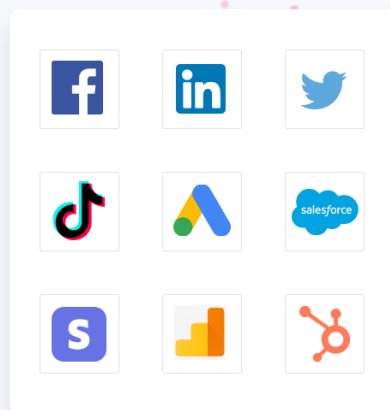


Get full control of your marketing data. 🖱️

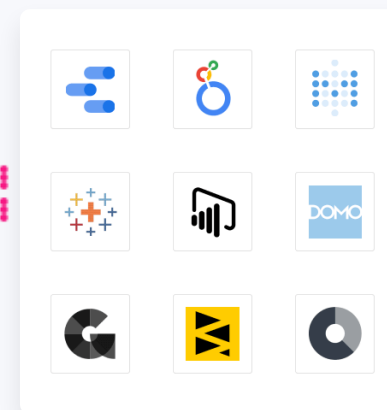
Let's talk



Take full control of all your marketing data



improvado



SHARE



How to Create Your Own Attribution Model

CHRISTOS VISVARDIS

SEPTEMBER 15, 2021



Allocating marketing budget efficiently is problematic without knowing the contribution of each marketing channel to a user's lifetime value and your overall conversions.

**Take full control of all
your marketing data**

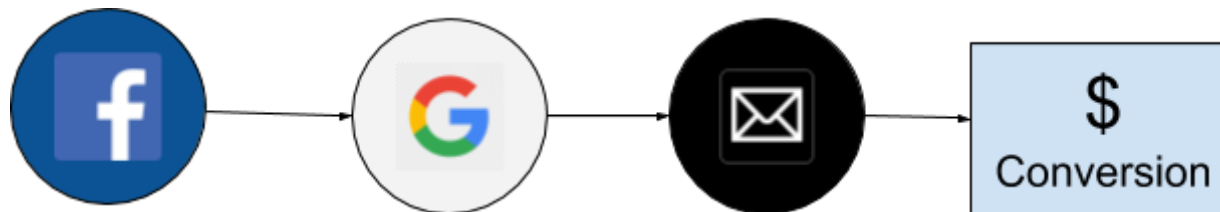
200+ data sources under one
roof to drive business growth.



EXPLORE

The problem arises from the fact that analytics tools and marketing tools report the **last click** that occurred and are biased towards **their own data**. Plus, they can't always detect other channels efficiently.

As companies and marketing budgets grow, customer journeys become more complex and convoluted since the number of marketing channels they use **increases**.



Here's an example. Let's think about a conversion path in which a user first interacts with a Facebook ad, then performs a Google search for the product and finally opens the company's newsletter before converting.

In this case, the same conversion would be reported differently in different tools.

- Facebook would get credit in Facebook Ads Manager, and yet
- Google would get credit as the last paid channel in [Google Analytics](#).

As seen in the example above, relying on third-party tools for attribution **will not reveal the whole customer journey** and will have inherent biases because of the data and the way it is treated under different analytics tools.

Create your own attribution model

The goal is to create your own custom attribution model. This way you can give credit where credit is due when it comes to the different channels on your customer journey.

Having a data-driven model helps you to:

- Allocate your budget to the most effective channels
- Target the most valuable users based on their lifetime value.

Collect data for your attribution model

In order to create an attribution model you need all the customer paths (converting and non-converting) that occurred on your website / app.

A path consists of the touchpoints (clicks) that the user interacted with during a typical conversion window.

Touchpoints are captured through the UTM tags used in your campaigns.

Capturing UTM Tags

UTM tags are parameters appended in the URLs of marketing campaigns that do not modify the destination of the URL but pass on information that can be captured by analytics tools.

For example, a URL from a Facebook campaigns would look like:

[https://www.example.com/?
utm_source=facebook&utm_medium=social&utm_campaign=nameofyourcampaign](https://www.example.com/?utm_source=facebook&utm_medium=social&utm_campaign=nameofyourcampaign)

This URL directs people to example.com. Everything after the ? is not part of the web address but passes on parameters to identify information for the origin source of the traffic. These parameters fire when the above URL is loaded.

