



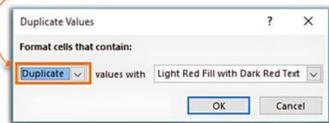
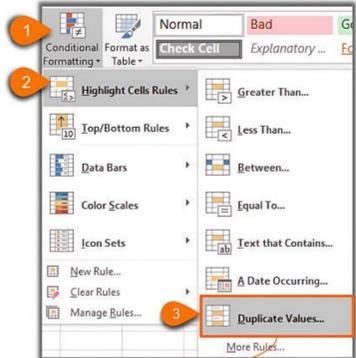
Excel Find and Remove Duplicates 7 Ways!

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

1 Highlight Duplicates with Conditional Formatting

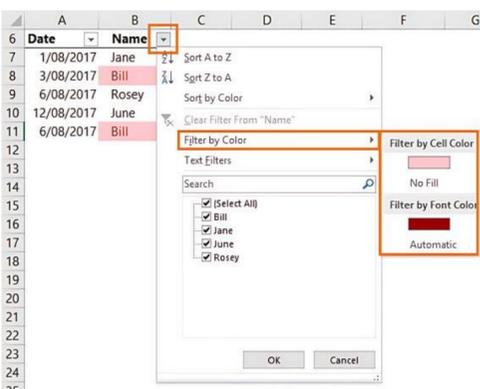
Conditional Formatting can quickly highlight duplicates in a column:

1. select the column or columns of cells containing the suspected duplicates.
2. Select Home tab.
3. Select Conditional Formatting.
4. On the Insert tab, select Highlight Cells Rules.
5. Then select Duplicate Values.



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:



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.

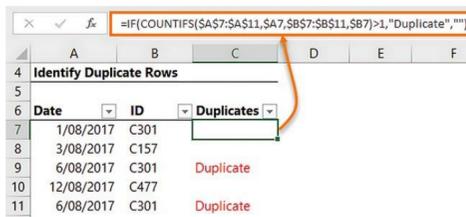
Cons: Duplicates remain in the dataset, and that may be exactly what you want, but if you just want to get rid of them, then keep reading.

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

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:

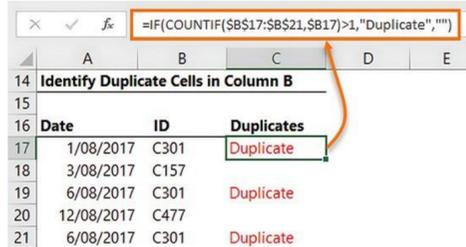


The formula in cell C7:

=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:



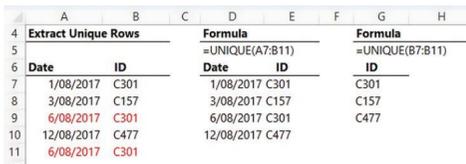
Pros: Column containing duplicate tag can be used in PivotTables or other formulas to ignore or focus on duplicate rows.

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

3 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 UNIQUE function to extract a distinct list from multiple columns.

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



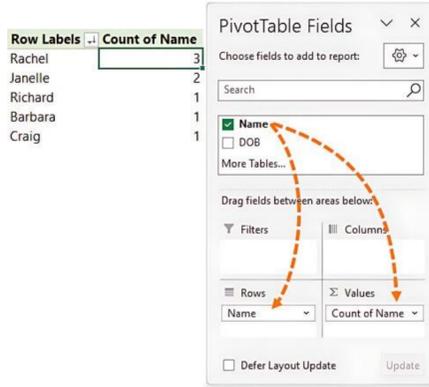
Pros: Quick and easy to use.

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

4 Identify Duplicates with PivotTables

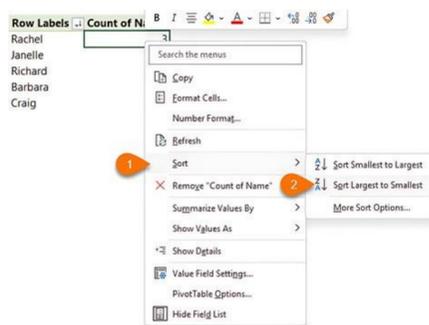
A PivotTable is an excellent way to quickly identify if you have any duplicates in a column.

