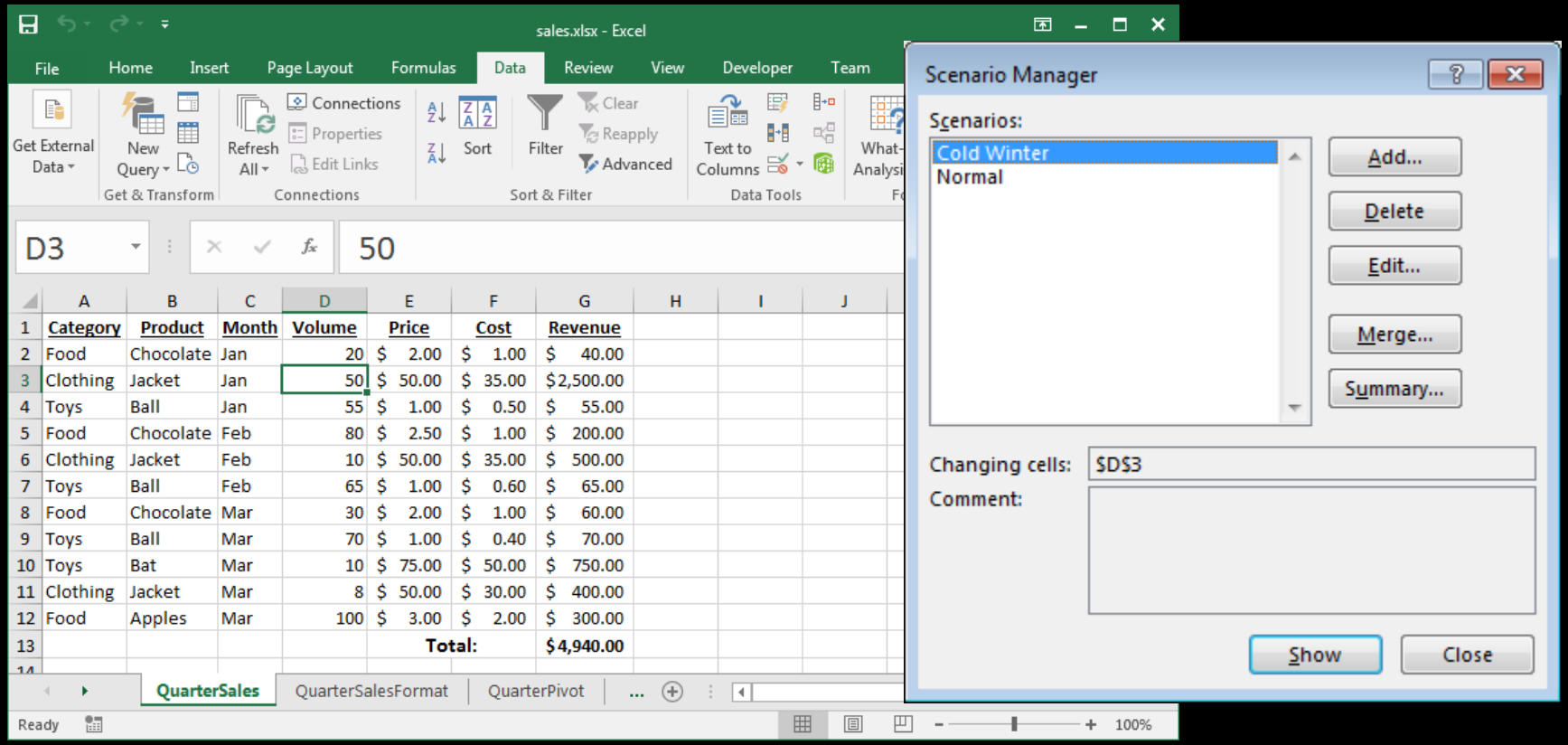






What-If Scenarios Example

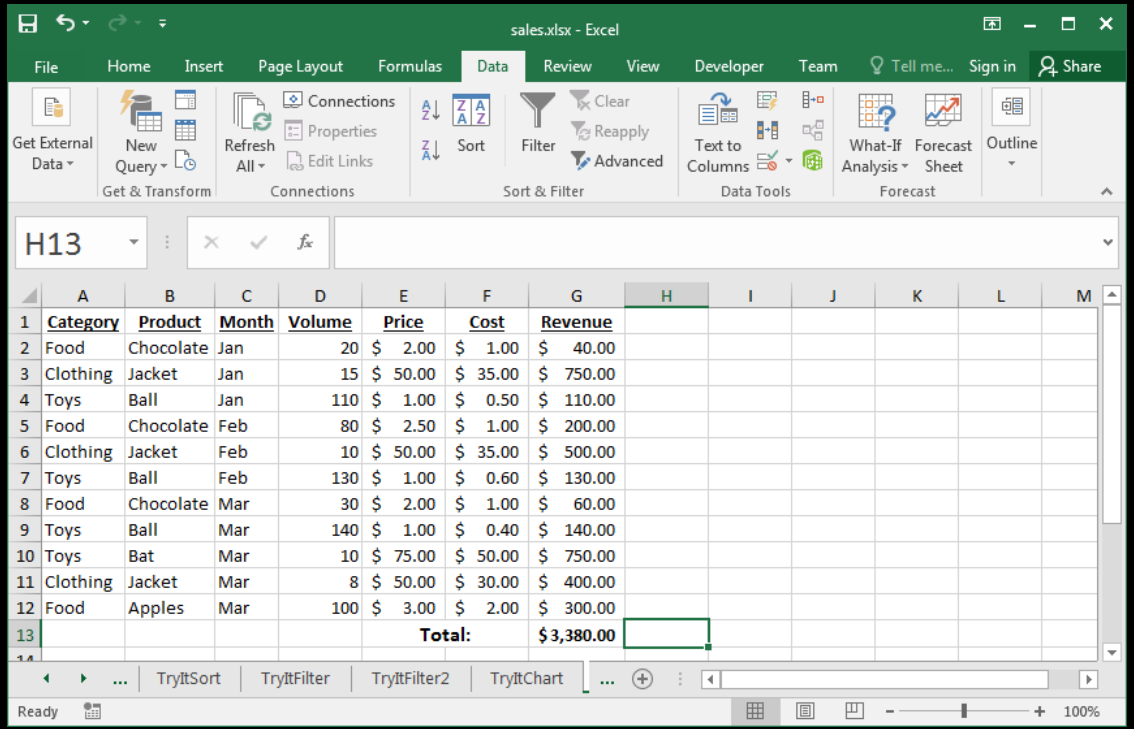
User can easily select scenario and see the result.



