DRIVER ANALYSIS
ARE YOU ANSWERING THE RIGHT QUESTION?

"Driver analysis" is a term which describes a set of related techniques that can be used to help organisations know which elements of the customer experience have most impact on crucial outcomes such as overall satisfaction, recommendation or NPS, and loyalty behaviours such as retention.

The different techniques available have different strengths and weaknesses, and unfortunately customer experience data has some unique properties that make it difficult to analyse effectively. That means that many organisations end up using an approach that is not really best suited to the business question they want to address.

In this article we'll look at how driver analysis works, and how the different strengths and weaknesses of different approaches mean that choosing the right one depends on being clear on what question you're asking. We'll start by explaining what driver analysis is, and what some of the business questions it can help you address are. We'll also take a quick not-too-technical look at how it works. Finally, we'll finish with some recommendations as to what a gold standard approach would consist of.

As specialists in customer experience research, we spend a lot of time analysing customer data in order to help our clients understand what the key drivers are. Depending on the needs of each client that analysis ranges from the very simple

to the very complex, but over time we've settled on a few favourite tools that work reliably well.

What is it?

What exactly is driver analysis? People talk about it as a singular thing, but the truth is that a whole range of different statistical techniques are commonly used under the umbrella term "driver analysis". The idea is to understand how much impact different aspects of the customer experience have on some outcome variable. Often that outcome is overall satisfaction or intention to recommend, but there's no reason that it couldn't be any variable you're interested in, such as retention or share of spend.

But what exactly do we mean by impact? If we break it down to the simplest possible example, what we're looking at is the relationship between two variables, which we'll call "X" and "Y". You can see in figure 1 that we have 8 pairs of scores for X and Y. Is there a link between them? Well they're not identical, but you can probably see that people have tended to give similar scores – either both low, both middling, or both high.

It's a lot easier to see what's going on if we plot those pairs of points on a scatter graph, with the values of X on the X axis and the values of Y on the Y axis, so that we can see how the two variables work together. How strong is the link? If we put a line of best fit through the dots we can see that there's a strong relationship (the line is nearly at 45

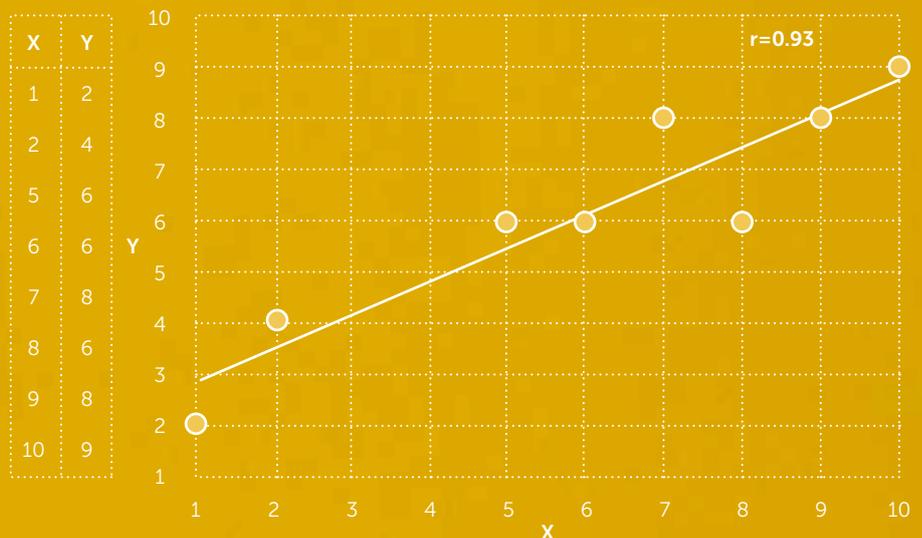| X	Y
1	2
2	4
5	6
6	6
7	8
8	6
9	8
10	9

r=0.93

Fig.1: Correlation between 2 variables

degrees), and the points are clustered close to the line. We can summarise that relationship in a single number – the correlation coefficient, which is referred to as "r", and varies from 0 to 1. In this case the correlation, albeit on a very small sample size, is extremely strong at over 0.9.

