

QUERYCLICK

The Complete Guide to Marketing Attribution

A MARKETER'S GUIDE TO REBUILDING MARKETING ATTRIBUTION

www.queryclick.com

Contents

04 Marketing attribution: the basics

09 Marketing attribution models explained

15 The problem with current attribution solutions

18 Why your data is the foundation for your attribution success

26 Marketing attribution: a guide to your choices

33 A new approach to attribution: visit-level attribution

38 The powerful data views you need to accurately drive marketing ROI

40 Closing thoughts

Overview



The purpose of attribution is deceptively simple: to most fairly share the value of a goal conversion across all touchpoints that may have influenced that conversion. However, in the real world the complexity of touchpoints and media opportunities in marketing create an attribution challenge, even in a perfect world of data availability.

The customer journey is now very often a highly complex one. And being able to attribute the impact of specific marketing touchpoints is crucial, as pressure from internal stakeholders - including finance and the boardroom - to link marketing to revenue, and prove ROI intensifies. Unravelling the impact of specific touchpoints on conversion is priority number 1 for marketers.

In fact, its relative importance was summed up when Google's marketing evangelist, Avinash Kaushik, described solving attribution to identify incremental value across all media channels as the biggest problem facing analytics in 2020. So, attribution matters.

In this guide, we are going to take a close look at **some of the key aspects of attribution** including:

- What marketing attribution actually is
- Why it matters more now than ever
- An introduction to the main attribution models including some of their limitations
- Why data is key to all of this – and some of the challenges and opportunities around collecting data across websites, offline media, social platforms and CRM/CDP/ERP
- How techniques like Machine Learning and Deterministic and Probabilistic matching can help overcome limitations in current attribution approaches
- How all of this provides effective attribution at a Channel, Campaign and Impression level to better focus and optimise your marketing efforts so that they are more closely aligned to the customer journey itself



Marketing attribution: the basics

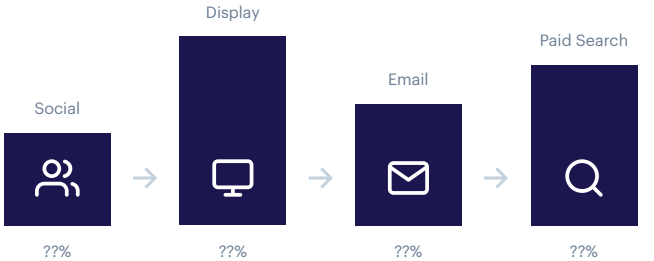
Marketing attribution: the basics



So, what is marketing attribution?

Put in its simplest context, marketing attribution is the process of determining which marketing touchpoints and activities are contributing to conversions – which could come in a variety of shapes and sizes from sign-ups, to downloads, to purchases of a product or service – or some other meaningful conversion types.

In a complex marketing environment where marketers are using an ever-increasing mix of channels simultaneously it is about unpicking the impact that channels, campaigns and even individual creative executions are having across media like Display, Paid Social, Search (Organic & Paid), email and offline media.



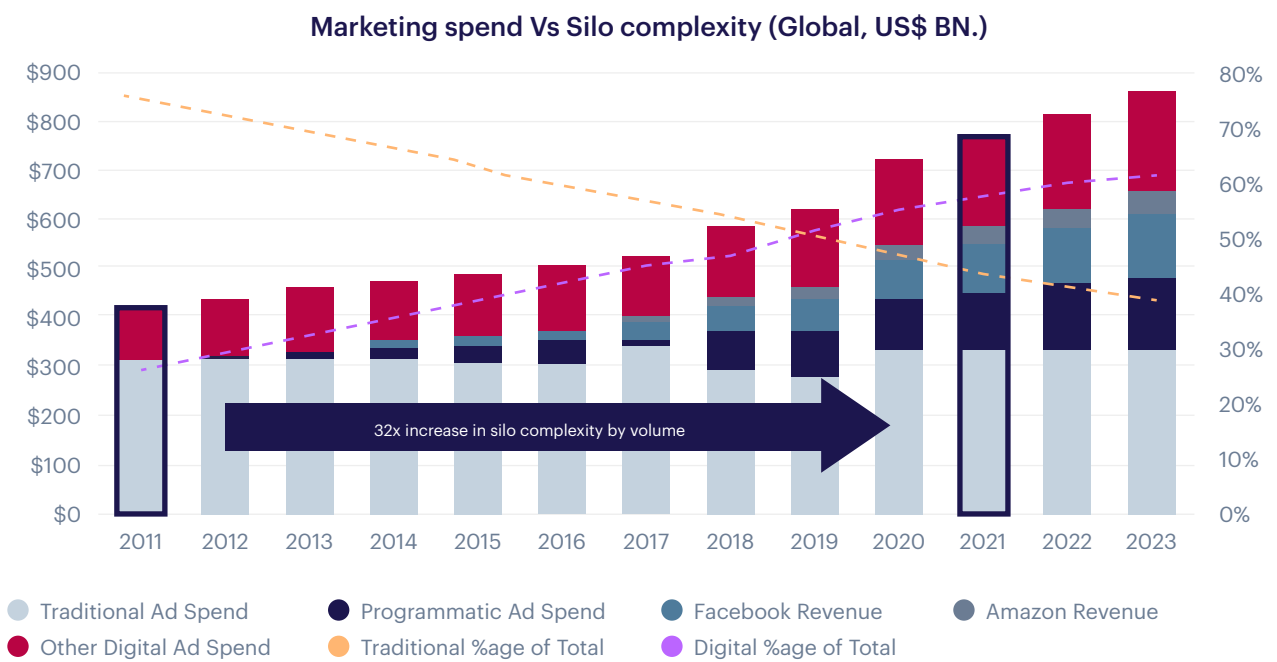
An attribution model, or combination of models, is a defined set of rules that helps you determine how conversion credit is given to these different marketing touchpoints across the customer journey. They come in all shapes and sizes – but are split broadly into two types: single-click and multi-click. But more detail on those later.

What's important to state here is that, if you don't have effective marketing attribution in place, then it's a fairly safe bet you don't know what is – and just as importantly, what isn't – working across your marketing mix. So accurately assessing and improving "true" marketing performance isn't really possible for you.

Why marketing attribution matters. More than ever before.

So, attribution matters - and it always has.

At the most fundamental level the sheer scale of current investment in marketing demands effective attribution. In the UK alone marketing spend will top £21.3bn in 2020



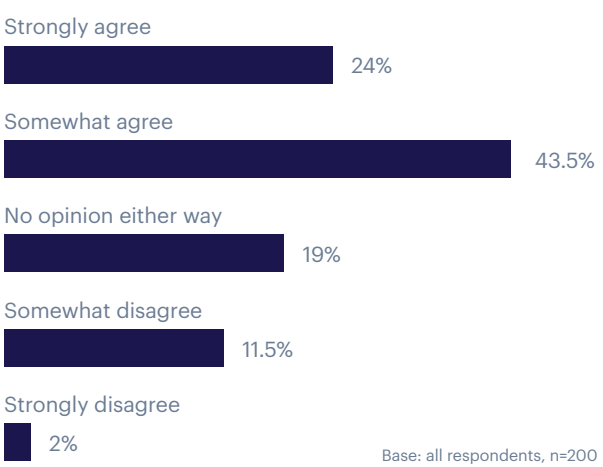
Source: eMarketer Global Digital Marketing Report & Statistics.

and digital is a growing part of the mix. By 2021 the volume of data in silos involved in delivering just digital marketing, will have **increased in complexity 32-fold compared to a decade previously.**

The figures are staggering and, as a result, many businesses are struggling to sustain growth while maintaining or reducing media spend budgets. Which – and to echo Avinash Kaushick - makes solving attribution the largest opportunity any marketer can engage with today.

But there are also a number of other key reasons why having an effective marketing attribution solution in place is more important than it has ever been, as it enables you to:

"Internal stakeholder pressure restricts my option to employ marketing activity that has a longer payback period than last-click measures."



Base: all respondents, n=200

Link marketing spend directly to growth and revenue

Which is non-negotiable now. The pressure for marketers to spend accurately, efficiently and to prove marketing effectiveness in driving growth is greater than ever. And is only going to intensify, so directly tying activity to revenue is no longer a "nice-to-have". It's essential.

Keep full control over the nature of marketing spend

Not only is there broad evidence that internal stakeholder pressure on spend - from outside of marketing - is increasing but there are also worrying signs that it is directly influencing the nature of spend. In fact, research carried out by QueryClick points to the fact that 67.5% of Marketing Directors report that internal stakeholder pressure actively restricts the option to employ marketing activity with longer term payback.

