Applications of Multi-Touch Attribution Modelling

Leonel Fernández

leonel.r.fer@gmail.com

July 2020

Dissertation to apply for the degree:

Master's in Management & Analytics

Supervised by:

Ph.D. Santiago, Gallino

Abstract

The Digital landscape has evolved vastly since the early 2000s in terms of analytical tools and tracking software. With the Rise of 4G to 5G, smartphones have become the norm when surfing through the web. New problems arise in terms of measuring business performance like Cross-Channel and Multi-Channel Attribution. Companies are selling more products and services on their Websites and marketplaces than ever before. Brands must become digital natives and translate all of their offline business into the internet.

When Brands invest in multiple marketing channels and those channels mix up in the Customer Journey, new measurement problems arise. Based on the current standard methodology on web analytics, companies track their conversions (signups, subscriptions, orders) and assign each channel's attribution using simple heuristics. In other words, simple decision models. It has been vastly studied that **single-touch attribution does not perform well under complex business scenarios** like those observed nowadays.

Attribution modeling has been a hot topic in the last decade due to **the rise of Machine Learning and data mining**. Nowadays, there are two current trends. The problem can be analyzed from a Machine Learning standpoint, understanding that it looks like a Classification problem with a Binary Outcome (0/1). On the other hand, Shapley Values and Game theory also adapt efficiently to the question, where every player gets credit for contributing to conversions.

Given that there are different **state-of-the-art models** which perform better than others and that multiple papers are trying to improve robustness, predictive accuracy, interpretability, this thesis **will focus primarily on applications and interpretability of the model**. Most of today's Marketing Managers and teams find it extremely hard to use and apply these types of models due to the complexity of the topic and black-box models, which have little to no interpretability. The idea is to **encourage more companies into the MTA** landscape to test their models and optimize them specifically for their industry in this work. Additionally, to my knowledge, there is no research on Markov Chains applied to Subscription Business Models that are substantially different from E-Commerce Customer Journeys.

Keywords: Multi-Touch Attribution, Logistic regression, Markov chain, Google Analytics, Adform, Digital Analytics, Click Stream, Data-Driven Modelling

Preface

This thesis is an extension of my interest in Marketing Analytics since I first started to work in the industry. In 2010 I landed my first job as a digital marketing intern at an agency where I learned about Google Analytics, Programmatic Ads, and Advertising. Years later, I had a transition towards Data Science, Big Data, and Analytics, given that the Digital World is growing each year exponentially in terms of platforms and information.

I would like to thank my family for the constant support and understanding of the hours and effort that a Master's degree requires.

Martin Baigorria has been a source of inspiration for me. He pushed me to get into the Data Science world, where I could find a new passion. Since day zero, it has been a game-changer with hundreds of new doors and opportunities arising every day.

Special mention to all the professors of the Di Tella Institution that inspire us every day.

Table of contents

Abstract **2**

Preface **3**

Introduction **5**
- The Omnichannel world 7
- Motivation 8
- Problem Description 8
- Thesis Objective 11
- Literature Review 12

Attribution Modelling **14**
- Heuristics 14
- Marketing Channels 15
- Machine Learning 17
  - Markov Chains 18

Case Study **21**
- Dataset 21
- Data Preparation 21
- Exploratory Data Analysis 24
- Model 27

Results **32**
- Predictive Accuracy (ROC) 34

Applications **38**
- From Aggregated to Individual Level 39
- Evaluating Conversion Value 41
- Impression Level Data 43
- Campaign Level Information 44
- A/B Testing Attribution Models 46
- ROAS Modifiers 47
- LTV 50

Business Recommendations **52**

Conclusions **56**
- Limitations of this study & Further Research 58

List of Abbreviations and Acronyms **60**

References **61**
  - Articles 64

Tables & Figures **65**

Introduction

In Marketing and Behavioural Economics, the Customers Journey topic has been widely studied with lots of papers, experimentation, books, and more. It is well known that customers spend time before buying any type of product, which differs within industries, products, and more.

Customers perform some fundamental analysis between alternatives, evaluate cost/benefit analysis functions, and test them out. It is important to remark that as the products' price increases (or the decision becomes of higher importance), we tend to rationalize the purchase even more as the room for any mistake decreases. An example is buying a Car, an Apartment, or choosing a school for kids.

The holy grail for marketers is to **understand all the motivations and touchpoints contributing to a customer buying** one brand's product against competitors. By understanding the underlying logic, they can optimize their budget to enhance the most critical touchpoints, increasing Total Revenue, and Total Profit as they become more efficient with marketing investments.

The attribution problem occurs in people's regular daily life. It is often hard to adjudicate the right reasons for different outcomes. For example: If we want to lose weight, and we know the marginal contribution of each food we ingest, we could then focus on the most significant foods that will generate the biggest impact. If doctors could know all the marginal effects of the relationship between actions/foods and Breast Cancer, it would be relatively easy to change those habits or activities.

As noticed, the problem is that these Cause-Effect relationships are impossible to understand with a high degree of confidence. There are multiple methodologies and experimentations in statistics, economics, and general Science to understand these relationships, but it is still hard to grasp them. In the field of Behavioural Economics, there is plenty of research pointing out that customers tend to be irrational in some cases and that their preferences seem to change given some cues *[1-Kahneman]*.