	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	50	$ 50.00	$ 35.00	$2,500.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$4,940.00

Try it: What-If Scenario

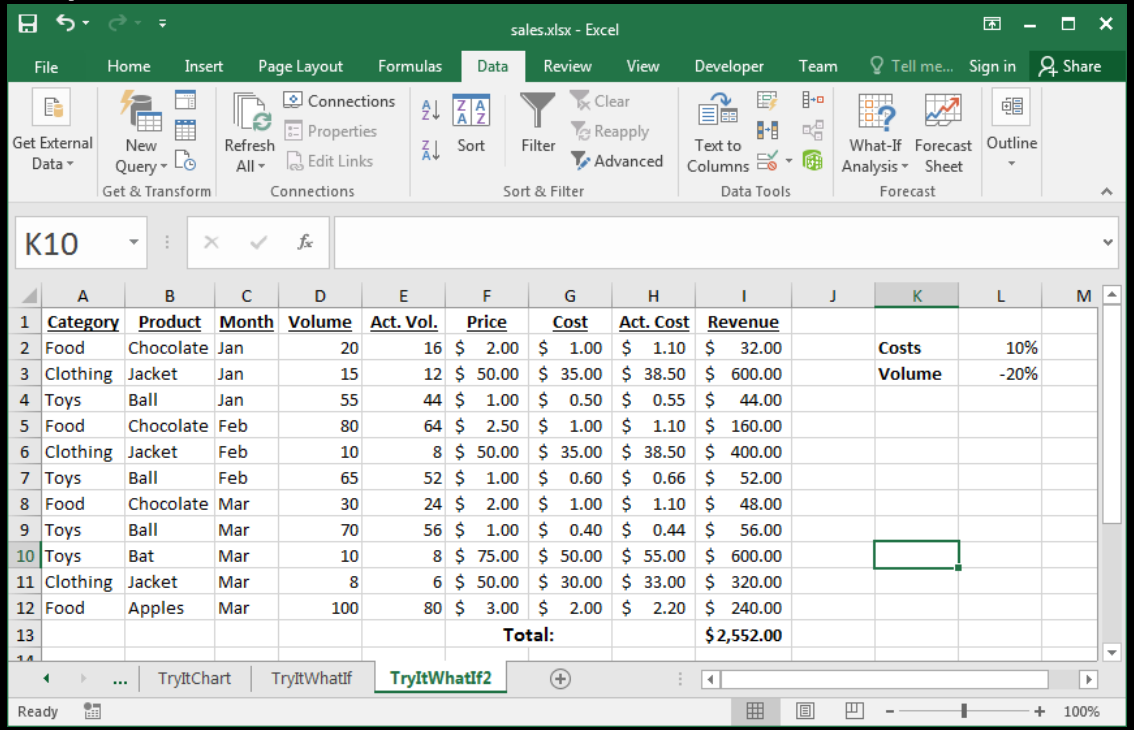
Question: Create a what-if scenario that wherever balls are sold, the volume is double than normal.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	Product	Month	Volume	Price	Cost	Revenue						
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00						
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00						
4	Toys	Ball	Jan	110	$ 1.00	$ 0.50	$ 110.00						
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00						
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00						
7	Toys	Ball	Feb	130	$ 1.00	$ 0.60	$ 130.00						
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00						
9	Toys	Ball	Mar	140	$ 1.00	$ 0.40	$ 140.00						
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00						
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00						
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00						
13					Total:		$ 3,380.00						

Try it: What-If Scenario Challenge

Question: Create a what-if scenario that all costs go up by 10% and volume down by 20%.

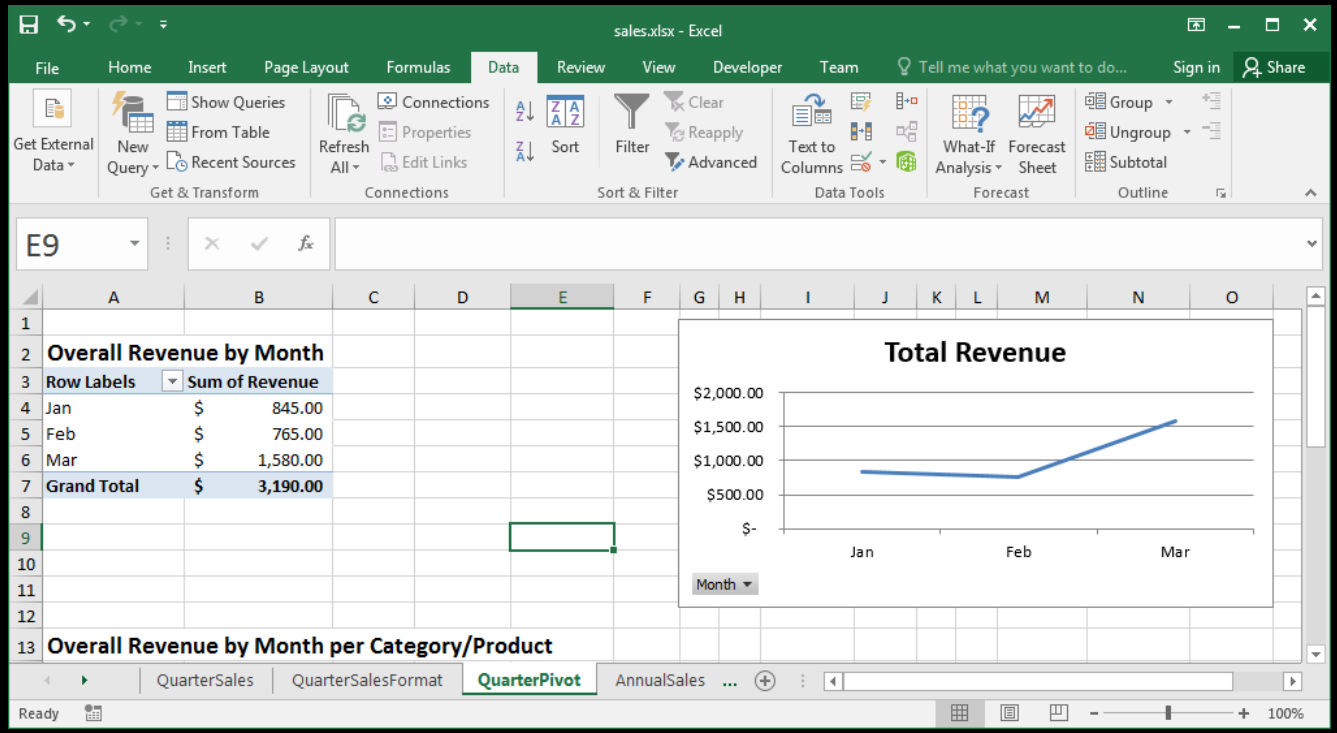


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	Product	Month	Volume	Act. Vol.	Price	Cost	Act. Cost	Revenue				
2	Food	Chocolate	Jan	20	16	$ 2.00	$ 1.00	$ 1.10	$ 32.00		Costs	10%	
3	Clothing	Jacket	Jan	15	12	$ 50.00	$ 35.00	$ 38.50	$ 600.00		Volume	-20%	
4	Toys	Ball	Jan	55	44	$ 1.00	$ 0.50	$ 0.55	$ 44.00				
5	Food	Chocolate	Feb	80	64	$ 2.50	$ 1.00	$ 1.10	$ 160.00				
6	Clothing	Jacket	Feb	10	8	$ 50.00	$ 35.00	$ 38.50	$ 400.00				
7	Toys	Ball	Feb	65	52	$ 1.00	$ 0.60	$ 0.66	$ 52.00				
8	Food	Chocolate	Mar	30	24	$ 2.00	$ 1.00	$ 1.10	$ 48.00				
9	Toys	Ball	Mar	70	56	$ 1.00	$ 0.40	$ 0.44	$ 56.00				
10	Toys	Bat	Mar	10	8	$ 75.00	$ 50.00	$ 55.00	$ 600.00				
11	Clothing	Jacket	Mar	8	6	$ 50.00	$ 30.00	$ 33.00	$ 320.00				
12	Food	Apples	Mar	100	80	$ 3.00	$ 2.00	$ 2.20	$ 240.00				
13						Total:			$ 2,552.00				