Prove your value to the business and build stakeholder trust

Effective attribution builds credibility and trust with the key stakeholders closest to the marketing function including the boardroom, finance – and, perhaps most importantly, sales. Securing the buy-in needed to support your marketing vision and fast-track future plans.

Secure and ring-fence budgets based on sound marketing analysis

Marketing budgets are under the microscope more now than any time previously. By accurately being able to show how marketing is contributing you have the capability to ringfence and even grow your budget.

Improve your marketing ROI

Ultimately, effective attribution helps you drive improved marketing ROI by finding the parts of your program that are working - and those that are not. This the allows you to redirect spend to where it is going to have the biggest revenue impact.

Reduce reliance on perceived short-term media ‘fixes’

For many businesses, media like Pay-Per Click is the main channel. In fact, everyone loves it because it provides immediate results. Switch it on and revenue flows when you spend. But it is no longer the panacea it once was. High spending competitors are not only eroding market share but driving up CPAs. So, ROI is being driven downwards. Effective attribution opens the door to wider media opportunities, with longer payback cycles – because you can pinpoint the data you need to make the business case for it more effectively.

Take control of marketing complexity

Finally, better attribution enables you to take control of what is ultimately a more challenging and complex media environment.

Where customer journeys are taken on and offline, and research can be done across a multitude of devices including smartphones, tablets, work and home PCs, and mobile, effective attribution enables you to bring together the data silos across this complex environment to define a single, data-driven view of the customer journey.

And confidently build your marketing approach around it.



Marketing attribution models explained

Marketing attribution models explained



In this section we are going to take a closer look at attribution models.

One of the key things to understand upfront about attribution models is that there isn't a definitive, one-size-fits-all answer for everyone. And, in practice, identifying which model is right for your brand or your business is going to be impacted by a range of different factors from the stage, scale and maturity of your marketing activity to how much time, budget and effort you are able to apply to it.

Below we take a whistle-stop tour of each in turn. Including where they might work well and where they also have limitations.

In broad terms they split into two broad types – single-channel attribution and multi-touch attribution. Let's look at these in turn.

Single-channel attribution

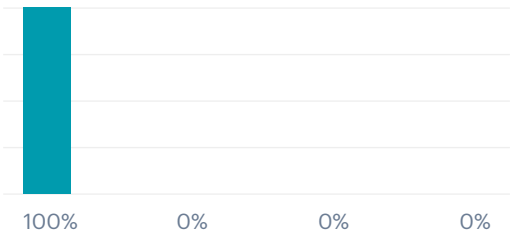
These models are relatively simple, and work on the basis of allocating 100% of the credit for a revenue conversion or other customer behaviour to a single touchpoint with the customer, either at the beginning or the end of the customer journey. All other touchpoints in between are ignored for attribution purposes.

↑ **Works well** Where you have a simple funnel environment with very few marketing activities. And businesses who are scaling from a low base may well employ them early in their growth.

↓ **Possible limitations** However, they oversimplify and ignore the impact of multiple touchpoints on complex, longer customer journeys.

There are two types of single channel models – First and Last-touch attribution.

First-touch attribution



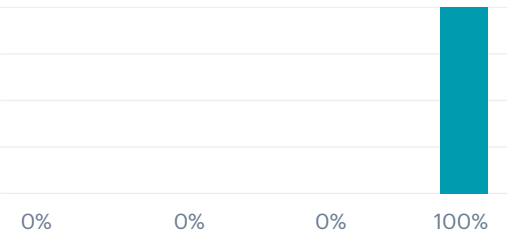
Built around the premise that the first touchpoint with the customer is the crucial determinant in a customer's decision to do business with you.

So, 100% of the credit for revenue is directly attributed to the first interaction in the customer journey. For example, if a customer comes in on an eBook download it gets all the credit – regardless of all the other touchpoints in between that and the sale.

↑ **Works well** Where you have a simple, early stage marketing model that is heavily focused on growth and quickly scaling your business. And where there is a strong focus on growing top of funnel.

↓ **Possible limitations** However, it is going to be much less appropriate where you have a lot going on in your marketing funnel and a more complex and extended buyer journey (of 90 days plus).

Last-touch attribution



Similar to First-touch attribution, Last-touch is predicated on allocating 100% of the credit for conversion to a single touchpoint.

However, in this case it allocates credit on the basis of the last touchpoint prior to purchase or revenue. And similarly, it gives zero credit to other interactions along the way.

↑ **Works well** Where you have a heavy focus on optimising conversion around the bottom end of the funnel.

↓ **Possible limitations** But again, it's going to be much less appropriate for analysing longer and more complex journeys accurately.

So, taken together, Single Channel attribution models do have their benefits – and you may well start your journey using them in early growth stage scenarios in smaller businesses

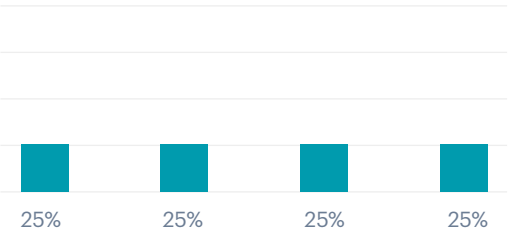
– but it is also highly likely that even an initial spurt of growth will see you outgrowing their usefulness. As your marketing evolves and matures, your need for more sophisticated attribution is likely to follow suit.

Multi-touch attribution models

These models take a broader view of attribution and attempt to place a relative value on the impact of all touchpoints across the customer journey, although these are weighted differently depending on the model you select.

If you are operating with a more complex model, longer cycles and a multitude of channels it's likely that this is where you find yourself gravitating to. Let's take a quick look at each of the commonly used multi-touch models in turn.

Linear

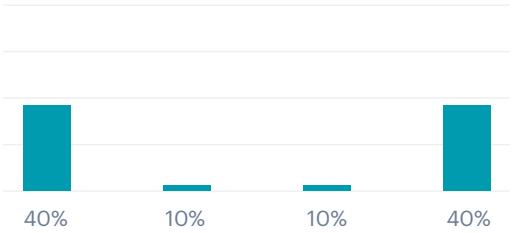


Very often you don't want to give single channels or campaigns all the credit for conversion and a Linear model moves the principle on by applying equal weighting and credit to all contributing touchpoints.

↑ **Works well** Where you are looking to measure a campaign holistically and get a feel for the customer journey, including seeing which channels contribute across it

↓ **Possible limitations** Where you already understand the customer journey and have a more detailed need to dig into the data on individual channels to understand which ones are critical so you can focus your effort on them

U-shaped

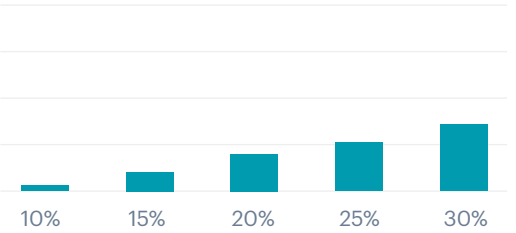


Takes view that some touchpoints have more influence on the way to conversion than others and scores them separately. Here first touch and lead conversion matter most and are both heavily weighted with 40% of the credit being applied to each. While the remaining touchpoints are given an equal share of the remaining 20%.

↑ **Works well** Where you are looking to identify what drives channel acquisition or have heavy focus on nurture

↓ **Possible limitations** If you are using Google Analytics but have a buying cycle that extends over 90 days you can potentially lose the data view you need

Time decay

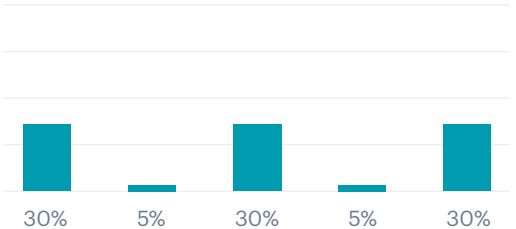


Also weights each touchpoint differently and is driven by the assumption that touchpoints closer to sale should get more credit. The assumption being they have a heavier impact on conversion as the buyer moves down the funnel.

↑ **Works well** With a short sales cycle less than 90 days and if you are looking to identify what drives conversion in the lower funnel

↓ **Possible limitations** Not so appropriate for longer B2B customer journeys where it ignores the fact that customers can make important decisions earlier in the process

W-shaped



An extension of the U-shaped model with the addition of weighting for the opportunity stage. Here first touch, lead conversion and opportunity creation share 30% credit each, with 10% being shared out between the other touchpoints.