Correlation underpins all of the more complex techniques we'll look at, so it's important to wrap your head round the basics of how this works. A strong relationship doesn't necessarily mean people giving precisely the same score for both variables, it means there is a consistent pattern in the way people score them…in statistical terms they tend to "covary", or move together.

That was a highly simplified example. What does it look like in practice? Something more like figure 2, with a lot more data, and a lot more noise in the relationship. You can see a few people who have scored 1 for one of the variables and 10 for the other. Overall, though, there is again a tendency for people to score X and Y in a similar way. That tendency gives us a correlation coefficient of 0.4, a low to moderate correlation in customer research.
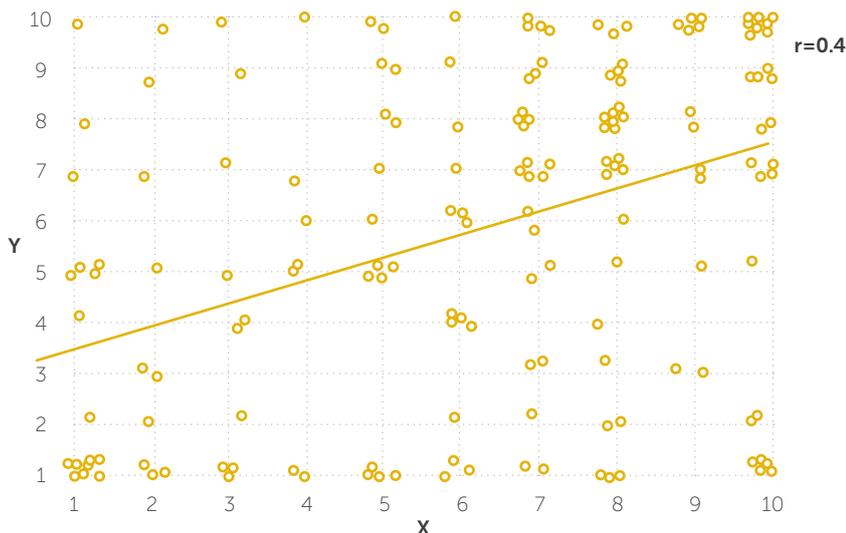

Fig.2: A real-world correlation

We can interpret that in three ways:
- *First, we can say that X has a significant, but not strong, impact on Y. Improving X should lead to an increase in Y, but there's a good chance another predictor would have a stronger impact.*
- *We could use it to make predictions, so if I know that your score for X is 8, I know that your score for overall satisfaction is likely to be between 6 and 7.*
- *We could also use this chart to answer "what if?" questions, specifically "If I increase the score for X by 1pt, how much would Y go up by?"*
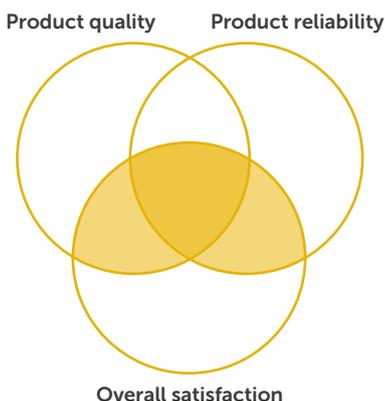

Fig.3: Correlation as shared information

Simple correlations like this are extremely useful…but they do have a fatal flaw. What we're doing with correlation is looking at the information in our outcome variable, in this case overall satisfaction, and the information in our predictor variable, let's say product quality.

You can think of the correlation coefficient as the proportion of that information that is shared between variables. If we repeat the exercise for another predictor, let's say product reliability, we'll find that as well as correlating with overall satisfaction, there is also a strong correlation between the two predictors, as shown in figure 3. Which makes sense, doesn't it? You're unlikely to think my products are great quality if they keep breaking down.

That means that some of the overlap with overall satisfaction ends up being double counted. This is a phenomenon known as "multicollinearity", and it's the curse of customer research. In practice, every single item on the questionnaire is correlated with every other item.

The real art of driver analysis is to unpick all this overlapping information and give the credit to the correct predictors. So, what is driver analysis? It's a set of tools designed to

help you pick apart that mess, so that you can understand the strength of the relationship between your outcome variable and the set of predictors.