Pivot Tables

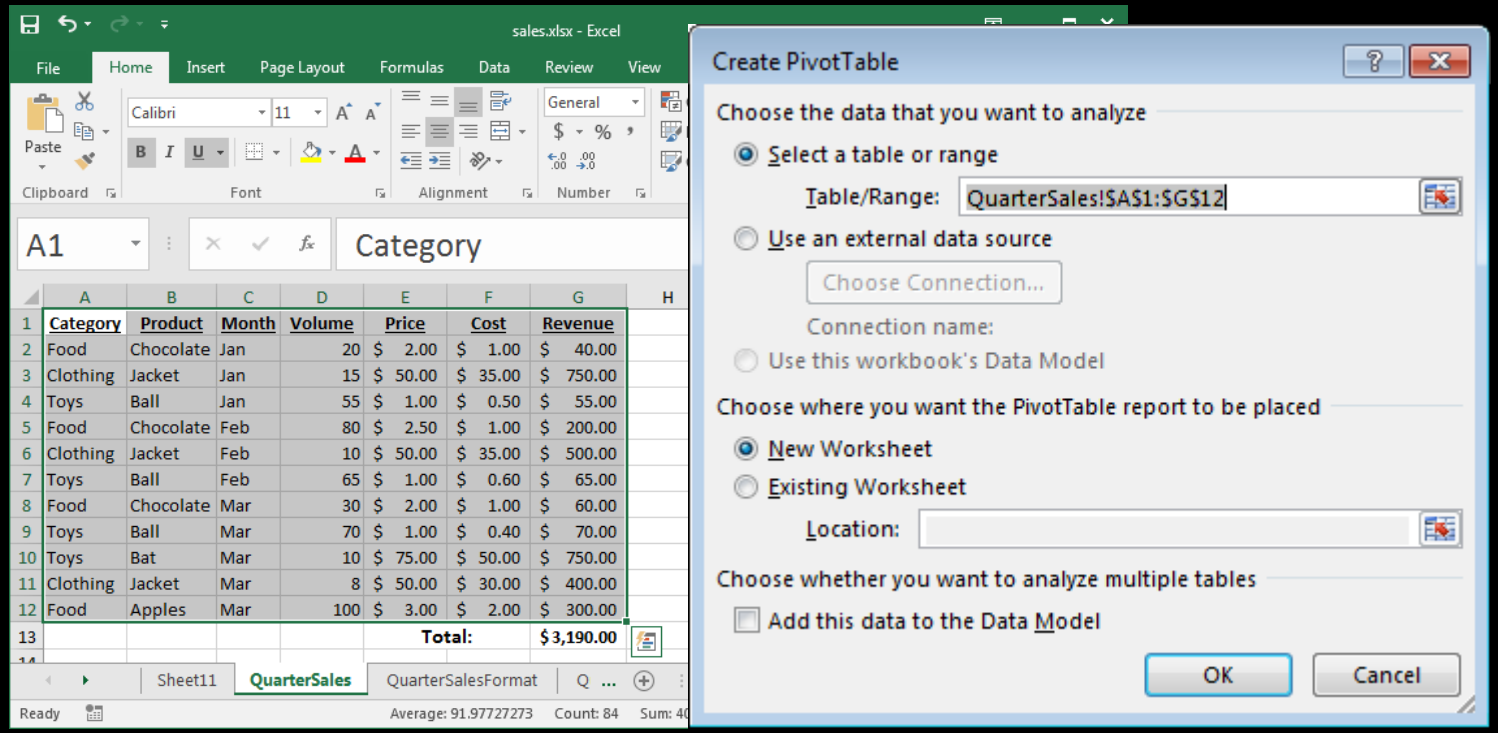
Pivot tables allow for easily aggregating and exploring large data sets.

- For example, our data set can be summarized by revenue by month.



Creating a Pivot Table

To create, select the data and then `Insert`, `Pivot Table`.



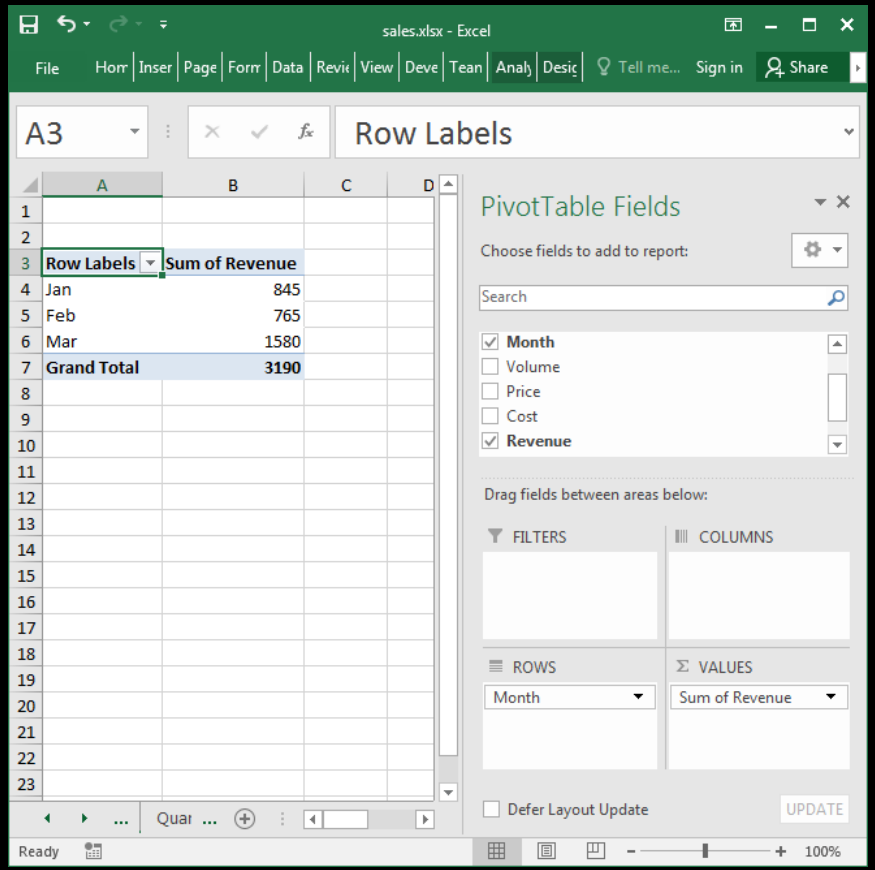
	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$3,190.00

Creating a Pivot Table

Add fields to pivot table.