↑ Works well Where you are looking to identify touchpoints at the 3 critical points that help you build audience, create leads and then drive to conversion

↓ Possible limitations If you are using Google Analytics but have a buying cycle that extends over 90 days you can potentially lose the data view you need

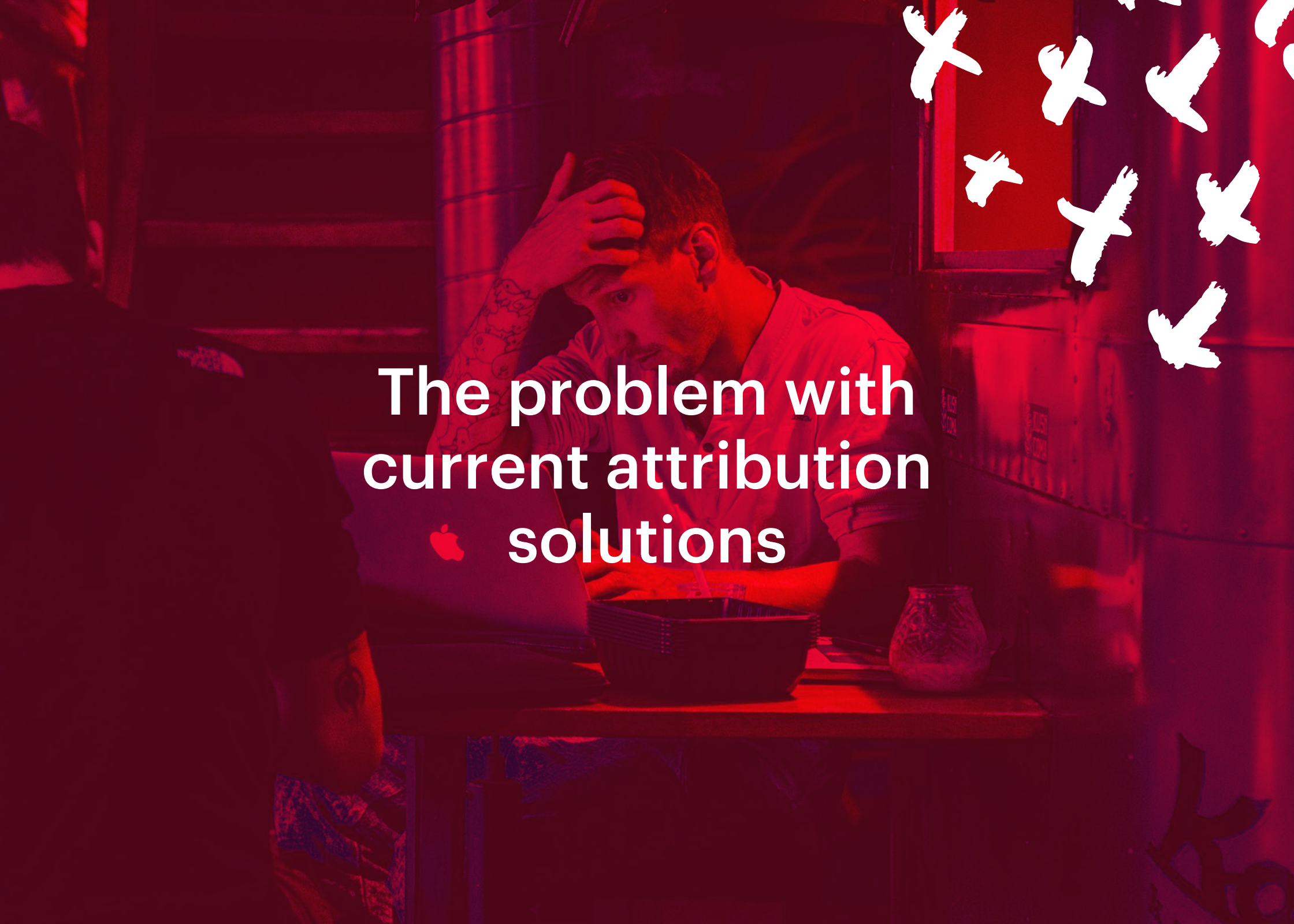
Custom model

This is where you take the view that you need a very personalised view for your marketing needs. And the discussion here opens up into a wider consideration of more complex attribution models that include Statistical or Data Driven Attribution, Media Mix/ Econometrics Attribution and Probabilistic modelling. Which we will return to in more detail later on.

↑ Works well Where you have a very direct need to tie things back to your end goals and need the flexibility to create tailored weightings to create a highly refined and accurate view of things

↓ Possible limitations Requires an adequate allocation of time, thought and resource to make it effective. So not to be undertaken quickly or lightly

For now, it's probably just important to know that, in practice, this is going to be a blend of elements from the models above. Customised in a way that works for your own marketing scenario.


The problem with
current attribution
solutions

The problem with current attribution solutions



The section above provides a good overview of marketing attribution models, each with their benefits and limitations. However, the reality is that existing flawed attribution models are part of the problem marketers face.

For a couple of key reasons:

Historical and oversimplified views: At the most fundamental level almost all current attribution models are historical in nature and all the evidence is that they are woefully inadequate in terms of doing what they should. For example, Last-click is the default attribution model for Google Analytics, and is a complete oversimplification where 100% of the value of the conversion is given to the last marketing engagement. In essence, this ignores all the previous points in the buyer journey: almost the same as applying no attribution at all.

Lookback windows are too short: Even tools like Google 360 have limitations due to their short lookback window and only partial attribution in relation to what is very often a long and complex customer journey – where the customer is influenced by a broad spectrum of media.

“Attribution is broken and it’s screwing up your entire approach to digital marketing”

So, rebuilding attribution – in a way that reflects the complexity of the customer journey – has to be a key goal for all marketers. Unfortunately, "in the wild" the vast majority of marketers - in fact, according to research by QueryClick, over 90% of those responsible for improving marketing ROI in the UK - are unable to get to sophisticated attribution models such as these. Or even simple rules-based models such as First-click, Position Based or Time-Decay Weighted. In an effort to provide better outcomes than a basic Last-click view.

But flawed attribution models are only one part of the story. It's also highly likely that **80% of your core analytics data is wrong** too.

So, the quality of "data in" to your attribution model has a huge bearing on the quality of your analysis. Which is why we are going to take closer look at this in the next section.

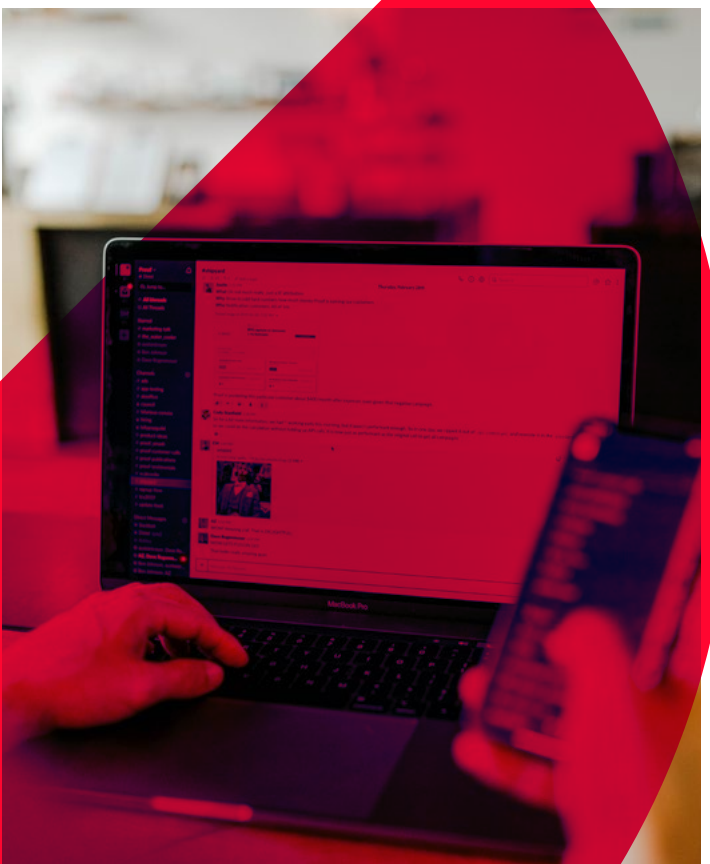


DOWNLOAD



Why your data is the foundation for your attribution success

Why your data is the foundation for your attribution success



Our extensive work with clients around data enrichment consistently tells us that poor models are only one part of the issue with attribution.

Just as important is the quality of the underlying data that you are feeding into your attribution model – and in purely layman's terms if you put rubbish in, you will simply get rubbish out. Let's take the practical example of Google's approach to collecting the vast majority of digital marketing data which is based mainly on a pixel plus cookie-based approach. The problem is that measurement in this model is rudimentary and based on tracking a device accessing a specific web property. Not around identifying an individual on the complex, multi-device journey they are on.