It will give you a clear sense of which of the predictors is most important, and the ability to predict what effect changes to any of those predictors would have on the outcome.

What questions can driver analysis address?

Those statistical questions translate to important business questions. Unfortunately it's not always easy for organisations to articulate which questions they are most interested in.

There are basically three questions that driver analysis can help you with:
- *What's important to customers?*
- *What's making a difference?*
- *What would happen if we changed the predictor (perhaps by improving satisfaction or reducing hold times)?*

Those are all good questions, but unfortunately they are not necessarily best addressed by any one single technique. Let's have a look at some concrete examples, and how they relate to the questions we might be interested in.

Importance

If you want to know what matters most to customers, then the best way is to ask them – stated importance tells you what matters. So sometimes the best approach to driver analysis may be not to use driver analysis!
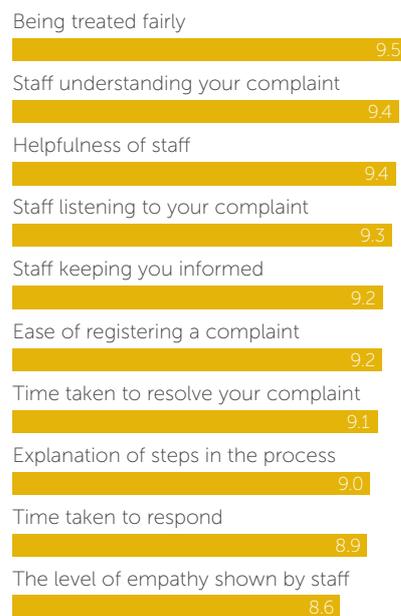

Fig.4: Stated importance

Impact

If you want to understand the strength of the links between each variable and overall satisfaction, then a simple correlation (like we've already looked at) is a good bet. It's easy, doesn't need huge sample sizes, and requires you to make very few theoretical assumptions about the nature of the data (which becomes more of a concern with some more advanced techniques).
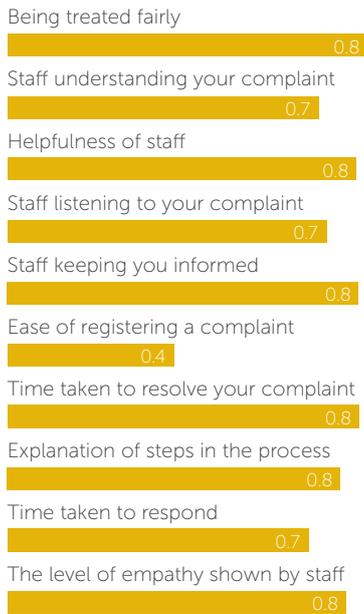
Being treated fairly
| 0.8 |

Staff understanding your complaint
| 0.7 |

Helpfulness of staff
| 0.8 |

Staff listening to your complaint
| 0.7 |

Staff keeping you informed
| 0.8 |

Ease of registering a complaint
| 0.4 |

Time taken to resolve your complaint
| 0.8 |

Explanation of steps in the process
| 0.8 |

Time taken to respond
| 0.7 |

The level of empathy shown by staff
| 0.8 |

Fig.5: Impact/correlation

Correlation puts an upper bound on how strong the link between each predictor and the outcome is and, if you have a good questionnaire, in most cases the conclusions you draw from it will not be radically different from more complex techniques.

A rounded view

Combining importance and impact (which we use as a synonym for correlation) into one chart, as in figure 6, gives a useful picture of both what matters most to customers and what's making a difference right now. Dividing the chart into a quadrant can identify areas which are high in importance but low impact. We call those "Givens" or "Hygiene factors", meaning that as long as there is not much dissatisfaction there is probably little to be gained from investing in further improvement.

Where stated importance tends to be relatively slow to change, impact coefficients (and other driver analysis techniques) are often very dynamic – reacting to changes in your performance and that of your

Fig.6: Importance & Impact combined

competitors. This matrix view helps you to understand how customers see all aspects of their experience right now.