Field may either be:

- Row value
- Column value
- Cell value (aggregated)
- Used in a filter



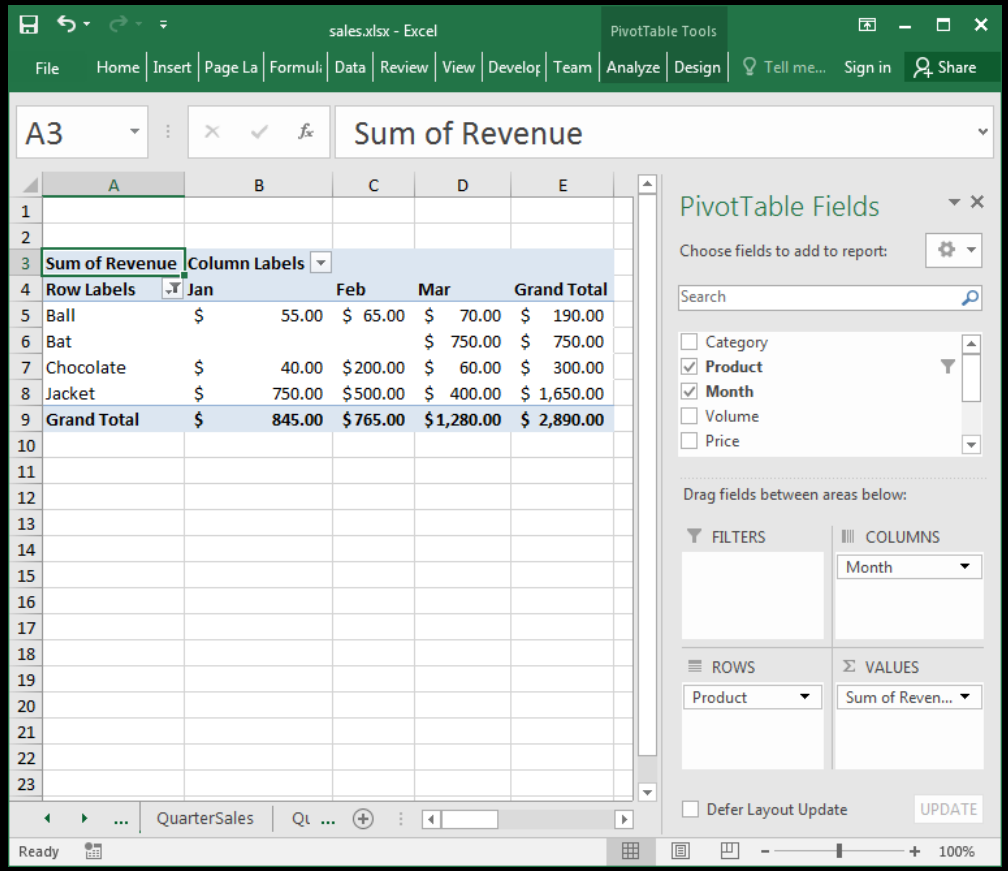
Creating a Pivot Table Example

Products are rows.

Months are columns.

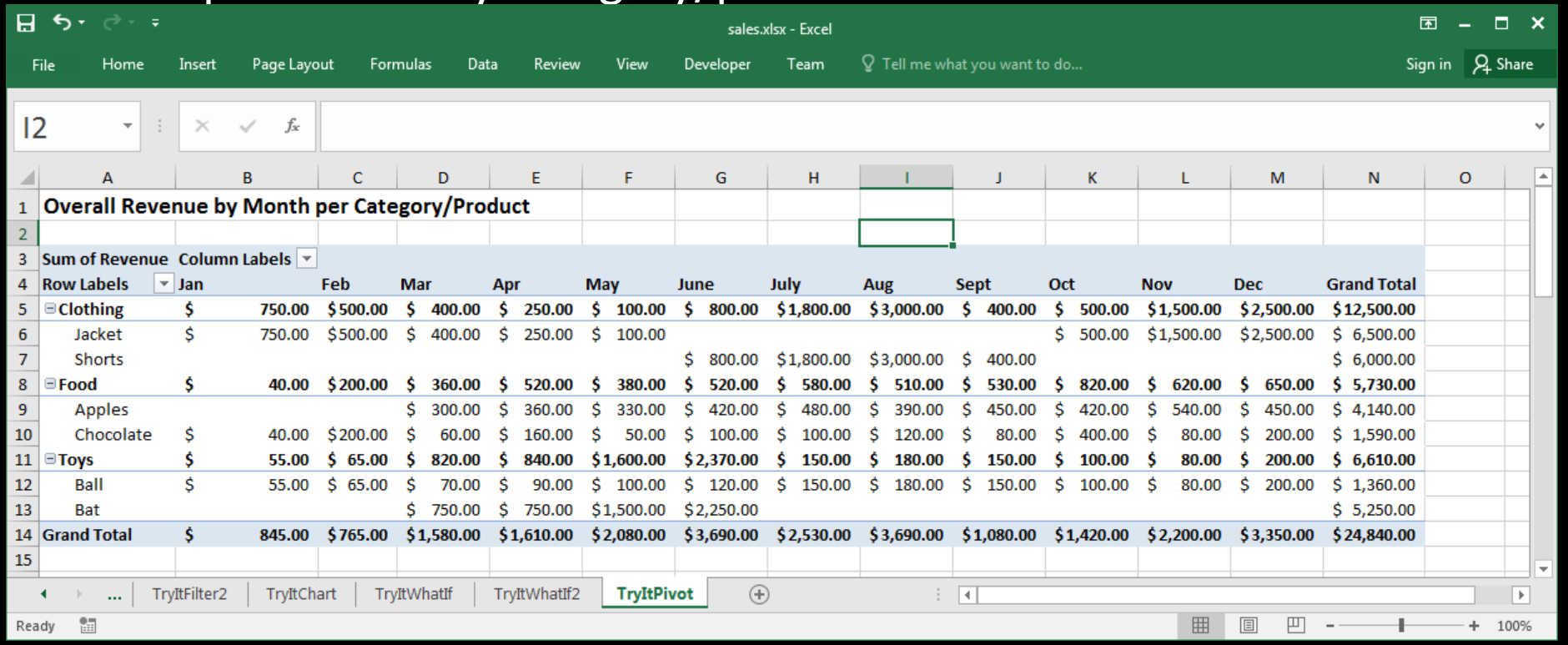
Each cell is a sum of revenue per product for that month.

Filter on product.



