

How to analyse logistics data

I've been asked about **how** I analyse logistics and other data, as a result of previous articles, and it's a thought-provoking question! I'm obviously not going to replace a textbook on the subject with one article, but here goes – my thoughts on the subject.

I tend to be analysing data for two reasons - looking back, understanding what has happened and why, and then using those lessons to look forward. To what extent can what has already happened be used for planning the future through extrapolation? In a day-to-day operation, you will also be interested in the here and now – are you ahead of your targets, and do you have alerts in case they may not be achieved?

Getting your ducks in a row

Any analysis has to start with the data. The first thing you need to do is to understand the data – what have you got, and what sort? There are various sorts of data: IDs, like account or customer numbers; categories such as product types, people or the originating DC; numbers like the number of orders in a day, or volumes; locations such as customer addresses or warehouse slots; dates & times, which is an interesting type as the data can be both sequential and also a category, such as day of week or a particular shift.

The next thing you need to do is some preliminary analysis to understand the quality of the data – is that outlier an error or is it really important! Not going to say too much here as I know I've already mentioned this topic in SHD Logistics in June last year, but do keep a history diary, an incident log or whatever you want to call it – noting any specific reasons affecting that day's operation: bank holidays, weather, sport, promotions, traffic...

But some of those outliers are not down to some operational quirk, they are actually mistakes. I had to query a case quantity that had been input as over 1Bn. When someone at the logistics provider told me that the key had got stuck down, they plainly weren't expecting me to believe them, but I was able to tell them that I'd been in the same situation after accidentally spilling Coca-Cola in my keyboard!

Something else worth noting when preparing data. Are there any fields that you could potentially use to connect to other data? Date is one of the obvious ones. You might be able to link your data to other sources including external ones, giving the values of key variables, for example: exchange rates, diesel, interest rates, inflation...

I'll leave maps and plans to one side for this article, and I'll talk about graphs in a moment. But just while you have your data in an ordinary Excel spreadsheet (particularly report-style worksheets, perhaps created using pivot tables), you have features like conditional formatting and sparklines which can add visual analysis without doing very much work – always a bonus!

Different ways of looking at the same thing

In logistics we look at a lot of data sequentially, creating graphs of volumes, metrics, or performance indicators with a date or time axis along the bottom. It's easy these days, to put one or more of those measures on a dual axis – a second measure of size on the right hand side of the graph, aligning the two measures roughly on the graph, so you can match the peaks and troughs. If I do that, I find it helpful to include 'LH' and 'RH' as part of the key/legend, so that the reader can quickly see which measure is on which axis. I don't always follow my own rules, but it's also helpful to make

the font on the axis the same colour as the line on the graph, again helping that boss who isn't as data-savvy as you.

Creating your graph with date along the bottom is great for showing trends – how stuff is changing over time - but I'm often interested in the relationship between two sets of data regardless of the date. In those situations, scatter graphs can be really useful, particularly for forecasting – how does the number of roll cages to process relate to the sales forecast, regardless of what day it happened. It's important to consider cause and effect before creating your graph. It's usual to put the factor that is thought to be the cause or driver along the bottom axis (in this case the sales) and then the effect (the number of cages to process) up the side axis.

If you do this, you can then use the trend function in Excel, or similar, to give you the equation that relates cages to sales (and if you're going to use this as an equation in a formula, make sure you format the trendline label to give you more digits after the decimal point than is automatically displayed). You can also get another useful figure from the trendline, R-squared. This gives you a feel of how closely related your two sets of figures are. Zero means totally random, whereas you would only get '1' if all the points on your graph were miraculously on a straight line. So 0.95 would tell you that the cage numbers are very closely related to sales, whereas 0.35 would tell you that using the sales forecast to predict cages would be a poor decision. NB R-Squared has nothing to do with the R you hear about with reference to Covid.

Maxing the content

Graphs can tell you an enormous amount about your data and may be able to tell you even more if you put some thought into how you set them up. In most desk-top analytic apps, such as Tableau, you can get 5-way analysis easily – using the X-axis, Y-axis, Colour, Size and Shape for different attributes. Colour by Day of Week, so you can see if Sunday is very different from other days. Use different shapes for different DCs. Even the axes can be compound, for example, roll cages per hour, although sometimes less is more!

And that's a quick version of how I tend to go about analysing logistics data.

Kirsten Tisdale
26 February 2021 for April 2021

NB There was no specific graphic used for this article – a generic stock image of a graph was used as the illustration, so this has been omitted from this version

Here is the article on the web archive:

<https://web.archive.org/web/20210226155216/https://www.shdlogistics.com/data-analysis/how-analyse-logistics-data>