So, the seeds of success in effective marketing attribution are sown firmly in the way that you collect the data to feed into your model.

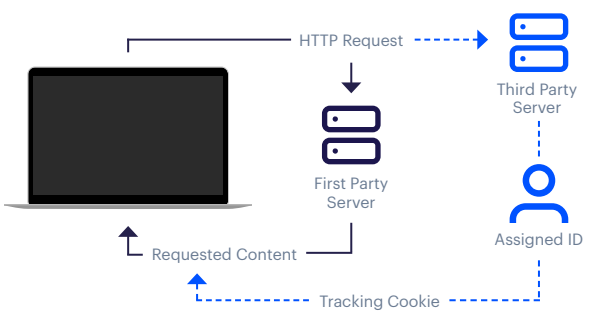
Challenges and opportunities in data collection

At this stage it is worth considering some of the challenges and opportunities around data collection across a number of channels and techniques including:

- Websites
- Offline media
- Mobile and social
- Fingerprinting & 1st Party Enrichment
- CRM, CDP & ERP

Limitations in the cookie/pixel approach

The vast majority of digital marketing data is what is called 'deterministic data' and is collected using a pixel working in tandem with a cookie. With the exception of log-file analytics, which uses data collected directly by the server hosting the web property.



A variety of data including:

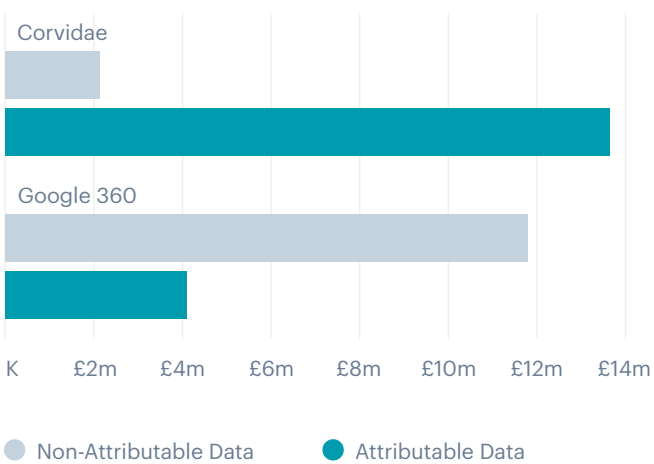
- measurement of dwell time
- previous site visited
- any ad campaign that may have been encountered by the visitor
- any repeat visit made by that same person

is all captured in this way.

And this is the point at which the world's most commonly used analytics packages start to break down and generate incorrect or incomplete data. The cause? What is being measured is a device accessing a web property. Not the person themselves. Who will use multiple devices in the course of even simple transactions.

Cookies are supposed to enable 'joining' of multiple sessions generated by devices interacting with a web property but, in reality, they do a poor job, as is shown in the following chart:

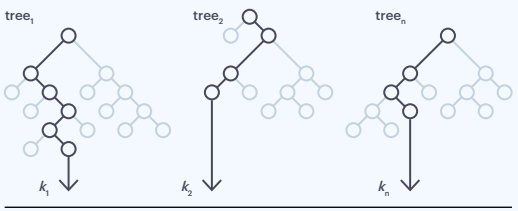
Daily attributable data (i.e. Repaired broken sessions & enhanced cross device)



TECH FOCUS

Random forest modelling

Random forest is a regression technique that uses multiple decision trees and merges their predictions together to get a more accurate and stable prediction rather than relying on individual decision trees.



In a Marketing Attribution context – and when combined with the Machine Learning capability of tools like QueryClick's Corvidae platform - it offers the possibility of taking deterministic data gathered from the clickstream and applies Machine Learning to rebuild it in conjunction with techniques like device characteristics to significantly improve attribution and identify individual customers journeys. As well as providing the predictive data insight needed to target incremental growth by using propensity to convert models to drive revenue.

This chart shows that using deterministic analytics data only (i.e. pixel and cookie data) to create a picture of the individual's behaviour behind their multiple devices. Even where advanced session joining algorithms are employed - as with Google 360 - generates **data that is around 80% incorrect**.

The alternative approach being used by Corvidae in this example, is a Machine Learning model approach, which takes the core deterministic data returned into the clickstream, and rebuilds it using random forest probabilistic modelling (see Tech focus) and regression analysis.

This approach is not applied by existing Web Analytics suppliers such as Google, IBM and Adobe, as compute intensity is high and initial model development has until very recently involved manual 'tuning'.

This analysis highlights why even core 1st party web analytics data for web properties fails to provide good data to base marketing decisions on. But more complex customer journeys involve a plethora of media and it's worth now considering some of the challenges – and opportunities – in data collection across mobile applications, CRM/ERP and offline, to build a clearer view of attribution across the customer journey.

Difficulties in integrating offline data

For most businesses, marketing activity is a blend of offline and online activity. Offline increasingly offers rich data as measurement moves beyond traditional panel-based approaches to incorporate Beacon technology, opt-in Wi-Fi Tracking and RFID data. These technologies, deployed compliantly, enable good data mapping to digital marketing data.

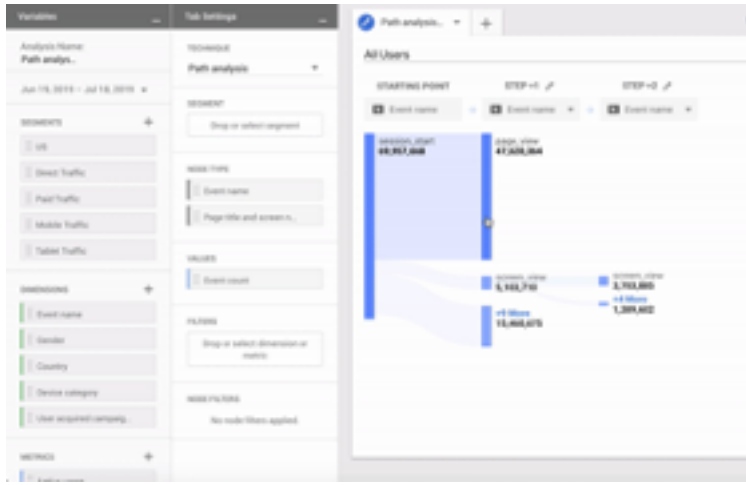
The largest marketing reach touchpoints outside of store are linear TV and Radio activity. Which typically offer either panel-led, econometrics data. Or increasingly purchased, direct from manufacturer data -

for example from Smart TVs - for inclusion in marketing mix analysis.

There are challenges with econometrics data including the inability to look beyond channel level impact as part of A/B channel or messaging strategy adjustment, a lack of immediacy in terms of delivering data outcomes and the fact that they operate in a silo to all other data collection strategies. Which creates an increased issue around cannibalisation. In addition to being expensive and less appropriate for smaller budget media spends.

However, if they are applied to enrich an already unified marketing strategy, econometrics can be a valuable tool to unlock customer behaviour, environmental influence, competitive intelligence and meaningful signals for targeting.

Challenges and opportunities in mobile application & social data collection



Data collection in-app or by a social platform is more tightly associated with an individual who is required to login. Which creates a much more robust measurement environment.

However, the challenge today with activating mobile app data is building good quality data to connect it to. For example - if purely deterministic data collection is being used in the wider web, then the individual's app

TECH FOCUS

“Walled Gardens” on Social Media platforms

The broader definition of a Walled Garden is a closed Ecosystem in which all operations are controlled by the ecosystem operator.

In media terms, this is very much the approach that highly successful social platforms like Facebook, Twitter, YouTube etc. are taking in marketing terms. By keeping their information and data analytics firmly closed for privacy but - just as importantly, commercial value reasons. In practice, this means reporting only aggregated data with no identifiers to aid the process of identifying individual customer journeys.

It is only by using the power of Machine Learning in Attribution platforms like QueryClick’s Corvidae that you can begin to use sophisticated datasets from social – using data on page views, location details, device details, on-site behaviour etc. - to derive specific user journeys.

based activity which it is being “joined onto” will remain 80% incorrect for the reasons we highlighted in the section above.

In addition, practical issues around how social platforms utilise “share” functionality across web properties. Combined with the fact that significant volumes of retargeting activity occurs against social platform pixels - essentially short-cutting the need to understand the customer journey and simply targeting them based on their demographic or sociographic profile and proximity to a conversion event - means individuals are already some way down the conversion path.

Social platforms also typically take a ‘walled garden’ approach to data availability for analysis outside of their own analytics (see Tech Focus). Which is geared towards enabling native advertising to their users.