The parameters used to identify the sources of traffic are:

- **utm_source:** The channel that brought the traffic (e.g. Google, Facebook etc)
- **utm_medium:** The type of traffic (e.g. social, paid search etc)
- **utm_campaign:** the name of the campaign

Note that you can manually choose the UTM tags for your campaigns, but you can also assign them dynamically. For example, [this article](#) shows you how to use dynamic UTM tagging for your Facebook campaigns.

If you don't use UTM tags for your campaigns, the [analytics tools](#) that you use will capture the URL of the referrer but will label it as organic instead of paid.

Capturing user level data

The second thing to consider is user-level data, i.e. the ability to identify users by assigning a unique ID to them if possible. In that case, you'll be able to have a proper cross-channel and cross-device attribution and not rely just on cookies.

Analytics tools like [Mixpanel](#) or Heap give you the ability to identify device IDs and user IDs, and use the combination of the above dimensions to identify multiple devices for each user across the customer journey.

Capturing conversion & revenue data

Since the attribution model data consists of the converting and non-converting paths of your users, along with the touchpoints that you will get from the UTM tags, you will also need the conversions and the respective value (revenue) for those conversions, likely found in your CRM.

These data points will help you calculate the conversions, revenue and ROI for each marketing channel.

Pull all your data together

In order to create your own custom attribution model without relying on the attribution models your analytics tools provide, you will need to import all the data we described above into one place, ideally a database or [data warehouse](#).

An easy and efficient way to do this is to use an ETL tool like [Improvado](#) which enables you to connect all your marketing data in minutes, saving massive amounts of time and developer resources.

Once you have all the metrics and dimensions required for the attribution model imported into your database, you should consider a few factors for your model.

Decide on an attribution window

Decide on an attribution window based on your data and business considerations like the purchase cycle for your products. The attribution window is the time period during which a purchase should be credited to a touchpoint that happened within that period.

Different industries have different purchase cycles and that affects their attribution window.

For example, it takes much longer for a customer to decide on the purchase of a vacation package worth thousands of dollars than to buy an inexpensive t-shirt. Generally, expensive purchases have long cycles that might take months and

dozens of touchpoints to complete, whereas cheaper and impulse purchases might take only a few hours from the first touchpoint to the conversion.

Analytics tools like [Google Analytics](#) provide reports that help you see the distribution of users based on how long it took them in terms of time and number of touchpoints to convert.

	Time Lag in Days ?	Conversions ?	Conversion Value ?	Percentage of total	
				Conversions	Conversion Value
	0	9,586	\$15,337.60	61.67%	61.67%
	1	577	\$923.20	3.71%	3.71%
	2	376	\$601.60	2.42%	2.42%
	3	337	\$539.20	2.17%	2.17%
	4	258	\$412.80	1.66%	1.66%
	5	226	\$361.60	1.45%	1.45%
	6	272	\$435.20	1.75%	1.75%
	7	217	\$347.20	1.40%	1.40%
	8	192	\$307.20	1.24%	1.24%
	9	168	\$268.80	1.08%	1.08%
	10	170	\$272.00	1.09%	1.09%
	11	154	\$246.40	0.99%	0.99%
+	12-30	3,012	\$4,819.20	19.38%	19.38%

Time lag report in Google Analytics

Building your Attribution Model

There are two widely accepted data-driven models for attribution:

1. Shapley Value
2. Markov Chains.

The inputs needed for both models are the touchpoints and the conversions, which as stated above are part of the data that you will import into your database.

Using the Shapley Value Attribution Model

Shapley Value - named after the Nobel Prize-winning economist Lloyd Shapley - is a game theory model for cooperative problems. In other words, the model tries to assign credit to different parties that contributed to a total value. This is also the question we're trying to answer with an attribution model, namely how much credit every marketing channel should get for making a user convert along the path.

The Shapley model is also the one used by Google for their own data-driven attribution model in Google Analytics 360, however by creating your own model you will have better control over your data and will avoid the biases that Google Analytics might have by giving more credit to Google Search.

In order to calculate the contribution of a channel under the Shapley Value model, we compare all the different permutations of paths and touchpoints that occurred. For example, we take two paths that differ by a single touchpoint and we assign the difference in total value to that extra touchpoint, since it is the only difference between the two.



