

The Power of Excel 2003: Identifying Duplicates



For years I searched for a reliable way to identify duplicates in an Excel spreadsheet. I tried everything I knew, including MS Access's "Find Duplicates Query Wizard" - to no avail. Finally, through networking at a conference, I came upon the answer below. Finding duplicates in Excel a three step process: inserting a new column with a formula, and then (here's the key), turn Automatic Calculation OFF and Manual Calculation ON so that when you Sort on the new duplicates column, the formula does not recalculate automatically. Lastly, once you've identified the duplicates and deleted them (or copy/pasted the unique's), you delete the column with the formula and turn Automatic Calculation back on.

Here's what to do:

- **Sort the datasheet by the field that has possible duplicates** (EmplID, SSN, for example) and then **Insert a column** to the left of the field that has the possible duplicates.
- In the 1st cell in the new column **type a heading** (Dups) adjacent to the duplicated field (SSN) field header. Then **type the following formula**: =IF(B2=B1,"DUP","") in the cell adjacent to the 1st possible duplicate, substituting your column name (A, B, C) in the formula. **Note**: the last part of the formula after the coma is double quotes ("").
- **Copy the formula** down to the appropriate cells – **Here's a cool Excel Trick**: place your mouse on the fill handle (the black square at the bottom right corner of the first formula), and when you see the thin cross-hair, double click the mouse. Excel will "auto-copy" the formula all the way down until it finds a blank space in the column to the left of the formula column. This will save you from having to manually copying the formula!
- **Select Tools, Options** from the **Menu Bar**.
- **Select the Calculation Tab** (as shown below).
- Click to **enter a checkmark in the "Manual"** choice for the Calculation Tab of that dialog box. This HAS to be done **before** sorting!
- Click **OK**. If desired, you can now scroll through your list to view all dups before sorting.
- Click on any cell in the column with the formula in it (Dups) and **Sort** by that field.
- All the duplicates will be on the top – Delete if desired.

The screenshot shows an Excel spreadsheet with columns A, B, and C. Column A contains a formula =IF(B2=B1,"DUP","") in row 2, which is circled in red. The spreadsheet lists names in column B: LAST NAME, Actiony, Adukal, Akonmade, Alakigaa, Anderson, Anyenvui, Applas, Arriemel, Autachetchakul, Badepur, Baillargeoni, Baker, Venilla, Vartouhi, Vanessa, and Valerie. The Options dialog box is open, showing the Calculation tab with the Manual radio button selected and the Recalculate before save checkbox checked.

This view is gotten with Control + ~ (Control tilde)

	A	B	C
1		LAST NAME	
2	=IF(B2=B1,"DUP","")	Actiony	
3	=IF(B3=B2,"DUP","")	Adukal	
4	=IF(B4=B3,"DUP","")	Akonmade	
5	=IF(B5=B4,"DUP","")	Alakigaa	
6	=IF(B6=B5,"DUP","")	Anderson	
7	=IF(B7=B6,"DUP","")	Anderson	
8	=IF(B8=B7,"DUP","")	Anyenvui	
9	=IF(B9=B8,"DUP","")	Applas	
10	=IF(B10=B9,"DUP","")	Arriemel	
11	=IF(B11=B10,"DUP","")	Autachetchakul	
12	=IF(B12=B11,"DUP","")	Badepur	
13	=IF(B13=B12,"DUP","")	Baillargeoni	
14		Baker	
15		Baker	Venilla
16		Banks	Vartouhi
17	=IF(B17=B16,"DUP","")	Bardoff	Vanessa
18	=IF(B18=B17,"DUP","")	Barr	Valerie

Note: Once you have completed this and deleted your duplicates, it's a good idea to go back to **Tools, Options** and **change** the calculation back to **Automatic!**