However, these platforms offer a potentially rich data view of customers and it is now possible, using sophisticated Machine Learning techniques to ‘unbundle’ data from APIs - such as the Facebook one - and ingest their data into a wider analytics view.

Fingerprinting & 1st Party Enrichment of Data

One associated data collection technique - which is highly relevant to social profiling - is fingerprinting. Essentially a way to combine certain attributes of a device to uniquely identify it.

There are two common forms of fingerprinting: device and canvas.

Device-led fingerprinting

This approach uses “hit-level” data from a person’s device to determine its uniqueness, and then match it to subsequent visits. The level of data available is very high and this is a very accurate way of joining multiple sessions together when included in the type of probabilistic models we outlined earlier.

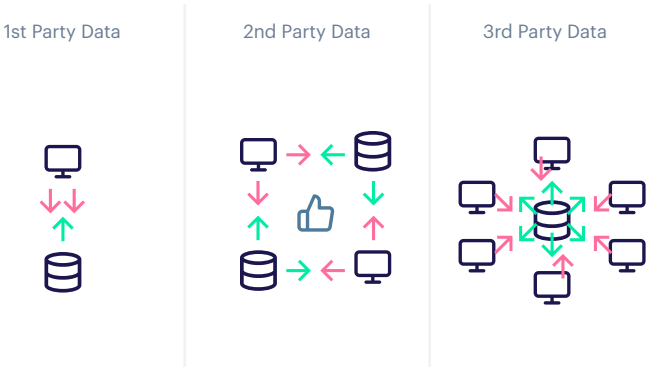
Canvas fingerprinting

A single device offers a wide range of data when preparing to render a web property and when combined with canvas fingerprinting it can become entirely unique. Although it does have potential data compliance issues to address.

Canvas fingerprinting data can also be collected via the delivery of ad inventory in the wider digital ecosystem. However the bulk of canvas render requests today are for ad visibility verification purposes. Facebook also has tracker pixels performing the same fingerprinting across 30% of the top 1,000 websites by visits globally. Giving their ad targeting data deep richness.

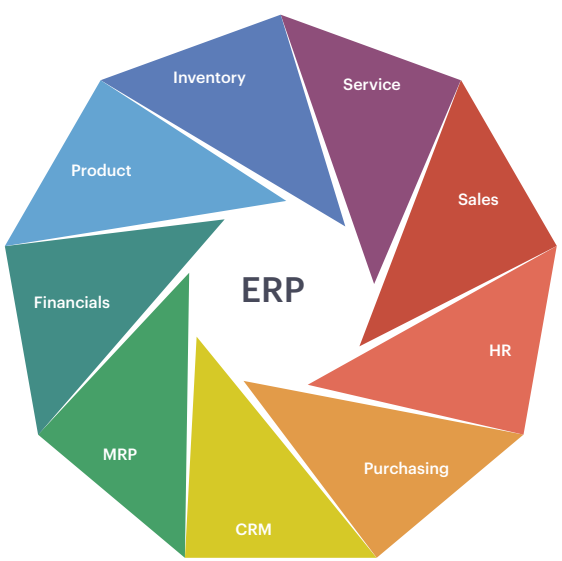
However, Pixels that are set from outside a web property's domain are already blocked by default for Safari users. And so, by default also on the vast majority of iOS devices. In addition, the world's most popular browser, Chrome, which is owned by Google, will be moving to the same default block position through the course of 2021.

All of this means that 3rd party data collection - essentially data collection from domains that are not the web property domain - will become ineffective, unless they switch to 1st party data collection processes.



Leveraging the wealth of data in your CRM, CDP & ERP

Similar to mobile application data collection, Customer Relationship Managers (CRM), Customer Data Platforms (CDP), and Enterprise Resource Planning (ERP) in some senses make data collection and achieving data accuracy simpler. As customer identification or sign in is usually a pre-requisite.



However, these systems' attempts at 'joining' or crossing over customer IDs into wider web analytics, is also affected by the 80% inaccuracy issue highlighted in an earlier section.

However, there is a wealth of potential data available to build into your analysis including:

CRM & CDP data

Across email, SMS, targeted ads, phone calls or direct mail

ERP data

Across clubcard schemes, Point of Service (POS) transaction data, inventory and stock availability data, returns and fulfilment cost data

That can be complemented by 3rd party data that can be associated with customer profile data to help drive:

- high quality customer segmentation and cohort analysis
- marketing automation to improve conversion and engagement and to increase the volume or frequency of purchases

As well as enabling a deeper understanding of marketing ROI. So, exploring what is possible here is well worthwhile.

Finally, be careful with data sampling and vendor selection

Before signing off on this section on the challenges and opportunities of data collection, it is worth making a small note on the impact of sampling and vendor selection. There can be issues with "sampling in" analytics data which differ from vendor to vendor. So, vendor selection when considering all of your data collection needs is a key part of getting attribution working for you.



Marketing attribution: a guide to your choices

Marketing attribution: a guide to your choices



Once you have identified and gathered your data, how you apply attribution to it is equally important. Marketers today have 4 options to get attributed data on their marketing activities.

- Single/multi-touch or rules-based attribution
- Statistical or data-driven attribution
- Media mix/econometrics attribution
- Probabilistic modelling

Let's take a quick look at each in turn.

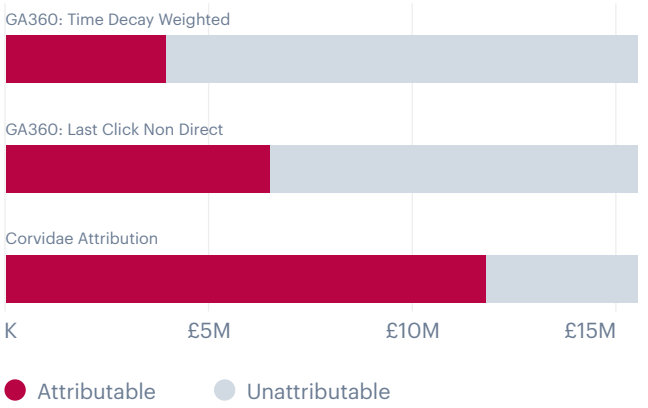
Single/multi-touch or rules-based attribution

We looked at these models earlier on and they include a range of options including First-touch, Last-touch, Linear, Time Decay and W/U shaped models.

These approaches are segmenting data from core clickstream data into scoring for each channel touchpoint and therefore typically offer only channel attribution outcomes.

Without rebuilding the clickstream, however, they suffer from increasing levels of inaccuracy as we can see in this side by side comparison using the same dataset as in the chart below – where attributable data identified by Google Analytics 360 was a fraction of that uncovered by QueryClick's Corvidae platform.

Daily attributable data (i.e. Repaired broken sessions & enhanced cross device)

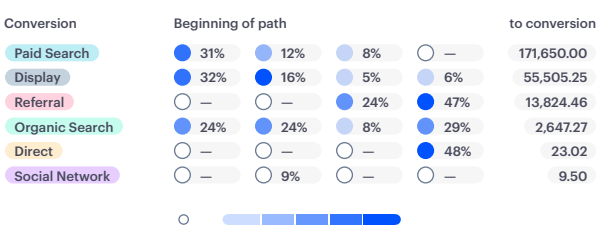


When used in combination with segmentation, channel attribution can still be powerful and enables budget reallocation across channels for paid marketing activity that will affect overall marketing impact. However, it offers no insight for individuals,

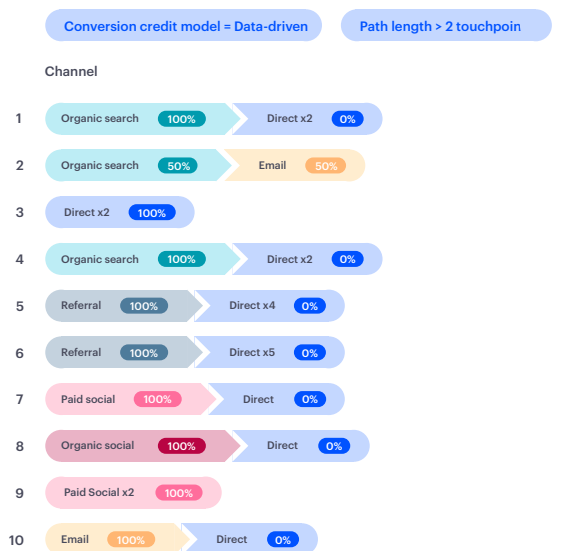
preventing the use of this attribution type for automated customer acquisition or insight on individual impressions. Or the impact of specific pieces of marketing content on that individual and their subsequent behaviour.