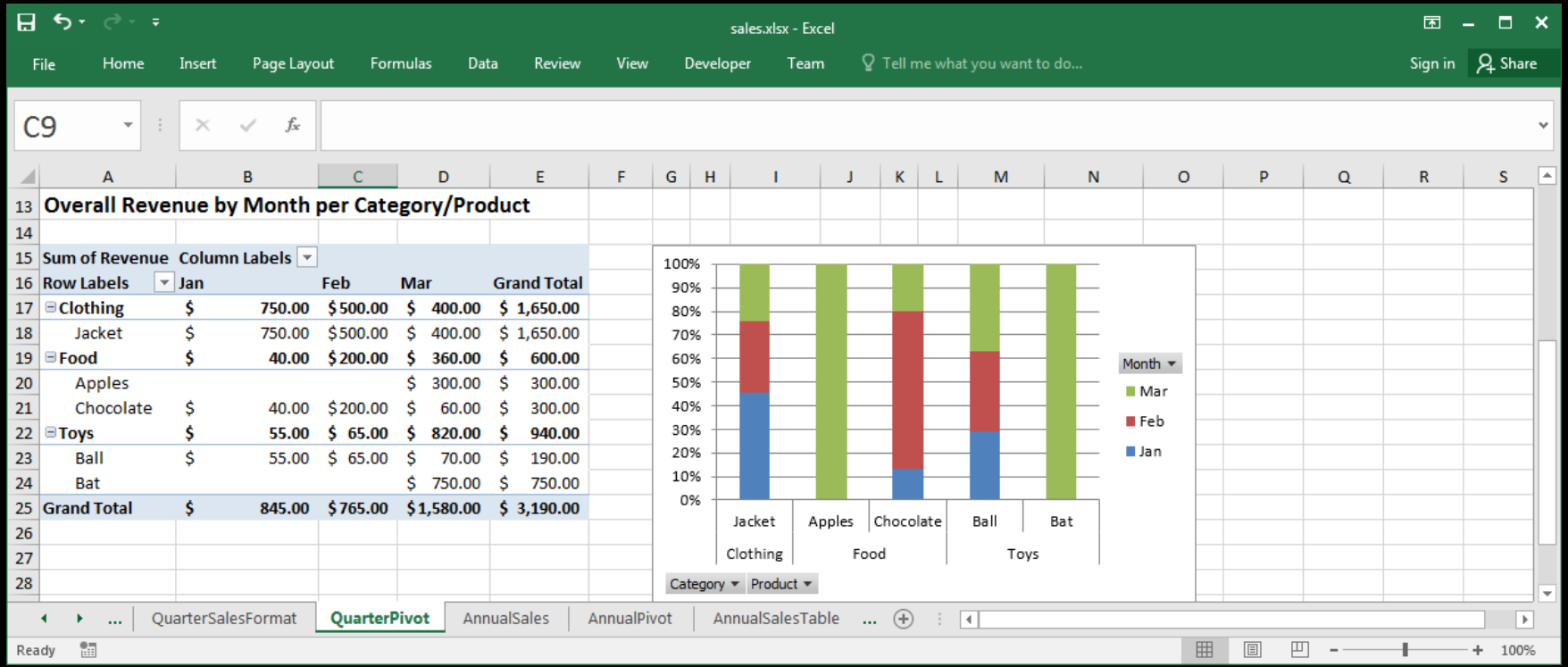
Try it: Pivot Table

Question: Create a pivot table using the annual sales data that shows revenue per month by category/product.



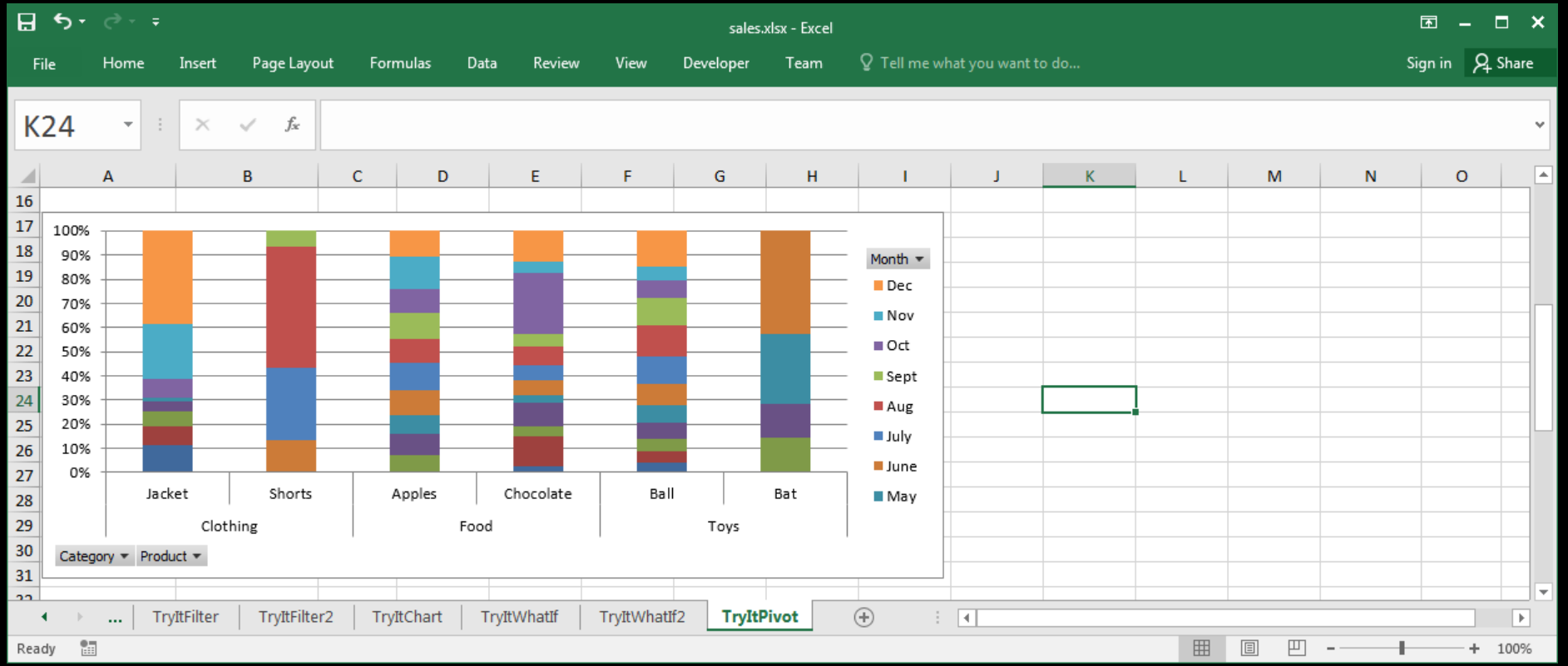
Pivot Charts

A ***pivot chart*** is a chart attached to a pivot table. Create it under `Insert` then `Pivot Chart`.



Try it: Pivot Chart

Question: Create a pivot chart for previous pivot table.



What-if and Pivot Tables Question

Question: How many of the following statements are **TRUE**?

1) A what-if scenario can have multiple cells change not just one.

2) A pivot table field can be used in `ROWS` and `COLUMNS` at the same time.

3) A pivot table field can be used in `VALUES` more than once.

4) In our sales spreadsheet example, if `Product` and `Category` are both used in `ROWS` then the order they are list does not matter.

5) It is not possible for a field that is a string to be used in `VALUES`.

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Conditions and Decisions

A ***condition*** is an expression that is either `TRUE` or `FALSE`.

Conditions are used to make decisions and perform different actions depending on the condition value.

Excel condition and decision functions:

- `FALSE()` – returns `FALSE`
- `TRUE()` – returns `TRUE`
- `AND(cond1, cond2)` – returns `TRUE` if both `cond1` and `cond2` are true
- `OR(cond1, cond2)` – returns `TRUE` if either or both of `cond1` and `cond2` are true
- `NOT(cond)` – returns `TRUE` if `cond` is `FALSE`

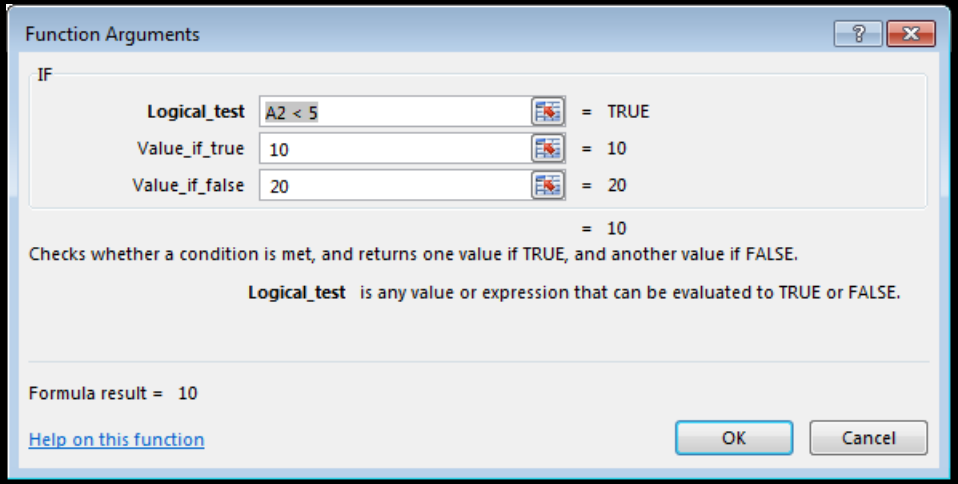
Decisions using `IF()`

The `IF()` function is used to make a decision based on a condition.

