

EXCEL <<<<<<<

By Bill Jelen

Excel 2013 Pivot Tables Offer Distinct Count

Getting a count of the unique values in a pivot table has always been difficult. Although you can easily build an Excel 2010 pivot table that is labelled “Count of Customer” for each separate sector you might have, this is really a count of the number of customer records. If you have 65 records for XYZ customer, it will be reported as 65 customers instead of one unique customer.

Excel 2013 offers an easy work-around, but only if you make use of the new Data Model feature. Most pivot tables are created using a pivot cache in memory. With Excel 2013, you have the option to create your pivot tables from the Vertipaq data model engine instead of a pivot cache. Among the many benefits of the data model is the ability to create a distinct count for a field.

To send your data to the data model, select a cell in your data set. From the Insert tab, choose Pivot Table. In the Create Pivot Table dialog, click the checkbox for “Add this data to the Data Model.” Click OK. This will give you a blank worksheet with a seemingly normal Pivot Table Fields dialog. The only indication that something is different is the addition of “ACTIVE | ALL” at the top of the

field list.

Now build the pivot table as you usually would. Drag Sector to the Rows area. Drag Customer to the Values area. The pivot table will show you the count of rows but not a distinct count of customers (see Figure 1).

Next, select one cell in the Count of Customer column. From the Pivot Table Tools Analyze tab, click on Field Settings. The Value Field Settings dialog appears, offering what appears to be the normal list of calculations: Sum, Count, Average, Max, Min, StdDev, and so on. One item missing, however, is the seldom-used Product choice. Scroll down to the bottom of the list, and you’ll see Distinct Count (see Figure 2). Select it. The result is the pivot table shown in Figure 3. Each sector reports the number of unique customers.

While this is a great trick in Excel 2013, the workbook isn’t backwards compatible with Excel 2010 or earlier versions. If you try to open the workbook in a previous version, you will encounter various messages about links to external data and invalid names. If you click through all of those and get to the pivot table, choices will be grayed out.

When you try to move a field in the pivot table, Excel 2010 will finally tell you that the pivot table was created in a later version and can’t be updated in Excel 2010.

Earlier Versions of Excel

If you need to calculate a distinct count in Excel 2010 or earlier, you’ll have to add a new field to the original data set and use formulas to help get the count. All of these formulas are fairly complicated, which makes the inclusion of Distinct Count in Excel 2013 a welcome addition.

There are a lot of formulas that can be used for this. One common approach is to count the number of times a customer appears in the data set and divide that into 1. So, if a customer appears five times, each occurrence will get the value of 1/5, which, when summed in the pivot table, will yield a total of 1 distinct customer. Assuming your customers are in cells D2:D500, you would use `=1/COUNTIF(D2:D500,D2)` to calculate the fraction for each record.

Another approach is a formula that

Figure 1

| Row Labels | Count of Customer |
|--------------------|-------------------|
| Communications | 84 |
| Energy | 78 |
| Financial | 84 |
| Healthcare | 4 |
| Manufacturing | 184 |
| Retail | 77 |
| Technology | 52 |
| Grand Total | 563 |

Figure 3

| Row Labels | Distinct Count of Customer |
|--------------------|----------------------------|
| Communications | 4 |
| Energy | 4 |
| Financial | 4 |
| Healthcare | 1 |
| Manufacturing | 7 |
| Retail | 4 |
| Technology | 3 |
| Grand Total | 27 |

looks at the customer field and detects if this is the first time that customer appears in the data set. For the first occur-

rence of each customer, apply a 1. For each additional occurrence, indicate a 0. One slow formula to do this is

$=IF(COUNTIF(\$D\$1:\$D1,D2)=0,1,0)$. The important part of this formula is the missing dollar sign ($:\$D1$) in the first argument of the COUNTIF function. This will create an expanding range that evaluates all rows from row 1 to the row above the current row, looking for other records that have this customer.

A faster formula to do this is $=(MATCH(D2,\$D\$2:\$D\$500,0)=ROW()-ROW(\$D\$2)+1)*1$. This formula finds the first occurrence of the customer and checks if the row number is equal to the current row number. While this formula is more difficult to follow, it calculates five times faster than the previous formula and 10 times faster than the first formula. **SF**

Bill Jelen is the host of MrExcel.com. He is currently on the road, taking his Power Excel seminar to IMA chapters across the country. Send questions for future articles to IMA@MrExcel.com.

Figure 2

