# Excel Find and Remove Duplicates 7 Ways!

**Download in hi-res here:** https://www.myonlinetraininghub.com/excel-find-and-remove-duplicates

# 1

Highlight Duplicates with Conditional Formatting

<u>Conditional Formatting</u> can quickly highlight duplicates in a column:

 select the column or columns of cells containing the suspected duplicates.
 Select Home tab.

3. Select Conditional Formatting.4. On the Insert tab, select Highlight Cells Rules.

5. Then select Duplicate Values.



Duplicate Values		?	×
Format cells that contain:			
Duplicate values with	Light Red Fill with D	ark Red T	ext 🔻
	ОК	Ca	ncel

**Tip:** You can change the format by clicking the drop down for 'Values with' (see image above).

Once the formatting is applied you can use filters (**Data tab** > **Filters**), Based on the cell fill color or font color to display or hide the duplicate values:



2 Identify Duplicates with a Formula You can add a column to your data table to tag rows containing duplicates. The formula below is looking for duplicate rows, i.e.

where both the Date and ID values are duplicated:

#### fx =IF(COUNTIFS(\$A\$7:\$A\$11,\$A7,\$B\$7:\$B\$11,\$B7)>1,"Duplicate","" D A 4 Identify Duplicate Rows ▼ ID ✓ Duplicates ✓ 5 Date 1/08/2017 C301 3/08/2017 C157 6/08/2017 C301 Duplicate 12/08/2017 C477 10 6/08/2017 C301 Duplicate 11

## The formula in cell C7:

### =IF(COUNTIFS(\$A\$7:\$A\$11,\$A7,\$B\$7:\$B\$11,\$B7)>1,"Duplicate","")

uses COUNTIFS to check both the Date and ID columns are the same, if the count is greater than 1, then 'Duplicate' is returned, otherwise the cell is left blank.

**Tip:** If you only want to check a single column, let's say the ID column, then you could use the **COUNTIF function** like so:

$\times \checkmark f_x$		=IF(COUN	TIF(\$B\$17:\$B\$21,\$B1	7)>1,"Duplica	ate","")
1	A	В	С	D	E
14	Identify Duplicate Cells in Column B		in Column B		
15				1	
16	Date	ID	Duplicates	)	
17	1/08/2017	C301	Duplicate		
18	3/08/2017	C157			
19	6/08/2017	C301	Duplicate		
20	12/08/2017	C477			
21	6/08/2017	C301	Duplicate		
	10. CONC. 10.				

**Pros:** Column containing duplicate tag can be used in PivotTables or other formulas to ignore or focus on duplicate rows.

**Cons:** Formula can be difficult to remember. Requires an extra column in your dataset. Could be cumbersome in large files.

# Extract Unique Values

If you want to retain the original data as it is, Microsoft 365 users and those with Office 2021 onward can use the <u>UNIQUE function</u> to extract a distinct list from multiple

# Identify Duplicates with PivotTables

A **<u>PivotTable</u>** is an excellent way to quickly identify if you have any duplicates in a column.

Place the field you want to check for duplicates in both the Rows and Values areas, in my case it's the Name field. The PivotTable gives you a list of unique names and the count:



**Tips:** sort the PivotTable Count column in descending order to bring the duplicates to the top;

right-click a cell in the values area.
 Select Sort.

3. Select Largest to Smallest :



Or filter the Count column to only show records greater than 1:



# Remove Duplicates Tool

We've looked at highlighting, extracting and tagging cells or rows containing duplicates, but sometimes you want to remove duplicates from the original data set, so all you're left with is a unique list of values. There are a couple of ways to tackle this.

Let's say we want to remove duplicate rows from the table below i.e. we want to retain row 7 with 'Produce' and 'Richard', but we want to remove one of the duplicate rows (9 or 11) containing 'Produce' and 'Rachel':

	A	В	C
4	Remove Dup	licate Rov	vs
5			
6	Department	Name	
7	Produce	Richard	
8	Dairy	Janelle	
9	Produce	Rachel	
10	Meat	Craig	
11 Produce		Rachel	←──
W	e can use the the Data tab	Remove of the rit	Duplicates tool

#### Data Review View Autor ditor Queries & Connections 絹 Flash Fill E Remove Duplicates S: Prop B Edit Links olumns 📷 Data Validation 👻 Queries & Connections Data Tools 7 Х Remove Duplicates To delete duplicate values, select one or more columns that contain duplicates. My data has headers Select All Columns ☑ Departmen ☑ Name

By selecting both the Department and Name columns I'm telling Excel that I want it to find duplicates where the values in both columns are the same. Note that I also have the 'My data has headers' box checked so it ignores my headers.

