

HOW TO REMOVE DUPLICATE ENTRIES FROM A DATASET IN EXCEL

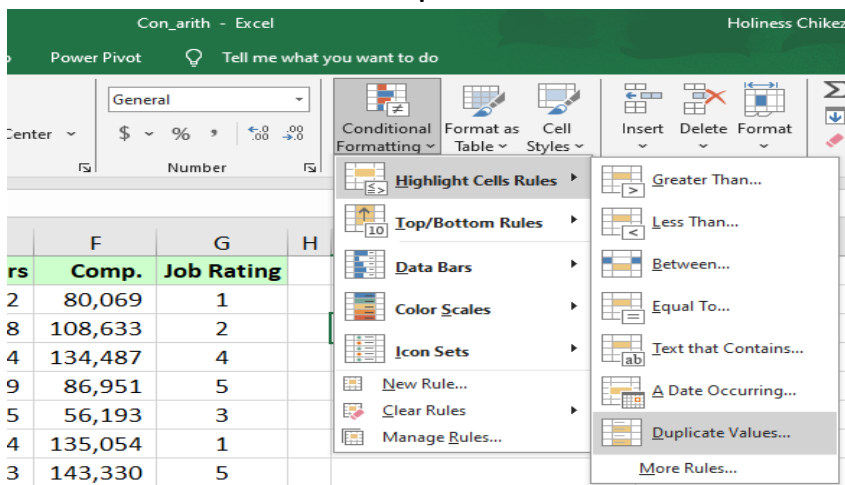
Duplicates are bound to be found in large datasets, especially when people have access to the same document, this makes data redundant. The larger the dataset the higher the chances of encountering duplicate records.

Before removing duplicate values, it is wise to find, highlight and review them before removing them.

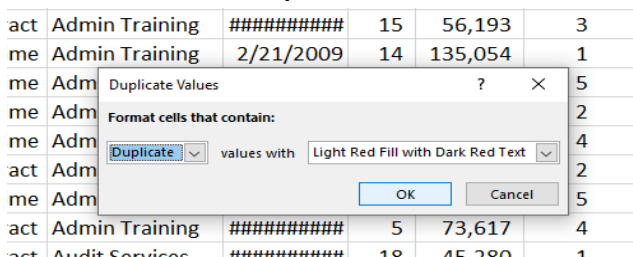
IDENTIFYING DULICATES

USING CONDITIONAL FORMATTING

1. Select the range of cell you want to check for duplicate.
2. On the Home Tab, click on conditional formatting, hover around the arrow and select duplicate values.



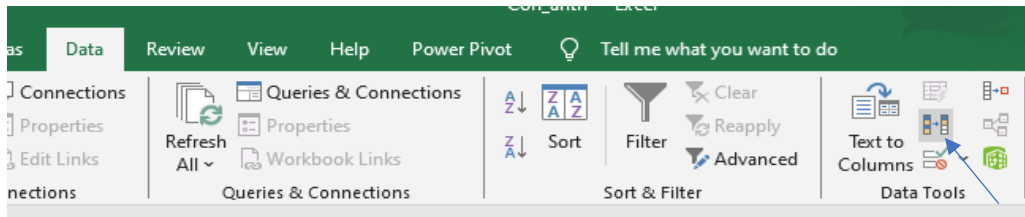
3. Customize how you want the values to appear.



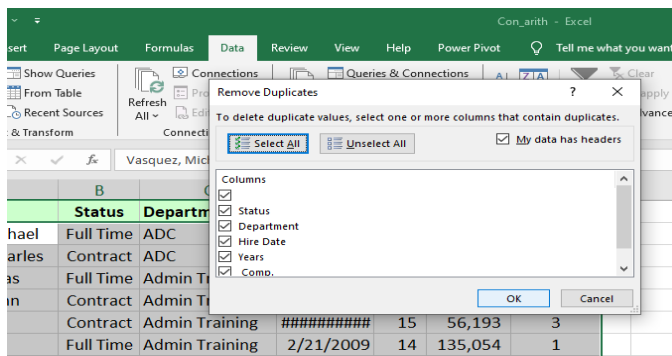
4. Click ok.