- `IF(condition, value_if_true, value_if_false)`

Example: If cell A2 is less than 5, return 10 otherwise return 20.

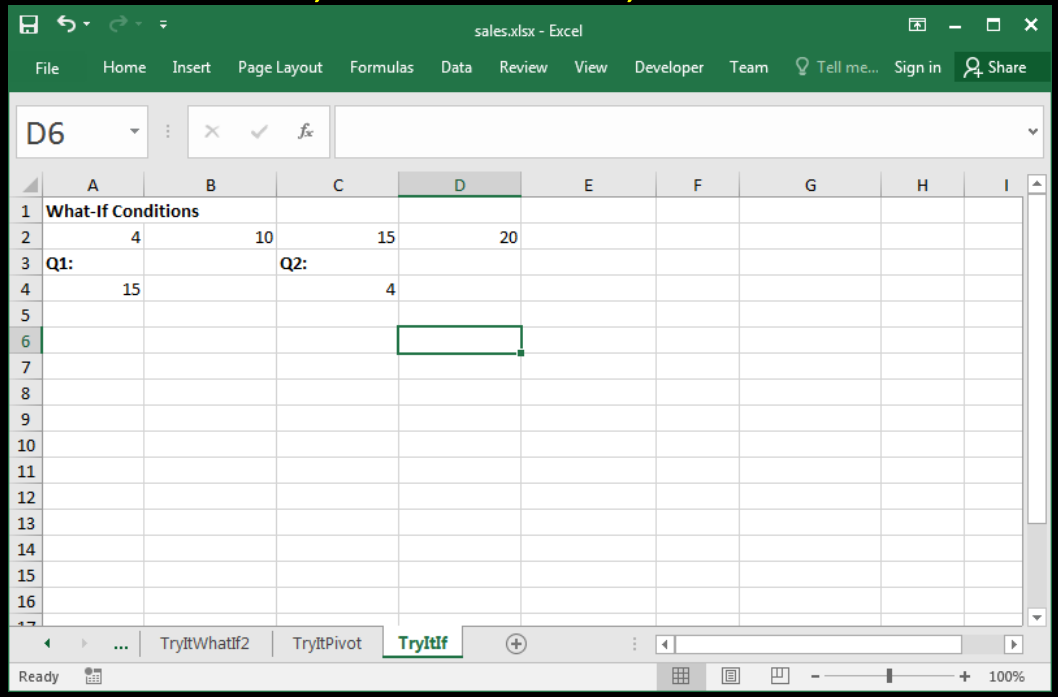
```
= IF(A2 < 5, 10, 20)
```



Try it: Conditions and `IF()`

Question: Create two conditions:

- 1) If cell B2 >= 10, then show C2, otherwise D2.
- 2) If cell B2 < 15 and C2 > 20, return B2*C2, otherwise if D2 < 10, return 1, else 4.



Decisions using IF() Question

Question: How many of these statements are **TRUE**? **A1=40, A2=10**

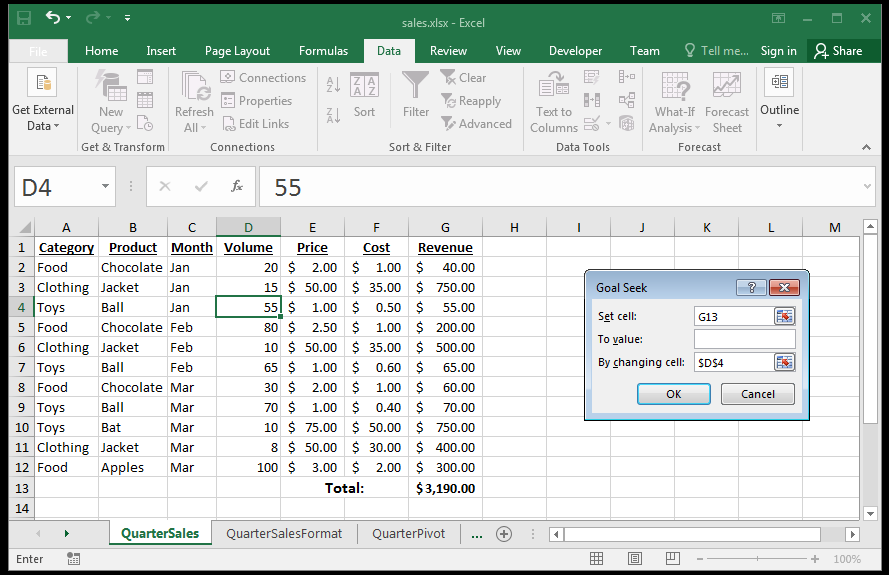
1) `=AND(FALSE(), TRUE())`
2) `=OR(FALSE(), NOT(TRUE()))`
3) `=IF(A1=40, 5, 10)` returns `10`.
4) `=IF(OR(A1=40,A2>10),1, 2)` returns `2`.
5) `=IF(A2=10,IF(A1=40,FALSE()),TRUE())`

A) 0 **B)** 1 **C)** 2 **D)** 3 **E)** 4

Goal Seek

Goal seek is used to have Excel solve for a variable given the target value of another cell.

- Example: How many balls would we have to sell in January to have total revenue for first 3 months of $4000? Answer: 865