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:

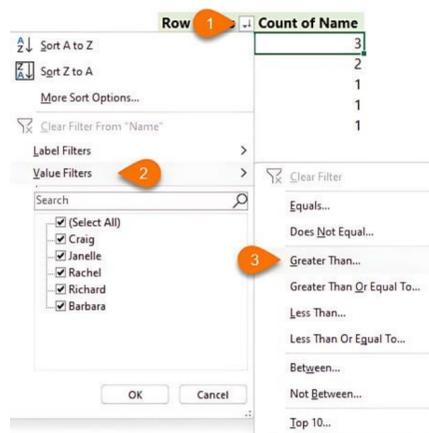


Tips: sort the PivotTable Count column in descending order to bring the duplicates to the top;

1. right-click a cell in the values area.
2. Select Sort.
3. Select Largest to Smallest :



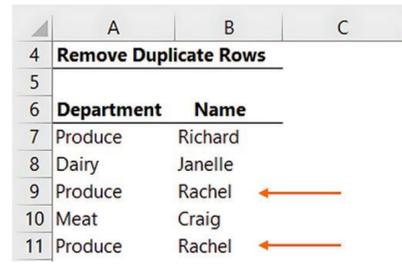
Or filter the Count column to only show records greater than 1:



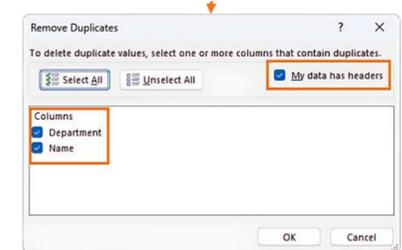
5 Remove Duplicates Tool

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':

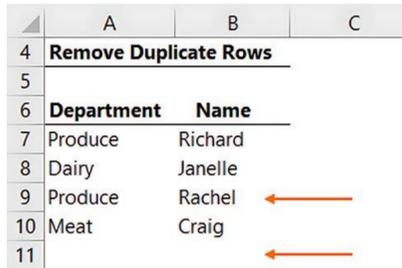


We can use the Remove Duplicates tool on the Data tab of the ribbon:



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.

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



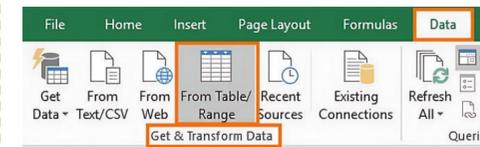
Pros: Quick and easy to use.

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

6 Power Query Remove Duplicates

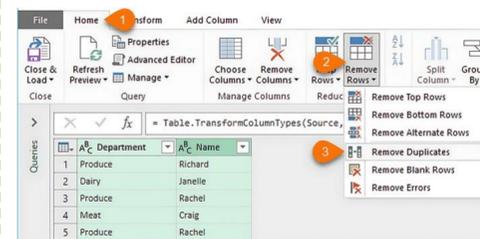
Format your data in an Excel Table then load the data into Power Query:

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 :

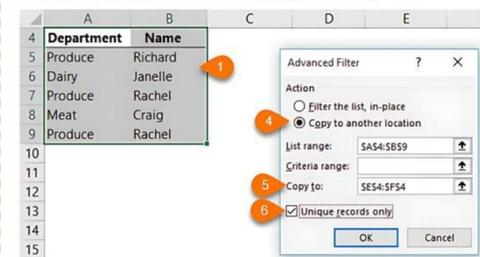
1. Select Home Tab.
2. Select Remove Rows.
3. Select Remove Duplicates:



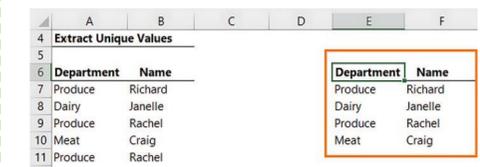
7 Remove Duplicates with Advanced Filter

Advanced Filter can extract a list of unique items from a column or columns.

1. First select the data.
2. Then from Data tab.
3. Select Advanced:



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:



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.



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

Mynda Treacy, SKILLS THAT GET YOU NOTICED AND PROMOTED

Follow me for tips and tutorials

myOnlineTraininghub

© Copyright 2023

You can View Steps on YouTube :