REMOVING DUPLICATES

1. Select the range of cells you want to check for duplicates.
2. Go to 'Data' tab, click on 'remove duplicate option'.

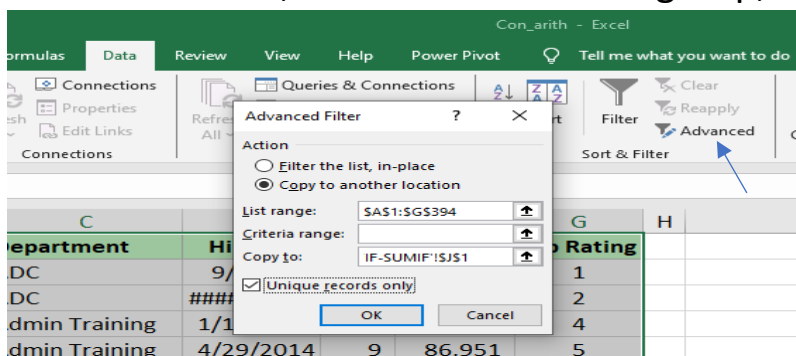


3. Check all the columns you want to remove duplicates from, and click ok.



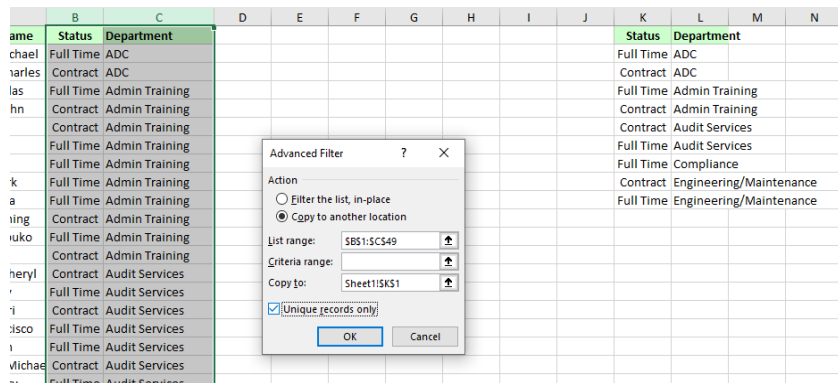
USING THE FILTERING METHOD

1. On the Data tab, in the sort and filter group, click advanced.



2. The advanced dialogue box appears, click copy to another location.
3. Select the range of cells you want to check for duplicates.
4. Click on the copy box and select on the sheet where you want your values to appear.

5. Check unique records only, and click ok.



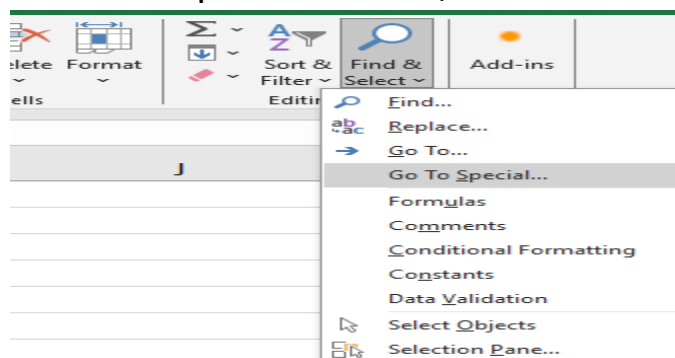
HOW TO REMOVE BLANK ROWS FROM A DATASET USING EXCEL.

One of the most common issues that arise when working with data is blank rows. These empty rows can disrupt your analysis and make your data hard to understand. But before removing them, we have to identify them.

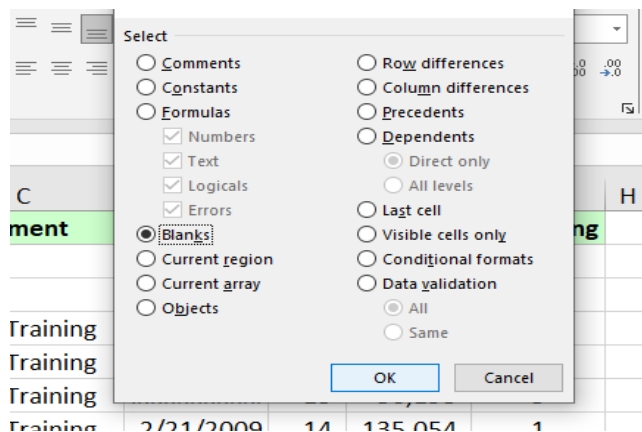
In this article, we will go over the step-to-step process of how to identify and remove blank rows.