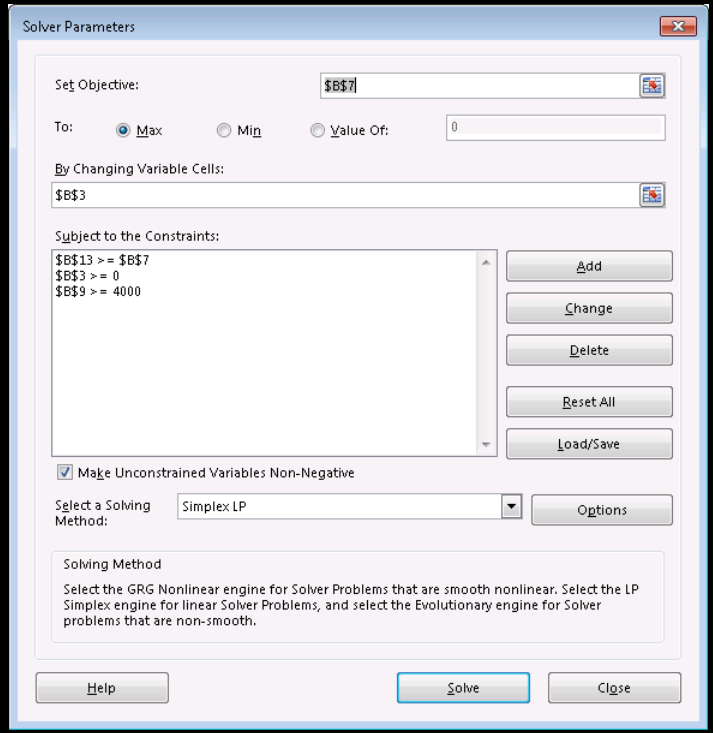


	A	B	C	D	E	F	G
1	Category	Product	Month	Volume	Price	Cost	Revenue
2	Food	Chocolate	Jan	20	$ 2.00	$ 1.00	$ 40.00
3	Clothing	Jacket	Jan	15	$ 50.00	$ 35.00	$ 750.00
4	Toys	Ball	Jan	55	$ 1.00	$ 0.50	$ 55.00
5	Food	Chocolate	Feb	80	$ 2.50	$ 1.00	$ 200.00
6	Clothing	Jacket	Feb	10	$ 50.00	$ 35.00	$ 500.00
7	Toys	Ball	Feb	65	$ 1.00	$ 0.60	$ 65.00
8	Food	Chocolate	Mar	30	$ 2.00	$ 1.00	$ 60.00
9	Toys	Ball	Mar	70	$ 1.00	$ 0.40	$ 70.00
10	Toys	Bat	Mar	10	$ 75.00	$ 50.00	$ 750.00
11	Clothing	Jacket	Mar	8	$ 50.00	$ 30.00	$ 400.00
12	Food	Apples	Mar	100	$ 3.00	$ 2.00	$ 300.00
13					Total:		$ 3,190.00
14							

Linear Programming with Solver

Solver performs linear programming to maximize or minimize a given function by changing multiple variables subject to constraints.

	A	B
1	Solver Example: Maximum House You Can Afford	
2		
3	House Value:	$ 500,000.00
4	Down payment:	$ 20,000.00
5	Mortgage amount:	$ 480,000.00
6	Amoritization in years:	25
7	Monthly mortgage expense:	$2,533.62
8		
9	Monthly income:	$ 5,000.00
10	Mortgage rate:	4%
11		
12	Maximum mortgage expense percentage:	30%
13	Maximum mortgage expense amount:	$ 1,500.00
14		$1,033.62



Analysis ToolPak

The Analysis ToolPak is an Excel add-in that has a set of statistical and data analysis tools such as ANOVA, covariance, regression, and t-test.

Analysis ToolPak is not installed by default.

- To install: `File → Options → Add-Ins`
- Select `Excel Add-ins` inn the `Manage:` box and select `Go...`
- Choose `AnalysisToolPak` and select `OK`

You should now see `Data Analysis` under the Data tab

Regression

Linear regression models the relationship between a dependent variable *y* and explanatory variables *X*.

- Simple linear regression has one explanatory variable: $y = Bx + \varepsilon$
- Used to fit a predictor model on observed data and also used to determine the strength of the relationship between *y* and *X* variables.

Trend lines are often calculated using linear regression.

The technique provides a way to determine patterns in the data set and model the data so that new values can be predicted.

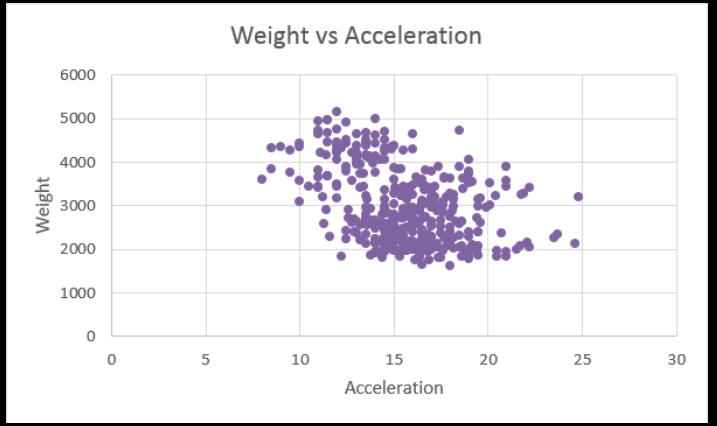
Regression in Excel

Excel provides a regression function that will calculate:

- R^2
- ANOVA table
- regression equation coefficients
- standardized and unstandardized residuals

Example: Given a data set of car weight and acceleration, determine if there is any relationship between them.

Scatterplot shows weak relationship with no strong patterns, and we would expect to see this shown in the regression analysis.



Regression Example

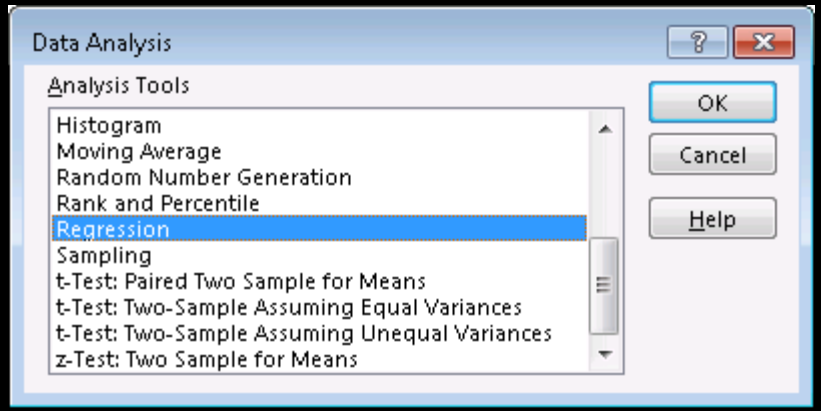
Regression computes constants *m* and *b* in formula:

weight = *m****acceleration** + *b*

Weight is the dependent variable and acceleration is the independent variable.

To start select, `Data Analysis`

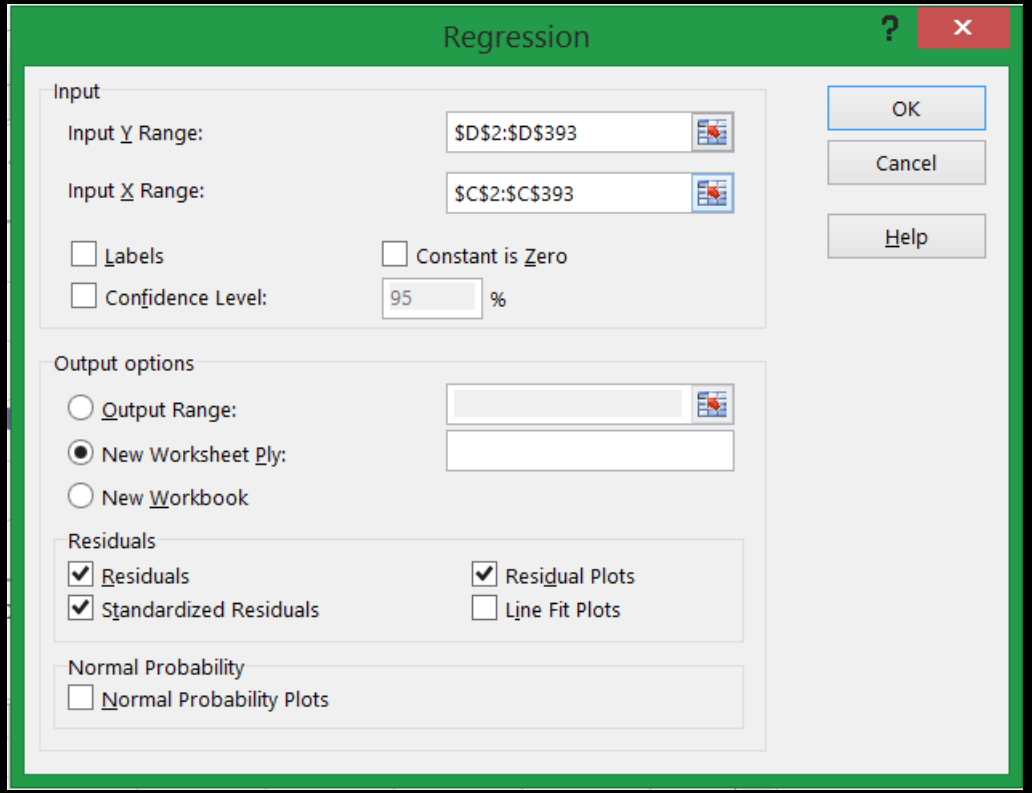
from the data tab and then select `Regression` and `OK`.