OK

Cancel

С

And I'm left with a list of unique rows:

# A B 4 Remove Duplicate Rows

### 6 Powe

**Power Query Remove Duplicates** Format your data in an **Excel Table** then load

the data into Power Query:

- 1. Select Data tab.
- 2. Select Get & Transform group ; From Table/Range:



This will load the data into Power Query and open the Power Query Editor window. In the Power Query Editor simply select the columns you want it to find duplicates for (hold **Ctrl** to select multiple columns, or **Shift** to select contiguous columns). **Then** :

1. Select Home Tab.

7

11

12

13

15

Select Remove Rows.
 Select Remove Duplicates:



Re	Remove Duplicates with Advanced Filter										
Advanced Filter can extract a list of unique items from a column or columns. 1. First select <b>the data.</b> 2. Then from <b>Data tab.</b> 3. Select <b>Advanced:</b>											
(	Data 🕗 w	View	Developer	Power Pivot	Q Tell						
Refresh All - Edit Links			$ \begin{array}{c}  A \downarrow \\  Z \downarrow \\  A \downarrow \end{array} $ $ \begin{array}{c}  Z A \\  A Z \end{array} $ $ \begin{array}{c}  Z A \\  Z \downarrow \end{array} $ Sort	CI To Re 3 To Ac	ear apply dvanced						
	A	В	C	[	)	E					
4	Department	Name					-				
5	Produce	Richard	-1	Advance	d Filter	?	×				

Action Produce Rachel O Filter the list, in-place Meat Craig Copy to another location Rachel Produce List range: SAS4:SBS9 Criteria range Ť Copy to: SES4:SFS4 1 Unique records only

	OK
--	----

**Pros:** Great for visually highlighting duplicates in a column while retaining them in the dataset. You can use filters to hide duplicates or focus on them.

**Cons:** Duplicates remain in the dataset, and that may be exactly what you want, but if you just want to get rid of them, then keep reading.

This method also doesn't highlight the row and only identifies duplicates in a single column.

	C	οι	ur	n	n

3

It's super easy to use, simply reference the column or columns you want to extract data from. The <u>UNIQUE function</u> spills the results.

	A	В	С	D	E	F	G	Н
4	Extract Unique	Rows		Formula			Formula	
5			_	=UNIQUE(	A7:B11)		=UNIQUE(E	37: <mark>B11</mark> )
6	Date	ID		Date	ID		ID	
7	1/08/2017	C301		1/08/201	7 C301	_	C301	
8	3/08/2017	C157		3/08/201	7 C157		C157	
9	6/08/2017	C301		6/08/201	7 C301		C477	
10	12/08/2017	C477		12/08/201	7 C477			
11	6/08/2017	C301						
11								

-			
6	Department	Name	
7	Produce	Richard	
8	Dairy	Janelle	
9	Produce	Rachel	-
10	Meat	Craig	
11			+

### Pros: Quick and easy to use.

**Cons:** Removal of duplicates is permanent. If your data gets updated then you need to run the Remove Duplicates process again.



In the Advanced filter dialog box (image above) choose to copy the list to another location (4 & 5), and check the box for 'Unique records only'. And voila, we now have two lists, the original, and the list excluding duplicates in columns E & F:

ĺ	A	B	С	D	E	F	
	Extract Uniqu	ue Values					
	Department	Name			Department	Name	
	Produce	Richard			Produce	Richard	
	Dairy	Janelle			Dairy	Janelle	
	Produce	Rachel			Produce	Rachel	
	Meat	Craig			Meat	Craig	
	Produce	Rachel					-



# So, there you have 7 ways

# to identify or remove duplicates.

Depending on my needs I like to use Power Query to remove duplicates, Conditional Formatting to visually highlight duplicate records and PivotTables to identify if large datasets contain duplicates.

# You can View Steps on YouTube :





Download in hi-res here: https://www.myonlinetraininghub.com/excel-find-and-remove-duplicates

 Mynda Treacy, SKILLS THAT GET YOU

 NOTICED AND PROMOTED

 in

 Follow me for tips and tutorials