Breaking down the drivers

The issue with correlation, as we've seen, is multicollinearity. There is a danger of double-counting some of the impact on overall satisfaction, particularly if the questionnaire includes too many questions which are very similar. This often happens, for instance, when the questionnaire has a large number of items about staff (friendliness, helpfulness, professionalism, etc.).

There is a statistical tool which aims to tackle this, and give us a measure of impact when looking at all the predictor variables together. This is known as multiple regression, and it's what people usually mean when they use the phrase "key driver analysis".

This combines the predictors together to give us an overall measure of impact for all predictors on the outcome variable, usually reported as R squared. In this instance the R squared is a very respectable 0.85, which can be thought of as meaning we can explain 85% of overall satisfaction.

Unfortunately, our problems with multicollinearity are not over. That overall 0.85 is useful, but multiple regression does not do a good job of understanding the contribution of each predictor when they're strongly correlated. As you can see in figure 7, instead of allocating the overlapping information fairly, it tends to award it all to one or two key drivers, which then seem to totally dominate, as "Being treated fairly" does here. If the question you want the answer to is "Which one thing should I invest in?", then this is a reasonable answer. But if it's "How important are all of these?", then it's extremely misleading.
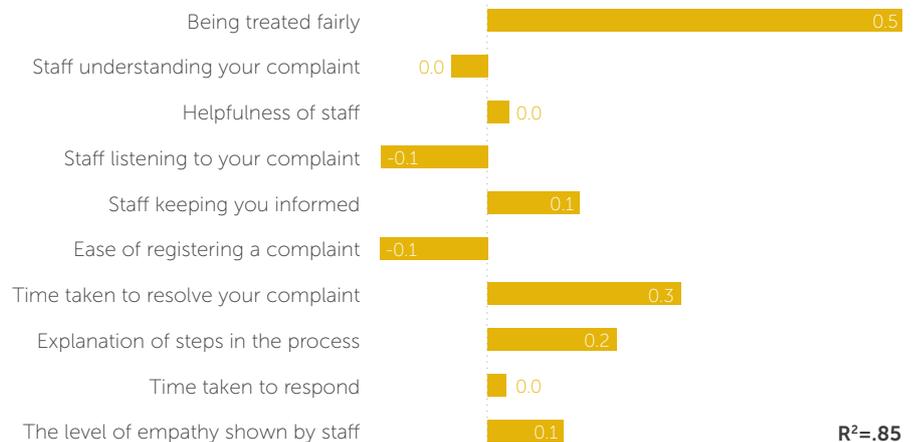
Being treated fairly — 0.5
Staff understanding your complaint — 0.0
Helpfulness of staff — 0.0
Staff listening to your complaint — -0.1
Staff keeping you informed — 0.1
Ease of registering a complaint — -0.1
Time taken to resolve your complaint — 0.3
Explanation of steps in the process — 0.2
Time taken to respond — 0.0
The level of empathy shown by staff — 0.1

$R^2 = .85$

Fig.7: Exaggerated key drivers with regression

Relative importance analysis

Is there a way of getting to a fairer breakdown of the contribution? Yes there is. The failings of multiple regression are well known, and a number of techniques have been developed to address it, collectively known as relative importance analysis.

Figure 8 shows the output of the gold standard technique, known as lmg, Shapley Value Regression, or True Driver Analysis amongst other terms.
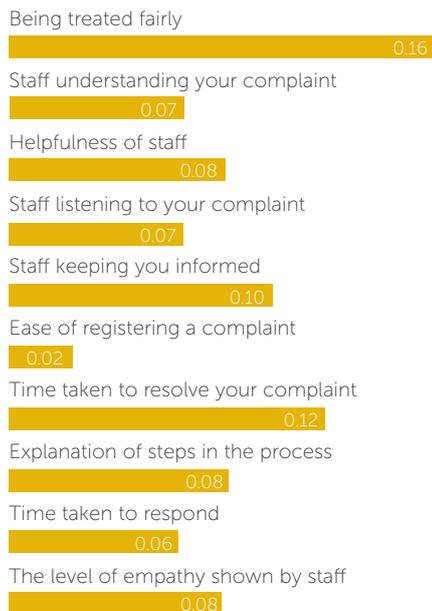
Being treated fairly
0.16

Staff understanding your complaint
0.07

Helpfulness of staff
0.08

Staff listening to your complaint
0.07

Staff keeping you informed
0.10

Ease of registering a complaint
0.02

Time taken to resolve your complaint
0.12

Explanation of steps in the process
0.08

Time taken to respond
0.06

The level of empathy shown by staff
0.08

Fig.8: Balanced drivers with relative importance analysis

As you can see, it gives us a picture of the relative contribution of each predictor which looks far more intuitively sound. Some drivers are more important than others, but the differences are not as extreme as multiple regression suggested.