Statistical or data-driven attribution

Increasingly marketers have access to 'data driven' models which essentially apply channel attribution rules that are determined by the data available to them. Instead of rules-based assumptions about channel mix, data-driven models determine the model's weighting based on existing data.



Even the free version of Google Analytics now offers this functionality, and it is increasingly available in digital ad tech like Google Ads and is based on using vendor specific tags – which have the very real limitation of not being aware of data collected elsewhere of course.



Channel attribution using data-driven approaches allow for path grouping like this in reports.

Data Driven channel attribution in GA and Adobe suffers from **the same data quality challenge** as rules-based options described previously – the modelling it relies on can only work with the supplied data. Which as

we have seen is inaccurately representing the individuals behind the clickstream. And so, the data-driven models are showing only part of the real picture.

There is still value and insight to be gained from these models, however. As they can give an understanding of the contribution of channels that are typically poorly served by Last-click - Paid Social in particular - in your marketing mix. As an overall measure of volume of interactions.

This means that relying on data-driven models in web analytics today is not advised for assessing revenue outcomes. Without additional processing from the likes of QueryClick's Corvidae platform.

Other types of statistical attribution are available in digital marketing, for example assisted conversions reports like the one shown here from Adobe. That can be usefully applied to get a better understanding of the volume impact of impressions on conversion paths.

But, again, should not be used to assess revenue outcomes.

	Orders Last Touch | Visitor			Assisted Orders		
Marketing Channel	Feb 1 – Feb 28	2,009		Feb 1 – Feb 28	2,009	
1. Email		624	31.1%		480	23.9%
2. Referring Domains		542	27.0%		355	17.7%
3. Social Networks		217	10.8%		141	7.0%
4. Paid Search		190	9.5%		136	6.8%
5. SEO		179	8.9%		103	5.1%
6. Display		174	8.7%		134	6.7%
7. Direct		42	2.1%		42	2.1%
8. None		41	2.0%		41	2.0%

Media Mix Modelling (MMM) / Econometrics

This approach applies statistical techniques to attempt to find incrementality in large data sets using mathematical techniques. Often relying on integrating panel data for offline activity.

In all cases, the object of MMM or Econometrics is to isolate a test case that is to be measured when it is applied in two or more variations. And the outcome is often expressed as a measure of 'incrementality' to aspects of the media mix:

For example, a business may ask what the cost per acquisition is for an offline catalogue. In creating a test to answer that question, different mixes of media channels will be used for cohorts of customers and the overall acquisition outcomes measured. This will involve mix variations including:

- the use of catalogue only
- catalogue plus some marketing channel
- and the same channels with no catalogue

By comparing the results of these variations, an overall value can be determined for the incremental acquisitions made thanks to the catalogue itself. And that result will be enriched with data about:

- the cost of delivery
- operational cost
- marketing costs and so on.

This gives a good and rich level of detail to the insight gained from the media mix modelling. Ultimately offering an actionable outcome that is statistically true for those tested cohorts for the period of time the test was run.

	Revenue	Profit Factor	Profit	Expense	Net
Catalog + Email	$12.00	30%	$3.60	$1.01	$2.59
Catalogs Only	$10.00	30%	$3.00	$1.00	$2.00
Email Only	$8.50	30%	$2.55	$0.01	$2.54
No Catalogs No Email	$5.00	30%	$1.50	$0.00	$1.50

A simple experiment example of test variations and measured metrics from an A/B or Media Mix test. This example shows a 2% incremental benefit from running offline catalogues with an email campaign. Source: kaushik.net

In this way we can understand that media mix modelling is not just a case of running A/B or multivariate testing - it is a science with a great volume of subtleties that can be used **to answer a particular question for a particular period of time**.

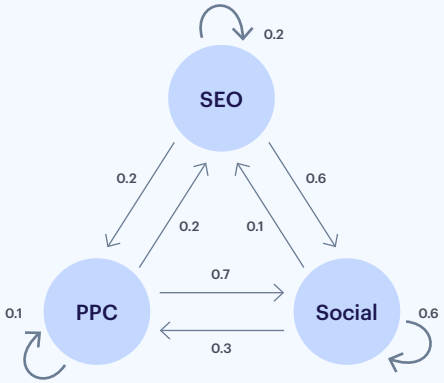
Due to the complexity and volume of data required to create balanced tests and models that are statistically sound, media mix and econometrics studies done correctly are an expensive, ongoing proposition. That should only ever be run in parallel to attribution activity that is intended to allow tactical optimisation day to day.

TECH FOCUS

Markov Chain modelling

Markov Chain is an example of a sophisticated, probabilistic modelling approach that looks at individual events in the customer journey and identify patterns of user behaviour to build journey profiles.

By evaluating the probability to move from one state to another in the customer journey.



For example, if a website visitor is identified as coming in on Organic Search, the Markov Chain would then understand and look at the other channels the visitor could have come in on and the likelihood they would come back from Organic. It would also look to see if they would come from Paid or if they would never come back or if they would transact.

This type of technique can be used to predict behaviours that lead to conversion, and the data can be fed back into the marketing model to optimise marketing performance.

In short: A/B Testing and Econometrics studies should be used infrequently to inform strategy. While attribution should be used to inform day to day decision-making in service of optimisation. Which leads us to our last option: probabilistic modelling.

Probabilistic modelling

Recent advances in Machine Learning and cheap deployment in SAAS architectures have allowed probabilistic data processing to be applied to data challenges in marketing day-to-day.

$$v(A) = \sum_{S \subseteq A} C(S),\ C(S) \geq 0$$

For example, Google Analytics uses Shapley Modelling for their data-driven channel models. This is based on Game Theory, and approaches marketing as a game, where each channel is a player in the game, and the set of all players/channels can be seen as though working together in order to drive the conversions.

A very broad brush explanation of a typical Shapley approach is to find an average incremental value for a particular channel touchpoint by looking at all conversion paths that exist, their conversion path rates and their mix of channels. Comparing the different mixes and conversion values brings you an average incremental value for each channel for each conversion path. Which is then aggregated into the reporting you see in the analytics interface.

Another common model is a Markov Chain model. Markov Chains are named after the mathematician Andrey Markov, and in essence, enable modelling a sequence of events in which the probability of each event depends only on the state attained in the previous event.

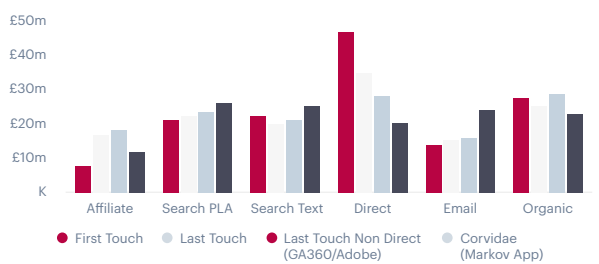
And practical application of the Markov Model can be powerful. Particularly when you are interested in channel level attribution.

Let's look a practical example. In the case below, a Markov Chain based approach

was used in conjunction with the Corvidae platform to assess and more accurately attribute revenue contribution by channel for a well-known Online Fitness retailer.

As you can see below, compared to Last Click Non Direct (the rules based model used by GA 360 and Adobe by default), **channel attribution has changed by £39.9m**.