Regression Example Settings

Settings:

- Response (dependent) data for the Input Y Range
- Columns for the explanatory (independent) data (X Range).
- For residual information select, Residuals, Standardized Residuals, and Residual Plots from the Residuals section.



Regression Example Results

	A	B	C	D	E	F	G	H	I
1	SUMMARY OUTPUT								
2									
3	*Regression Statistics*								
4	Multiple R	0.416839282							
5	R Square	0.17375492							
6	Adjusted R Square	0.171636343							
7	Standard Error	2.510965983							
8	Observations	392							
9									
10	ANOVA								
11		*df*	*SS*	*MS*	*F*	*Significance F*			
12	Regression	1	517.0999442	517.0999442	82.01491373	6.56562E-18			
13	Residual	390	2458.930566	6.304950169					
14	Total	391	2976.03051						
15									
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
17	Intercept	19.57266581	0.462860061	42.28635703	9.5445E-148	18.66265269	20.48267893	18.66265269	20.48267893
18	X Variable 1	-0.001353896	0.000149499	-9.056208574	6.56562E-18	-0.001647821	-0.001059971	-0.001647821	-0.001059971

R^2 * 100% = percentage of variation in dependent variable explained by independent variable

Coefficients for the regression equation

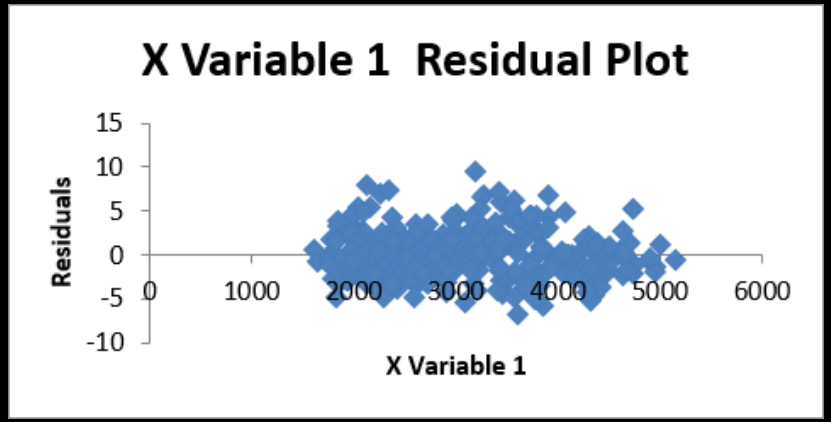
All of the output is put into a new sheet. Read the values off of the table and form the regression equation:

- weight = -0.001*acceleration + 19.572

Regression Example Results (cont.)

Below the previous tables are the predicted y values (from the regression equation) as well as the residuals and standardized residuals. All plots are placed to the right of the charts.

Observation	Predicted Y	Residuals	Standard Residuals
1	14.82861427	-2.828614269	-1.127947728
2	14.57272793	-3.072727927	-1.22529131
3	14.9206792	-3.920679196	-1.563423207
4	14.92474088	-2.924740884	-1.1662795
5	14.90307855	-4.403078548	-1.755786393
6	13.69540333	-3.695403327	-1.473591445
7	13.67780268	-4.67780268	-1.865336311
8	13.73466631	-5.234666311	-2.087393123
9	13.58167606	-3.581676065	-1.428241179
10	14.36016626	-5.860166257	-2.336819583
11	14.74873441	-4.748734406	-1.893621284

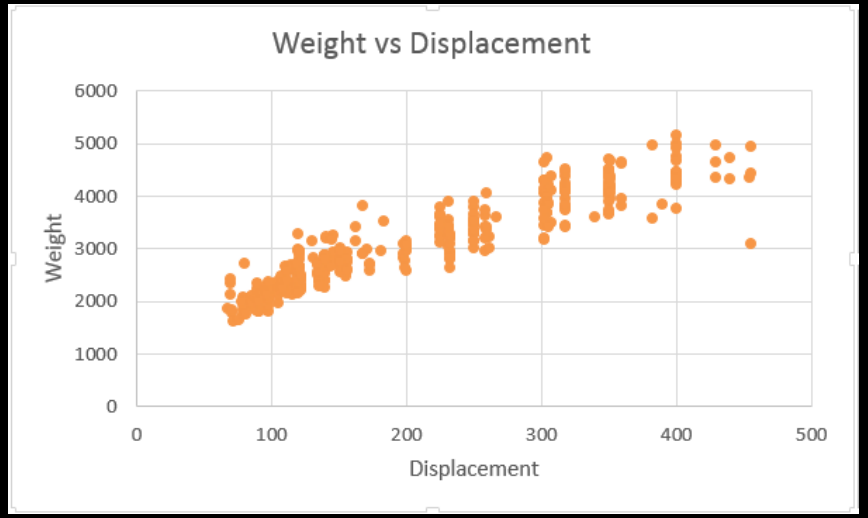


Expected a weak relationship and this is demonstated by the R^2 value.

- Only 17.4% of the variation in weight is explained by acceleration.

Try It: Regression

Question: Perform a regression analysis between weight (dependent) and displacement (independent) variable.



Conclusion

Spreadsheets are general purpose tools for data analysis that consist of a table of cells which contain data and formulas.

Formulas contain data values, cell references, and functions.

- Aggregate functions summarize multiple data values into a single value.
- Functions exist for statistics, string manipulation, lookup/indexing, and decisions.

Spreadsheets provide tools for data sorting, filtering, visualization using charts, and summarization (pivot tables).

- Also contain tools for what-if scenarios, goal seek, linear solvers, and statistical analysis tools.

Objectives

- Explain what a spreadsheet is.
- Explain how cells are addressed in a spreadsheet.
- List some of the ways to select cells in a spreadsheet.
- Define and explain: formula, function, argument, concatenation
- Use these functions: concatenate, lookup, index
- Explain the difference between an absolute and relative address.
- Explain how an aggregate function works. List some examples.
- Explain how to use conditional formatting.
- Explain how spreadsheets can be used as a database. Use sorting and filtering.
- Be able to create and edit charts and use chart features: trendlines, sparklines
- Explain the usefulness of: what-if scenarios, goal seek, solver
- Use and create pivot tables and charts.
- Evaluate and create conditions. Use IF() to make decisions.