Relative importance works by parcelling out the overall R squared, so it can also be interpreted in a very straightforward way as the percentage of the outcome variable that each predictor accounts for, meaning that in this instance "Being treated fairly" accounts for 16% of overall satisfaction, "Time taken to resolve" accounts for 12%, and so on. If we add up all those percentages it totals 85%… which is the overall ability of our list of items to account for overall satisfaction.

This is great, because it means that this is not only the gold standard in terms of methodological rigour, it's actually very easy to explain to people as well.

Inevitably, there's a catch. What relative importance techniques do not allow you to do is answer "what if" questions in

a straightforward way. They're the gold standard for explaining the contribution of predictors, but they can't make predictions.

Dealing with multicollinearity

Let's go back and start again. We've mentioned it a few times, and it's true that our big problem is multicollinearity. Is there a way to tackle that issue head on?

The first place to look is not analysis, but your questionnaire. The chances are that it's too long, and if it looks like you're measuring the same thing in two, or three, or more different ways, then it suggests that something is fundamentally wrong with your research. A questionnaire that discriminates badly between drivers is one that is painful for customers, and is difficult for you to action. It's a really unhealthy sign. So a questionnaire with a smaller list of more distinct items will work much better.

Within reason, though, multicollinearity is an inevitable part of customer research,

driven by an innate psychological bias we all share called the "halo effect".

One approach to this is to study the patterns in the way items on the questionnaire correlate with each other, and to group together those that seem very similar. A family of statistical methods known as "dimension reduction techniques", including factor analysis and principal components analysis amongst others, help you to do this in a data-driven way.

In this instance, although everything is correlated with everything else, the data suggest that there may be three main groupings: "Ease of registering" on its own, and then a bundle of process related items and a bundle of staff related items.

If we combine the information in those bundles, as in figure 9, we can soak up a lot of the collinearity that we've been struggling with, and create a more reliable model which reflects how customers think about their relationship with you.

Fig.9: A two-stage driver model

Those bundles can then be used in a two-stage model, combining dimension reduction techniques with multiple regression, to give us the benefits of multiple regression without the weaknesses. A number of different methods exist to construct and test models which look like this, of which the gold standard for customer experience research is Partial Least Squares Path Modelling, or PLS-PM.

By now you're probably wondering what the catch is. Apart from the fact that it's difficult, the main problem is that it requires

large sample sizes (of the order of 10-20 cases per variable). If you have the data, though, it's definitely worth considering.

In the real world

In the real world, dealing with real messy data, things are never quite as clear-cut as they seem in the textbooks. There are lots of potential traps, but I want to discuss three specific things which can make driver analysis difficult in practice.

Missing Data

First, and perhaps most serious, is missing data. All the people who answer "not applicable" instead of giving you a score, and even the people who have been asked one set of questions but not another because of routing on your questionnaire.

Some techniques, such as correlation, are reasonably robust in the face of missing data. Correlation uses pairwise deletion, so it includes all the data it can based on matching pairs of scores.

All the more complex techniques find it much harder to deal with missing data, because they default to something called listwise deletion, which means that anyone with any missing data is completely excluded from the analysis. That's a real problem, because it means even a scattering of missing values can easily leave us with hardly any valid cases to analyse.

There are ways to deal with this, from replacing missing values with an average (which biases any links towards zero), to replacing them with predicted values (which overestimates our certainty about how strong links are). The gold standard is something called multiple imputation, which is not used anywhere near as often as it should be.

In practice, the important thing is to examine your missing data so that you can make informed decisions about it, and often the best solution is to build separate models for separate groups of customers (such as those who have answered the questions about online versus those who have answered the ones about phone).

