

# DATA CLEANING IN 10 MINUTES



Hi-res PDF, Video & Practice File: <https://bit.ly/data-cleaning-10>



## 1 Identify & Remove Duplicates: Avoid wrong analysis

How?

1. Select the column with duplicate data > Home tab > Conditional Formatting > Highlight Cell Rules > Duplicate Values
2. Assess the highlighted duplicate values
3. Data tab > Remove Duplicates > Delete duplicate values

### 1. Identify duplicates

Date	ID	Name	Product	Quantity	Price Per Unit
1/31/2021 0:00	1	John Smith	Magic Wand	10	\$20.00
2/28/2021 0:00	2	Jane Doe	Unicorn Horn	15	\$10.00
4/30/2021 0:00	3	Anna Belle	Fairy Dust	25	\$10.00
5/31/2021 0:00	4	Chris P. Bacon	Bacon Scented Candle	30	\$16.67
6/30/2021 0:00	5	Peter Parker	Web Shooter	0	inf
7/31/2021 0:00	6	Mary Jane	Potent Potion	35	\$10.00
8/31/2021 0:00	7	Bruce Wayne	Bat Signal	40	\$15.00
4/30/2021 0:00	3	Anna Belle	Fairy Dust	25	\$10.00
5/31/2021 0:00	4	Chris P. Bacon	Bacon Scented Candle	30	\$16.67
6/30/2021 0:00	5	Peter Parker	Web Shooter	0	inf



### 2. Arrive at duplicate-free data

Date	ID	Name	Product	Quantity	Price Per Unit
1/31/2021 0:00	1	John Smith	Magic Wand	10	\$20.00
2/28/2021 0:00	2	Jane Doe	Unicorn Horn	15	\$10.00
4/30/2021 0:00	3	Anna Belle	Fairy Dust	25	\$10.00
5/31/2021 0:00	4	Chris P. Bacon	Bacon Scented Candle	30	\$16.67
6/30/2021 0:00	5	Peter Parker	Web Shooter	0	inf
7/31/2021 0:00	6	Mary Jane	Potent Potion	35	\$10.00
8/31/2021 0:00	7	Bruce Wayne	Bat Signal	40	\$15.00



## 5 Use Appropriate Number Formats: Minimize clutter

### Best Practices

1. Use "Short Date format" if time component is not required
2. Avoid Currency format – Use "Number" or "General" formats to declutter currency symbol from each row of your data
3. CTRL+1 > Select "Number" format > Check "Use 1000 Separator"
4. Use the currency symbol in the column headers



Date	Product	Quantity	Price Per Unit	Sales
1/31/2021 0:00	Magic Wand	10	\$20.00	\$200.00
6/30/2021 0:00	Web Shooter	0	inf	
7/31/2021 0:00	Potent Potion	35	\$10.00	\$350.00
8/31/2021 0:00	Bat Signal	40	inf	
2/28/2022 0:00	Unicorn Horn	30	\$36.67	\$1,100.10

Date	Product	Quantity	Price Per Unit	Sales \$
1/31/2021	Magic Wand	10	20.00	200.00
6/30/2021	Web Shooter	0	inf	
7/31/2021	Potent Potion	35	10.00	350.00
8/31/2021	Bat Signal	40	inf	
2/28/2022	Unicorn Horn	30	36.67	1,100.10



## 6 Find & Replace: Replace unwanted values in one-go

Find and Replace

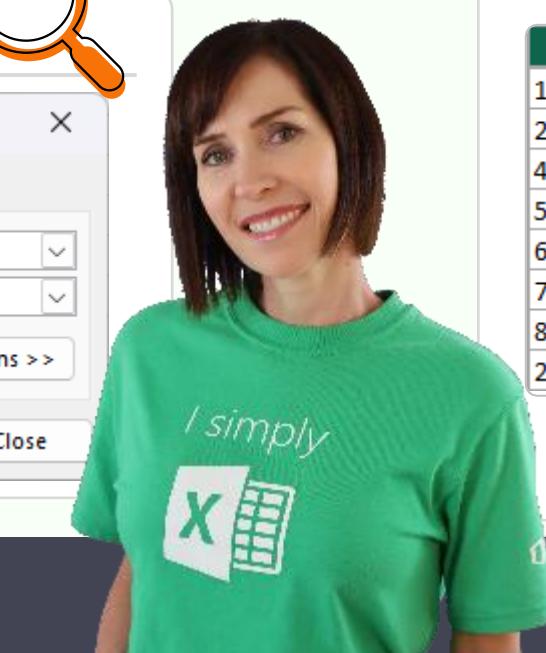
Find what: inf

Replace with:

