How to Cleanse Data in Excel

By Rich Malloy, Tech Help Today ([www.techhelptoday.com](http://www.techhelptoday.com))

When you import data into Excel, you often must “clean” the data to match certain requirements of your spreadsheets. Fortunately, Excel includes many tool for transforming data into more suitable formats.

We’ll be using a set of data files packaged in a Zip folder. To get the set of data files, download the following zipped folder: [XXXet-and-Transform-Data-Files.zip](https://drive.google.com/open?id=0B72BoLO2D_HFWEdULXg3TmJhQkU)

# Find & Replace

The venerable Find & Replace tool is not an Advanced feature of Excel, but it nevertheless comes in very handy for cleaning data. It is particularly powerful in that it can process a whole worksheet in a fraction of a second. By using a Macro, you can have this tool make a number of different replacements in a single mouse stroke.

Because most people already know how to use this tool, we will refer to a resource from Microsoft: <https://support.office.com/en-us/article/find-or-replace-text-and-numbers-on-a-worksheet-0e304ca5-ecef-4808-b90f-fdb42f892e90>

# The Substitute Function

If you need to replace one set of characters with another set automatically, you can use the SUBSTITUTE function.

### Description

**SUBSTITUTE( *Text*, *Old\_text*, *New\_text*, [*Instance\_num*] )**

*Text*: Text or cell that contains the text that needs to be replaced

*Old\_text*: Text that should be replaced

*New\_text*: The replacement text

*Instance\_num*: Optional; specifies which occurrence of the Old\_text should be replaced. If not specified, all occurrences will be replaced.

### Example

**Formula:** =SUBSTITUTE("The colour of the neighbour’s car", "ou", "o")

Result: The color of the neighbor’s car

*Note: If you know the position of the text to be replace, but not the actual text itself, you can use the REPLACE function.*

# Remove Unwanted Characters

Excel does not have a REMOVE or DELETE function, but we can make do with the SUBSTITUTE function. All we have to do is substitute nothing for something:

### Example

**Formula:** =SUBSTITUTE("98.6 degrees", " degrees", "")

Result: 98.6

# Remove Extra Spaces

Sometimes data has too many spaces. You can easily trim off the spaces with the TRIM function. The TRIM function converts any group of 2 or more spaces between words to one space and deletes all leading and trailing spaces.

### Example

**Formula:** =TRIM(" too many spaces ")

Result: "too many spaces"

# Display Duplicate Rows

Duplication can be annoying. Duplication can be — well, you get the idea. By adding two columns, you can quickly see which rows are duplicates.

### Procedure:

*The following assumes you have a range of data in columns A to D, and rows 5 to 30. In column E we will create a formula to concatenate the contents in columns A to D into one cell. Then in the next column (F) we will have a formula that counts how many times each cell in column E appears in that column. Note that this technique marks all copies of duplicate rows, including the first copy, which is technically not yet a duplicate.*

In cell E5, enter the formula: = A5 & B5 & C5 & D5

In cell F5, enter the formula: =COUNTIF(E:E, E5)

Copy these cells to the rest of the rows in these columns.

All duplicate rows will have a count greater than one.

# Remove Duplicate Rows

*Excel has a nice tool for deleting duplicate rows.*

### Procedure:

Select the range of cells that might contain duplicate rows.

In the Data tab, look in the Data Tools group, and click Remove Duplicates

If needed, uncheck any column that do not need to be checked for duplicates.

Click: OK

*All rows containing duplicate data will be deleted.*

# Split Text

There are three ways to split the text in a column of cells: The Old Way, the Formula Way, and the Amazing New Way. These techniques are very handy when you need to separate the first name from the last name in a cell.

### Text into Columns

*This tool can be used to separate, say, the first 5 characters from a cell, but most often it is used to separate words that are already separated by spaces. Note: to avoid overwriting data, be sure that there are a sufficient number of blank columns to the right of the column that needs to be split.*

Select the cells in a column that need to be split

In the Data tab, look in the Data Tools group, and click Text to Columns

Click the option button: Delimited

Click: Next

Check the checkbox for Spaces

Click: Finish

### Split Text Automatically by a Formula

*For this procedure, let’s say that cell A5 contains “John Smith”. Our goal is to get the first and last names in cells B5 and C5, respectively.*

*We’ll use the LEFT function to get the first name, and the RIGHT function to get the last name. To determine how many characters we will specify for each of these functions, we’ll use the FIND function to locate the space between the names. The location of the space character (" ") is 5, which will always be one more than the length of the first name.*

*To find the length of the last name, we need to find the length of the full name with the LEN function. Then, if we then subtract 5 from the length of the full name, we will know the length of the last name.*

Click in a blank cell to the right of the cell that needs to be split (i.e., B5)

Enter the formula: =LEFT(A5, FIND(" ", A5) – 1)

Click in the next cell to the right (i.e., C5)

Enter the formula: =RIGHT(A5, LEN(A5) – FIND(" ", A5) )

# The Amazing Flash Fill Tool

*The third way to split text in a cell is the Flash Fill tool, which not only can split text, but can also handle a number of other text manipulations. The best part of this tool is how it seems to guess what you want.*

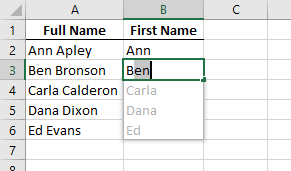
### Procedure:

Make sure that cells B2:B6 are empty

In cell B2, enter: Ann

In cell B3, type: B

*Flash Fill now seems to guess that you want to put first names in the rest of the cells in that column (see screenshot below).*

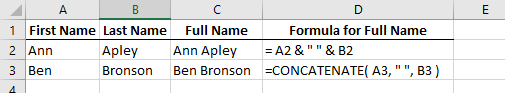


Press the Enter key to confirm that that is your intention.

*Note: Another way to initiate Flash Fill is to use the Flash Fill button in the Data tab.*

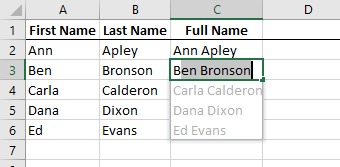
# Combine Text

This is the reverse of the Split Text procedure. A simple concatenation formula can do this automatically, or we can use the Flash Fill tool. There are two possible concatenation formulas:



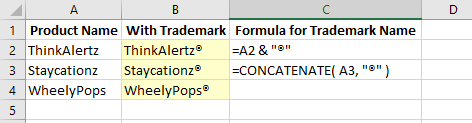
### Combining Text Via Flash Fill

The procedure is very similar to that for splitting text. Again, before starting, be sure that the destination cells (B2:B6) are empty.



# Add Text to Column

In this case you want to add the same text to every cell in a column. The processes are very similar to those for combining text. In this example, we are adding the Registered Trademark symbol to all the product names in the column A.



### Adding Text with Flash Fill

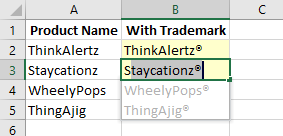
As before, make sure that all the destination cells (B2:B5) are empty

In cell B2, enter ThinkAlertz®

In cell B3, type S

Flash Fill will suggest entries for the remaining cells in the column

Press the Enter key to confirm those suggestions



# Converting Measurement Units

The CONVERT function can convert measurement units easily. For example, ounces to grams, centimeters to inches, or Centigrade to Fahrenheit.

### Example

**Formula:** =CONVERT(100,"cm”,”in”)

Result: 3.343

[End]