Sample size

Missing data compounds a second problem, which is that advanced techniques require large sample sizes.

Correlation coefficients, and all the measures of impact from more sophisticated techniques, have margins of error, just like any other figure you deal with in survey analysis. These margins of error are often larger than we'd like them to be, so sample sizes are something you should take seriously.

So what do I mean by a "large" sample? You shouldn't even think of looking at a correlation unless you have a minimum of at least 50 cases, and you should be aware that the confidence interval will be quite wide. To give a specific example, with a sample size of 50 and a correlation of 0.55,
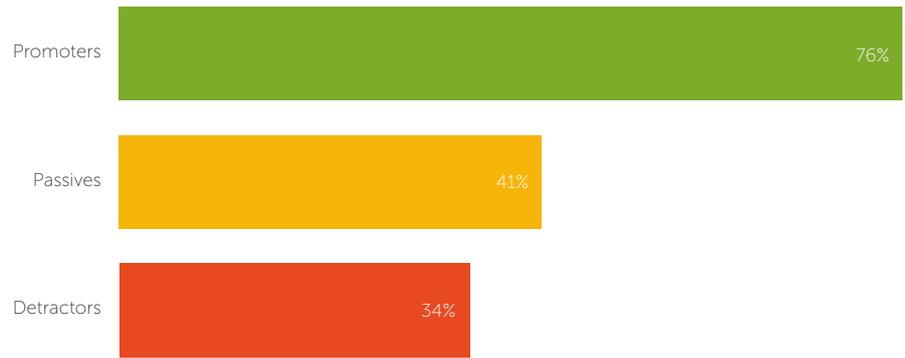

Fig.10: A non-linear link

your 95% confidence interval range is from 0.32 to 0.72. The correlation is statistically significant, in other words we're sure it exists, but we're pretty hazy about exactly how strong it really is.

With techniques such as multiple regression most researchers recommend a sample size of at least 10 valid cases per variable included in the model. That's not usually a problem at an overall level, but it can be an issue when you want to break the results down by subgroups such as customer type or demographics.

The assumption of linearity

One final thing to watch out for is that, if you think back to the scatter plot we started out with and the fit line we added, all of the techniques we've discussed assume that the relationships we're looking at are linear, and that may well not be the case.

It usually holds relatively well for items on a survey, but if you want to look at the link between survey data and actual customer behaviour, such as the repurchase rates shown in figure 10, then you need to make sure you look out for non-linear relationships such as this one. The links between your behaviour, customer attitudes, and then on to customer behaviours are often non-linear.

If you don't consider that possibility, there's a severe danger that you will underestimate how strong the links really are. I suspect this is one of the reasons that some organisations have a hard time proving the financial benefit of customer satisfaction.

Best practice

So we've looked at a whole range of techniques with some of their strengths and weaknesses. What does a best practice

approach to driver analysis look like?

- *Start with correlation. It's simple, it isn't as affected by missing data, and it makes fewer assumptions than more complex techniques. It will give you a good steer on what the drivers are, and frankly is probably the best measure for most organisations most of the time.*
- *Combine it with stated importance for a full understanding of how your customers see each aspect of the experience right now, and you'll have a tool that allows you to monitor the ongoing evolution of customer needs.*
- *Especially if your questionnaire is long, it makes sense to use dimension reduction techniques to look for patterns and groups. This can help you to understand how customers think.*
- *If you want to know what's making a difference to your outcome variable right now, then relative importance techniques are the best way to break that down. They're far superior to multiple regression for this, and you really should make the switch if you need to know what matters.*
- *If your questionnaire lends itself to breaking into bundles of related questions, or if you want to investigate more sophisticated causal modelling techniques, then partial least squares path modelling is the best technique for customer data. Be prepared for a lot of hard work, and make sure you have good sample sizes available, but if you can make it work you'll be doing some of the most robust analysis it's possible to do with customer data.*
- *Finally, whichever route you go down, remember that some of the most interesting links you investigate may be non-linear.*

Stephen Hampshire
Client Manager
TLF Research
stephenhampshire@leadershipfactor.com