Top 6 Channel attribution comparison by channel



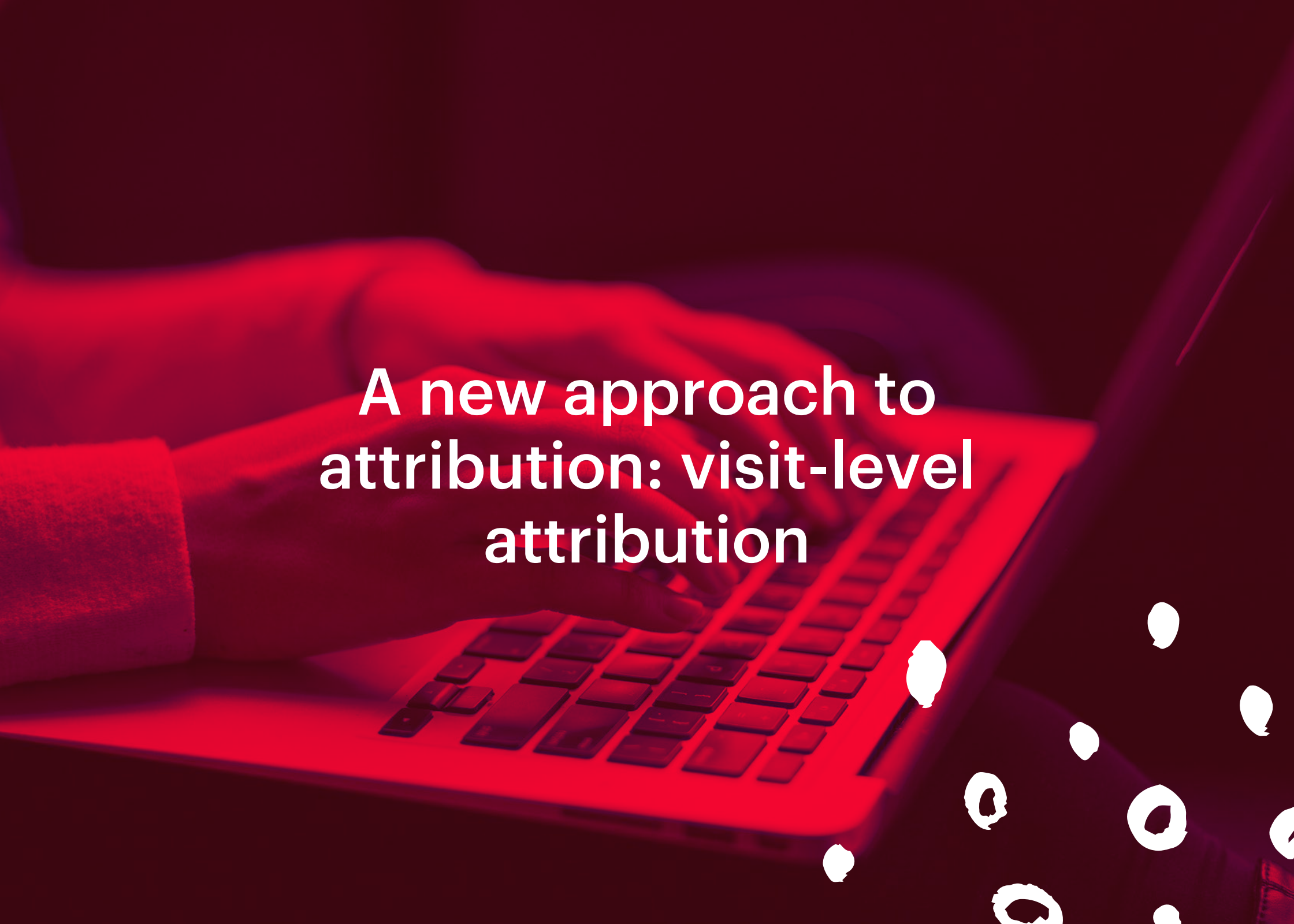
Given total turnover is only £136.7m across the measured channels, this **radically alters the ROI implications of media spend, particularly in Paid Search and PLA (Paid Search Shopping Ads)**.

Fundamentally both approaches, Shapley and Markov, are predictive models and have much in common, however they both have drawbacks.

For example, using Markov chains for marketing conversion paths will find lots of 100% probability value chains due to the very long tail of customer conversions that exist for all businesses - and are limited in their ability to look at more granular value answers than simply channel value.

Most importantly however, both are limited in that they must be re-modelled to provide more granular outcomes than just channel attribution. Which is computationally very expensive - and which also causes these models to run into data volume limits which prevent their practical functioning. Small Paid campaigns would not model with either approach for example. And understanding the engagement an individual visit has - and that individual's probability to convert at each individual stage of their own, unique conversion path - is not being modelled.

But what if it could be?



A new approach to attribution: visit-level attribution

A new approach to attribution: visit-level attribution

Advances in technology have made this a reality.



It is now possible to take an approach that learns from econometric modelling techniques but is focused on providing near-live data for tactical marketing insight using bleeding edge data science. And which leverages Random Forest Machine learning techniques and session-stitching capability plus the capacity to join any offered data source to fix the incoming clickstream data being generated by analytics platforms. And join data to any marketing channel online or off.

This - **visit-level attribution** which is used by Corvidae – is the creation of a fully attributed map of a customer journey. With a scored value against every piece of marketing collateral that individual has engaged with.

And has **3 clear stages** as follows:

- Stage 1 – Rebuild
- Stage 2 – Unify
- Stage 3 – Attribute

Let's take a closer look at each in turn.

Stage 1 – Rebuild

Using previously developed session stitching Machine Learning processes, data from almost any marketing channel may be added to the Corvidae system for inclusion. Because the system first creates a customer journey from the ground up using raw clickstream data, it is then possible to further 'stitch' in marketing content that overlaps with an individual's journey. Where supplied data overlaps in time, geolocation or other fingerprinted patterns match.

This scores each engagement a visit has - and the probability of its next engagement - based on existing data. This completely rebuilds core marketing data from the ground up around a true picture of customer behaviour.

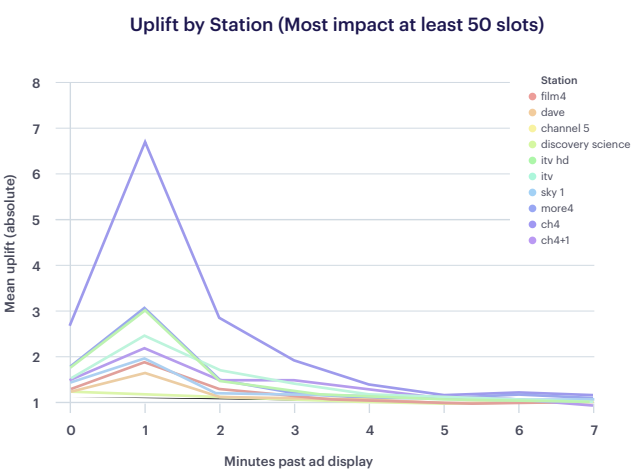
Providing the basis for building a single view of each individual's journey to conversion.

Stage 2 – Unify

This probabilistic approach then enables joining of data from offline activity, store activity and even enables the joining of third party 'walled garden' digital data such as data from Facebook, Google Ads or YouTube.

So, how does this work in practice?

Let's take an example – where rebuilt data in a platform like Corvidae can be accurately mapped very precisely to offline events



If supplied offline activity data is sufficiently granular, platforms like Corvidae can be used to apply further performance segmentation.

where geo and time data – from a range of offline media sources - overlaps with the 500 measured markers held against each customer across a two year lookback window.

Where predicted customer conversion behaviour is measured to have changed for a customer, and they have likely been exposed to an offline marketing event - a TV ad, or instore activity - then that activity is 'stitched' into their conversion journey.

When attribution modelling is subsequently applied that offline event is allocated value just as a digital channel would. Which opens up the possibility to quantify offline activity effectiveness, and allows digital and offline to be held to the same, unified, account for performance optimisation.

And the impact is significant in terms of the insight it can provide.

As a result of joining offline data to online, QueryClick were able to identify significant opportunities for spend reallocation in future

TV and Radio media buying campaigns for a major life insurance customer. In fact, as the table below shows, **39% of TV spend was identified as ineffective** and available for allocation to other channels for **no negative revenue impact**.

TV Spend Effectiveness: Reallocation Opportunities	
	Ineffective Spend
Total Yearly TV Spend	£2,264,000
Savings from Time adjustment	£430,160
Savings from Station adjustment	£311,753
Savings from Day adjustment	£152,209
Total	**£3,158,122**

More starkly, no correlation between radio activity and conversion impact enabled **£2.42m of radio media spend** to be reallocated with no negative revenue impact.

The rebuilding & joining phases above create the most complete possible journey for every identified individual, including an accurate attribution score for each and every content interaction for each and every individual.

Stage 3 - Attribute

While the customer engagement paths generated by the Rebuild and Unify phases are greatly improved data, they still represent a vast and complex relationship which must ultimately be scored to give a single attributed revenue outcome for each engagement point. Considering the incrementality or otherwise of a piece of content on one particular prospective customer journey.

And that is a complex task.

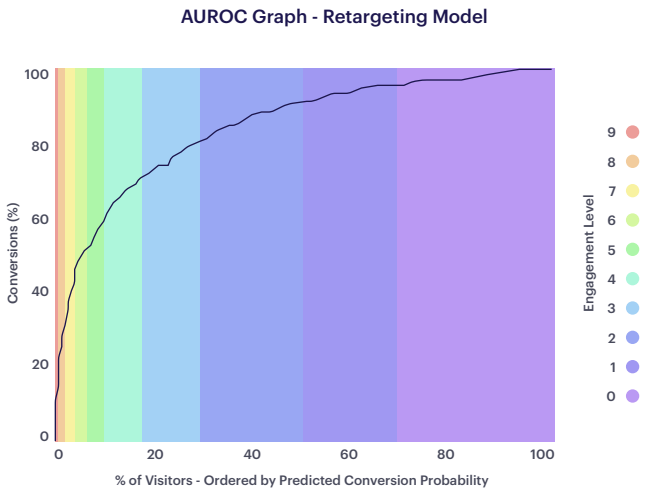
For example. Consider the challenge of valuing the incrementality of a single impression for a single piece of content - which may have many millions of impressions - on a single prospective customer. Who may well behave differently both before, and after, that impression than every other prospective customer. And will almost certainly be in a different conversion phase of their conversion path - or non-conversion path - from every other prospective customer again!

