About data cleaning

When working with data in Excel, one of the first things you should do is understand the data you have. What kind of data is it? How is it structured? Is it good quality data, or does it need cleaning up? This is often called “exploring” your data. You want to understand the characteristics of your data and also its limitations.

Exploring the data lets you look for trends or patterns as well as operationally or financially significant values, like negative or positive sales growth, an underwater stock option value, a negative inventory level, or any items sold prior to a certain date.  Becoming aware of such characteristics of the data would then let you ask questions about the source and quality of the data. Does a negative inventory level signify a database that reflects order quantity greater than fulfilled quantity as a negative inventory level? Or does the dataset or its source have errors in it?

When exploring your data, some of the things you can look for include

* Missing values
* Repeated / duplicate values
* Errors in the data or in calculations that drive values in the data
* Trends and patterns
* Values that are negative or positive
* Values that are larger or smaller than a cutoff value
* Values that occur in a certain date range

Conditional formatting is a technique in Excel to make data exploration visual. You format the data (its colors, fonts, highlighting, etc.) based on the criteria you select. This allows you to quickly identify values that are either a data quality problem (missing values, formula errors, nonsensical values, etc.) or a value that drives insights and decisions (i.e., sales growth below a set target; inventory below a set amount, which would trigger re-ordering; etc.)

Excel’s conditional formatting tools are easy to use but have a large and powerful array of options for criteria to base formatting on, including the ability to write your own formulas. You can even use the conditional formatting tools to format colorful and visually easy-to-read reports and simple dashboards of your data, creating visuals such as “heat maps” and even “Harvey ball” rankings to visually indicate data that meet specific criteria.

If you are experienced with Excel’s conditional formatting capabilities, then the task below will be very straightforward. But if you have not spent much time immersed in conditional formatting, now is the time to learn these techniques.