[Shapley Value example by Google](#)

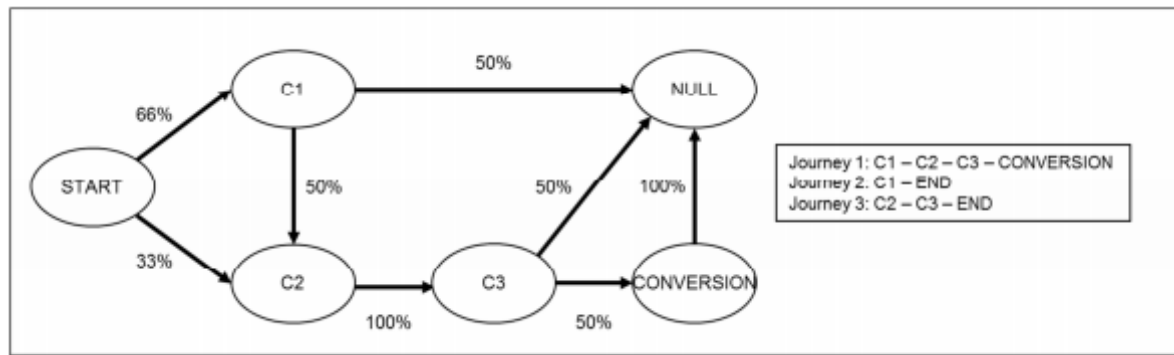
Then we compute all the permutations and we assign credit to each channel accordingly. Thus, the model calculates the probability of conversion when a specific channel is present in the conversion path.

Using the Markov Chain Attribution Model

The [Markov Chain](#) model - named after the Russian mathematician Andrey Markov - describes the sequence of various events and tries to make predictions based on them. Once again, we try to assign the probability of a user converting when exposed to various marketing channels.

The Markov Chain model assigns credit to marketing channels by calculating the removal effect. The removal effect depicts what happens when we remove a marketing channel from a path and see how many conversions take place without that channel.

By calculating all the different permutations of paths and the removal effects for every touchpoint, we end up with a probability to convert for each marketing channel.



[Source](#)

In both the Shapley and the Markov model, the output is a matrix of all marketing channels and a probability or credit for all conversions that occur thanks to each of those channels.

Channel	Last-Click Conversions	Data-driven model conversions
Paid Search	5	3.7
Social Media	3	4.4
Newsletter	2	1.9
Total	10	10

The above table is an example of the output of a custom attribution model compared to a standard last-click model. Note that the total number of conversions is the same for both models, but what changes is the allocation between different channels. Moreover, the data-driven model can have fractional conversions, since credit for a conversion is given to multiple channels.

You can also calculate the revenue and ROI for each of the channels since you have conversions, revenue and marketing cost in your database. This will help you allocate your marketing budget across channels.

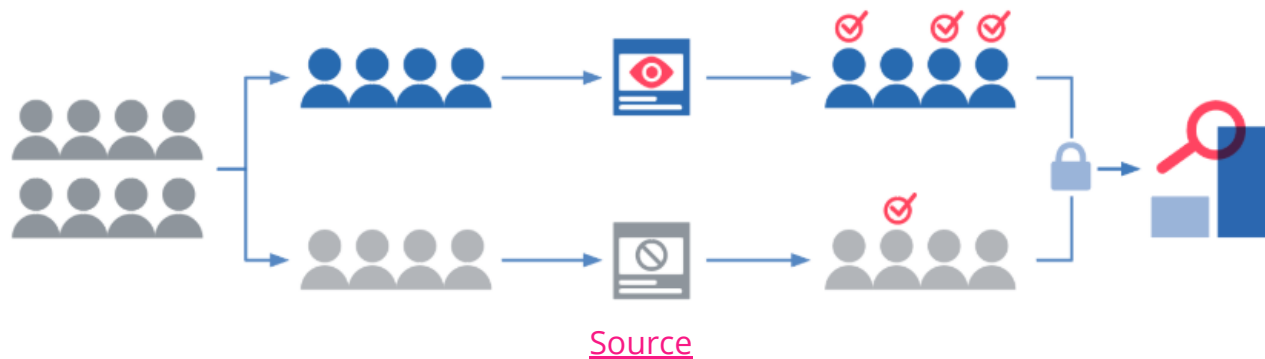
How to run a “Lift Test”

In the models and data mentioned above we talked about capturing touchpoints via UTM tags. UTM tags occur through clicks, which means that there are channels (mainly social media) that will be underrepresented due to the lack of impressions.