CTRL H

Replace All Replace Find All Find Next Close Options >

Product	Price Per Unit \$
Magic Wand	20.00
Unicorn Horn	inf
Fairy Dust	10.00
Bacon Scented Candle	16.67
Web Shooter	inf
Potent Potion	15.00
Bat Signal	inf
Gamma Radiation Serum	inf



## 2 Trim Extra Spaces: Avoid formula errors caused by rogue spaces

How?

=TRIM(D6)					
A	B	C	D	E	F
3	Date	ID	Name	Region	Trimmed Name
4	1/31/2021 0:00	1	John Smith	North	John Smith
5	2/28/2021 0:00	2	Jane Doe	East	Jane Doe
6	4/30/2021 0:00	3	Anna Belle	South	Anna Belle
7	5/31/2021 0:00	4	Chris P. Bacon	East	Chris P. Bacon
8	6/30/2021 0:00	5	Peter Parker		Peter Parker
9	7/31/2021 0:00	6	Mary Jane	West	Mary Jane
10	8/31/2021 0:00	7	Bruce Wayne	South	Bruce Wayne

TRIM Masterclass: <https://bit.ly/fntrim>



How?

1. Insert a new column
2. Use Trim function on the original data
3. Paste as values the trimmed data to the original data. (Microsoft 365 users can use **CTRL + SHIFT + V** to preserve formatting)

## 3 Fill Blank Cells: Prevent unforeseen analysis disruptions

Date	ID	Name	Region	Product
1/31/2021 0:00	1	John Smith	North	Magic Wand
2/28/2021 0:00	2	Jane Doe	East	Unicorn Horn
4/30/2021 0:00	3	Anna Belle	South	Fairy Dust
5/31/2021 0:00	4	Chris P. Bacon	East	Bacon Scented Candle
6/30/2021 0:00	5	Peter Parker		Web Shooter
7/31/2021 0:00	6	Mary Jane	West	Potent Potion
8/31/2021 0:00	7	Bruce Wayne	South	Bat Signal
2/28/2022 0:00	8	Bruce Banner		Gamma Radiation Serum



How?

1. Select the column with blanks: Home > Find & Select > Go To Special > Blanks (or **CTRL+G**)
2. Fill in a placeholder: e.g., TBA > Press **CTRL+ENTER**

Automate Data Cleaning!

POWER QUERY COURSE

Automate BORING Tasks

ENROLL NOW!

<https://bit.ly/moth-courses>

## 4 Handle Errors with IFERROR: Avoid messy errors

How?

1. Select data > **CTRL+T** to insert table (Optional step - enables easy-to-read structured references)
2. Wrap your formula in the IFERROR function to avoid error message when dealing with missing or non-numeric values such as "inf"

IFERROR Masterclass: <https://bit.ly/iferrorfn>



A	B	C	D	E	F
3	Date	Product	Quantity	Price Per Unit	Sales
4	1/31/2021 0:00	Magic Wand	10	\$20.00	\$200.00
5	6/30/2021 0:00	Web Shooter	0	inf	#VALUE!
6	7/31/2021 0:00	Potent Potion	35	\$10.00	\$350.00
7	8/31/2021 0:00	Bat Signal	40	inf	#VALUE!
8	2/28/2022 0:00	Unicorn Horn	30	\$36.67	\$1,100.10

A	B	C	D	E	F
3	Date	Product	Quantity	Price Per Unit	Sales
4	1/31/2021 0:00	Magic Wand	10	\$20.00	\$200.00
5	6/30/2021 0:00	Web Shooter	0	inf	
6	7/31/2021 0:00	Potent Potion	35	\$10.00	\$350.00
7	8/31/2021 0:00	Bat Signal	4		