IDENTIFYING BLANK ROWS USING THE SPECIAL FEATURE.

1. On the Home tab, click on Find and Select option.
2. On the drop-down menu, click on 'Go-to-Special'.

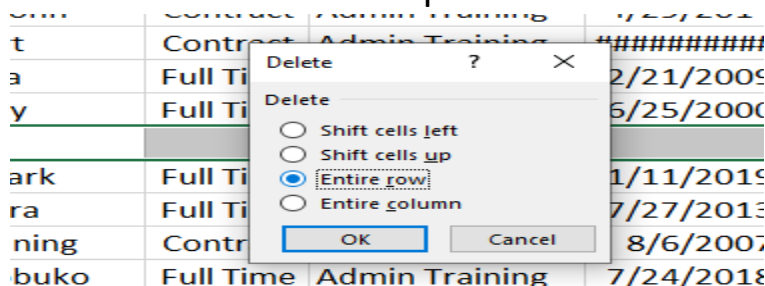


3. In the dialogue box, select 'blanks' and click 'ok'.



REMOVING BLANKS USING 'GO TO SPECIAL'

1. Select the entire sheet.
2. Click the 'Find and Select' option on the home tab.
3. Select 'Go to Special', click blanks and ok.
4. Right click on any of the highlighted cells and click delete.
5. Check the 'entire row' option and click ok.



FORMULA METHOD

1. Add a new column to your sheet and label it 'blank'.
2. In the first cell of the blank column, enter this formula
'=IF(A2="",0,1)' click enter.

=IF(A2="",0,1)								
	A	B	C	D	E	F	G	H
1		Status	Department	Hire Date	Year	Comp	Job Rating	BLANK
2	Vasquez, Michael	Full Time	ADC	9/6/2011	12	80,069	1	=IF(A2="",0,1)
3	Woodard, Charles	Contract	ADC	#####	8	108,633	2	
4	Ayers, Douglas	Full Time	Admin Training	1/16/2000	24	134,487	4	
5	Cameron, John	Contract	Admin Training	4/29/2014	9	86,951	5	

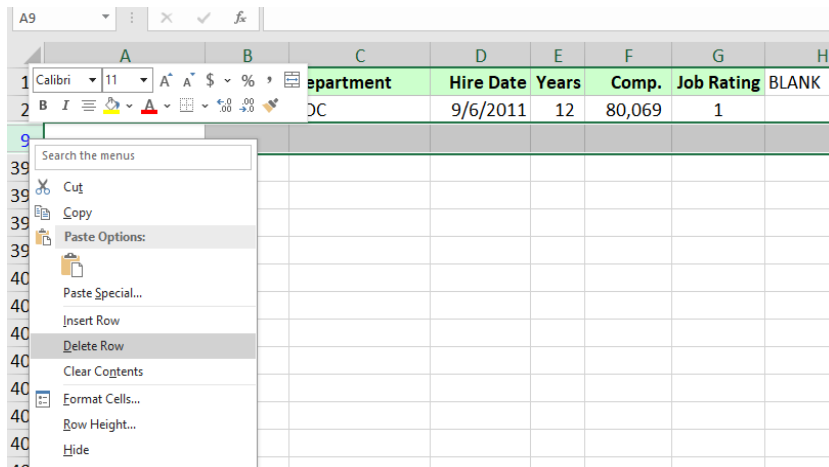
3. Autofill the entire column, by double clicking on the small box.

F	G	H
Comp	Job Rating	BLANK
80,069	1	1
108,633	2	
134,487	4	

4. Filter out the rows with '0' value in the blank column.

#	8	1	Sort Smallest to Largest
0	24	1	Sort Largest to Smallest
4	9		Sort by Color
#	15		Sheet View
9	14	1	Clear Filter From "1"
0	23	1	Filter by Color
9	5	1	Number Filters
3	10		Search
7	16		<input checked="" type="checkbox"/> (Select All)
8	5	1	<input checked="" type="checkbox"/> 0
#	5		<input type="checkbox"/> 1
#	18		
1	22		
9	24		
6	17		

5. Select the filtered data and delete the rows.



P.S: Is best to save a copy of your original data before making changes, so you can easily refer back when needed.