In order to incorporate impressions to your model, you should consider running lift tests for channels like Facebook and Instagram as they rely on impressions more than other channels.

A lift test is a randomized control test where we randomize an audience into a test and control group. We only show ads to the test group.

The difference in conversions between the two groups is known as lift or incrementality and represents the real impact of a channel's ads on the audience. Moreover, since this is based on the concept of randomized control trials, it also incorporates the concept of causality, meaning that we know that it was the ads that caused the extra conversions.



A good practice is to regularly run lift tests (e.g. once a quarter) so that you can see the effect of Facebook, Instagram and other impression-heavy channels and calibrate your attribution model accordingly.

Lift test vs attribution model

Both attribution models and lift tests are useful and should work in conjunction to give the best possible results. They both have their advantages and limitations, as you can see in the table below:

Lift Test

Attribution Model

Accurate results

Approximation based on model

One data point in time

Tool that can be used on a daily basis

Holdout

No holdout

Based on results, not on arbitrary rules

Rule-based unless you build a **data-driven** model

Causality

Correlation

Baseline (organic, brand effect) is taken into account

Gives little to no credit to organic

Impressions are taken into account (but not segregated)

Impressions hard to track (depends on channel)

Not all channels have lift tests (imperfect alternatives like matched market test, before-after etc exist)

Models all digital channels. Offline is problematic

Incorporating Offline Activities into your attribution model with “Matched Market Tests”

For offline activities (TV, billboards etc) it is recommended to run Matched Market Tests, where you take two similar geographic areas and use them as a test and control group to get results. The calculation of results is similar to a lift test but we have to acknowledge that this is not a perfect test, as the audiences are not randomized.

You can also employ before-after tests, where you compare two periods of time with different marketing activities.

Something we have to take into account when running all sorts of tests is duration and seasonality.

A rule of thumb is that tests should last for at least one week (and ideally 4 weeks or more), since there might be fluctuations of conversions for different days (e.g. a lot more conversions on weekends compared to weekdays). Moreover, you should avoid periods during which you experience big increases or decreases (e.g. Christmas, Black Friday).

Our recommendation:

Check out [The Best Facebook Ads Report Templates to Try in 2021 \[Updated\]](#)

[The Best Marketing Analytics Tools & Software for 2022](#)

[Funnel.io Alternatives and Competitors in 2021](#)

[Supermetrics VS Funnel.io VS Improvado - A Comprehensive Comparison](#)

Free Guide

How to Become a Marketing Analytics Master



Analytics Master

A tactical guide for marketing & analytics leaders

[DOWNLOAD NOW](#)

Company

[ABOUT](#)

[CUSTOMERS](#)

[PARTNER WITH
IMPROVADO](#)

[EVENTS](#)

[CAREERS](#)

[WE'RE HIRING!](#)

Products

[DATA EXTRACTION](#)

[DATA TRANSFORM](#)

[POWERFUL API](#)

[HOW IT WORKS](#)

[PRICING](#)

Explore

[INTEGRATIONS](#)

[CONTENT LIBRARY](#)

[USE CASES](#)

[DASHBOARDS](#)

[MARKETING TOOLS](#)

[REVIEWS](#)

Support

[HELP CENTER](#)

Improvado - The Enterprise Marketing Data Platform

Improvado automates the annoying parts of data management. No more manual anything. Just automate.



From the blog

[Top 12 Data Analysis Software: How to Choose the](#)

LEGAL

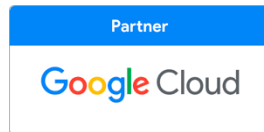
COMPARE TOOLS

IMPROVADO BLOG

One That Will Drive Your Growth

Data Ingestion: Where Every Healthy Data Ecosystem Starts

Data-Driven Marketing 101: How to Become a Digital “King of the Hill” with Data Analysis



San Francisco | **Headquarters**

2800 Leavenworth St, Suite 250, San Francisco, CA 94133

-

Moscow

2-Y Kozhevnicheskiy Pereulok, 12 stroenie 10, Moscow

Amsterdam

Keizersgracht, 1016EH Amsterdam

-

Tomsk

Lenin Ave, 94, 3, Tomsk