But it is possible, by applying advanced Machine Learning – and neural network based – techniques which are predictive in nature. Like the ones we have outlined to this point, and that are used in sophisticated attribution software like QueryClick's Corvidae platform.

So, what does that look like in practice?

Let's take a closer look at the example of accurately assessing conversion in retargeting activity for a major life assurer. In this case, Corvidae made available a Machine Learning based, predictive view of customer prospect conversion for ingestion into the client's Google's DV360 platform. In order to run an A/B comparison against the in-built 'data-driven' ML targeting offered by the client's existing platform.

And the results were outstanding. Corvidae was able to deliver **4x more conversions at ½ the CPA of Google's model**. Delivering a 50.5:1 ROI and clearly demonstrating the value in taking a predictive, neural network based approach to attribution.



Corvidae's ability to see the whole conversion picture of each individual allows very high performance retargeting.

Prospecting for incremental revenue via programmatic display was also highly successful, delivering a 3.8:1 ROI, and overall Corvidae's automated customer acquisition feeds returned 5.8:1 ROI against media spend in the highly competitive life insurance sector in the UK.

Programmatic Display Corvidae API Performance

	Spend	Revenue	ROI
Retargeting	£4,951	£254,770	50.5 : 1
Prospecting	£110,000	£529,900	3.8 : 1
Total	**£114,951**	**£784,670**	**5.8 : 1**

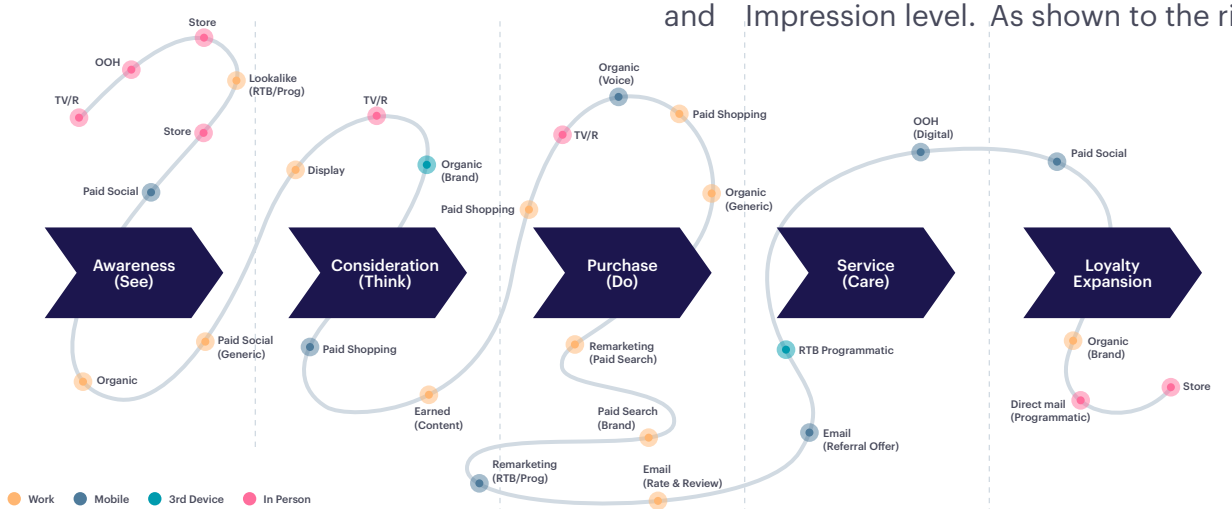

The powerful data views
you need to accurately
drive marketing ROI

The powerful data views you need to accurately drive marketing ROI

By building powerful visit-level attribution data you are able to unravel the complexity of the wide variety of touchpoints on individual customer journeys and accurately attribute the correct level of impact to each.

This enables you to make the direct link between marketing spend and revenue, providing the capability to significantly improve your marketing ROI by making informed judgements based on sound data analytics.

It also opens up access to a broader, tiered level of attribution at a Channel, Campaign and Impression level. As shown to the right.



Channel Level Attribution

ONGOING ROI

Removes 3rd Party Cookie dependency

Optimisation of media spend by channel.

- Spend reallocation (manual)
- Predictive Paid customer acquisition (automated)

Google Analytics

Campaign Level Attribution

ONGOING ROI

Optimisation to individual campaign level.

- Granular campaign performance data (manual)
- Bid adjustment (automated)

Google Ads

Microsoft Advertising

Impression Level Attribution

ONGOING ROI

- Optimisation of social display and video
- Optimisation of individual creatives

TikTok

Corvidae provides the ability to work at both strategic and tactical levels to improve your marketing programmes from the top down and bottom up.



Closing thoughts

Some closing thoughts



A range of factors from increasing media complexity to the need to directly link marketing to revenue mean marketing attribution is the key issue for marketers to solve right now. But our data shows marketers lack access to effective attribution and are facing a myriad of challenges including:

- seriously flawed existing attribution models that are **based on oversimplified and historical approaches** that simply aren't suited to the length and complexity of the customer journey
- an increasingly long and complex customer journey that is undertaken **across multiple channels and devices** both on and offline
- poor underlying data that is largely predicated on a Cookie/Pixel based approach that **generates data which is around 80% incorrect**

Founded in 2008, QueryClick is an independent digital marketing agency trusted by brands over 32 markets and focused on solving the attribution issue for leading brands including Tesco, Wiggle and Vitality Life.

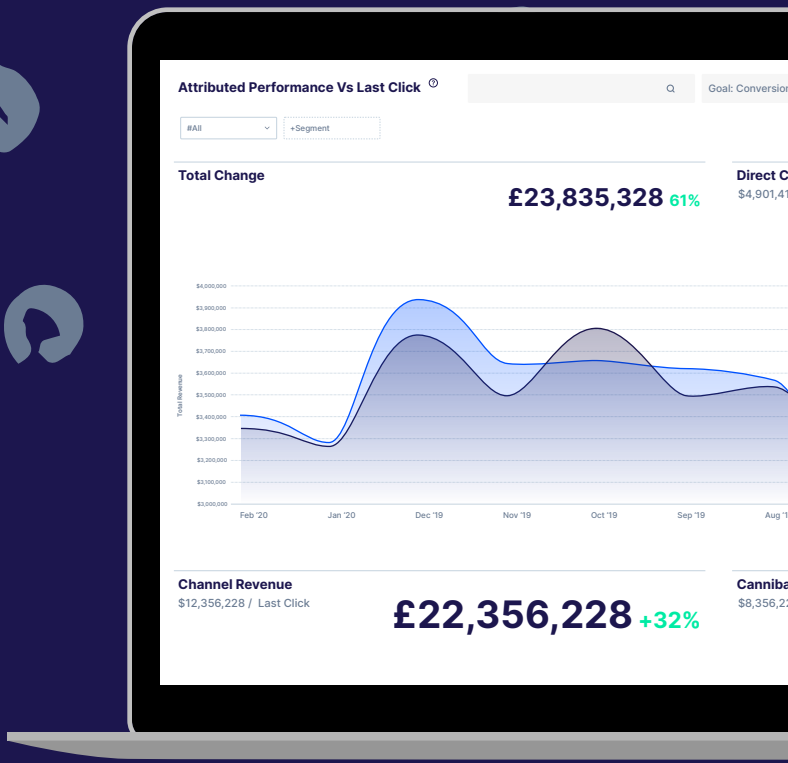
Reaching beyond Google 360 and Adobe Analytics, Corvidae is the only attribution solution which uses sophisticated Machine Learning techniques to completely rebuilds your marketing data to eliminate inaccuracy that prevents effective attribution and enables you to:

- view the complete customer journey **across on and offline media**
- **maximize marketing ROI** by eliminating wasted spend and optimising budget allocation
- correctly value content and **build top of funnel activity**
- automate customer acquisition and **convert customers based on probability to convert**
- enable **dynamic data science analysis and Machine Learning** in action

Talk to us about Corvidae

Corvidae is QueryClick's SaaS marketing attribution platform and the only attribution solution which completely rebuilds your marketing data, reaching beyond Google 360 and Adobe Analytics.

GET STARTED



"Corvidae reveals up to 334% more data for attribution than market-leading competitors. Providing game changing accuracy and predictive performance."