A classic example in the Attribution world is the Car Sale. If a customer talks to a salesperson and buys a car, it is straightforward to give the salesman commission. However, what if, before buying, the customer spoke to four salespeople on different days. Can the owner split the commission between the four? Is that accurate or true? Did the four contribute in the same percentage to generate the sale? Alternatively, maybe the last one was the one responsible for generating the deal. If the owner does not understand how the four salesmen interacted with the customer, he will probably distribute commissions mistakenly. By doing this, he would be rewarding the incorrect touchpoints for their effort. Thus, sales can not be maximized in the future as the incentives are incorrectly distributed to the wrong factors.

Going back to the main statement, Marketers and Management, in general, need to understand the factors contributing to either the success or failure of their selling proposition. Specifically, this paper will focus on a subset of this problem, **evaluating how different touchpoints contribute together to generate or not a transaction (conversion) in a digital store**. The players in this game will be every marketing channel that interacts in the Customers Journey. Not everyone will be considered, only the ones that are Paid Channels.

This thesis will explore this topic mainly focused on the Marketing Investment side. How are companies optimizing their budget by selecting almost hundreds of different advertising opportunities to increase their overall Return over Investments (ROI)? Nowadays, the solutions remain based on heuristics or gut-decisions within the management. Digital Marketing can track each campaign's performance but cannot see the big picture of how Multiple Touchpoints define the Customer Journey. If **Marketers can reward and invest in more efficient touchpoints, they can maximize revenue** and understand their customers with a higher level of confidence.

Focused mainly on Digital Marketing, which is easier to track, we will explore the different state-of-the-art Machine Learning solutions, focusing primarily on the **applications of multi-touch attribution models for enhancing ROI**. It is important to note that the real world is often more complicated than Online Data. Customers interact Offline/Online with brands, with stores, friends, and advertising in the streets, but all this complexity is hard to model. New startups tend to focus only on Digital Marketing, which reduces the attribution friction and makes the problem easier to model.

The Omnichannel world

The world has changed tremendously since research in the Marketing and Economics fields started to unfold. Customers' Journeys are exponentially more complicated than going into several stores to evaluate some products. The amount of products and services that can now be consumed worldwide is at least 100 times larger than in the previous century.

Consumers can access hundreds of products in seconds with their Smartphones and compare them almost instantly. Furthermore, the Ads industry has evolved to deliver Billions of impressions every second, making the market noisy and overcrowded (Mckinsey, 2020). The rise of 4G, 5G, and smartphones allow customers to search about almost everything while walking on the streets or commuting to their jobs. This means that what we understood as a simple journey has evolved into an entangled web where customers access information from different places like Smartphones, computer PCs, home PCs, tablets, Notebooks.

Luckily, infrastructure and tracking technologies are up to date with this type of problem. There are plenty of technologies that allow marketers to track efficient customers around the web. The current stack is based primarily on Web Cookies, letting companies know information like IP Addresses, Timestamps, Events, Device ID, Operating systems, and more. There are millions of log-level data being generated daily on almost every internet-connected device. Tools like Google Analytics, Adobe Analytics, Amplitude, and many more are taking the challenge of tracking internet behavior.

The real problem is trying to make sense of all the information that is generated. For every purchase made, there are contributing factors that need to get the credit. Even if nowadays companies can track efficiently the different touchpoints a consumer interacted with before buying, it remains challenging to understand how they mix up to lead to a conversion.

Motivation

Companies often **struggle to create the perfect Marketing Mix**. It is often viewed as a Marketing Problem where they study customers' behavior in an old-fashioned way with market research, focus groups, and more. With the rise of the Internet and Big Data, these problems can be transformed into Machine Learning problems and be solved with efficient algorithms that maximize the given function.

Often known as Budget Allocation Optimization (BAO) [2-Aggarwal, 2018], the topic is constantly being studied from a Finance perspective where the capital needs to be distributed accordingly to each asset to maximize revenue or profit. Marketing investment can be thought of in the same way, where there are allocation opportunities (channels, platforms, campaigns), and each one produces an outcome.

Almost 95% of companies are still investing billions of dollars with simple heuristics, which only contemplate one touch into their attribution models. I firmly believe that every company should move out of simplicity and interact with complex models and frame the problem like a Machine Learning one. There is a very famous phrase that says:

"Half the money I spend on advertising is wasted; the trouble is I do not know which half."
John Wanamaker *(1838-1922)*

Marketers need to be more innovative and become way more efficient when allocating spend. There is a huge trend to measure absolutely everything, but models to understand all the inflow of information have not evolved enough. In this research, we try to remove much of the complexity of the state-of-the-art models to focus on real applications that marketers can use to minimize money wasted in non-performing channels, campaigns, and assets. Thus, the motivation is to create new opportunities where MTA models can be applied in real-world scenarios that can profoundly change the industry.

Problem Description

The problem is within the bigger question: How can companies be more efficient in allocating their marketing budget over thousands of options? During the last decades, the tracking problem has been solved, meaning that the Digital world is fully traceable (cookies-based solutions). Furthermore, statistical approaches are trustworthy and reliable in the offline landscape, allowing marketers to get a sense of the return on ads.

Nevertheless, the problem remains as the tools available lack “intelligence” to generate good recommendations on where to spend the money when there are multiple channels, different types of ads, and several different customer journeys that grow exponentially.

The problem of attribution has been around for a long time in many different forms. There is an outcome, and there are “generators” or reasons for it. In short, causality. If the reasons are understood, then results can be replicated, expanded, or improved. Some examples:

- John gains 5kg of weight. Why? How much of the extra 5kg can we assign to the variables involved in this outcome. What % is due to flour products, animal fat, smoking, depression.
- Lionel Messi Scores a Goal. Why? Who else was involved in that goal? Iniesta passed the ball, Puyol recovered the ball, Suarez cleared the defense.
- A student scores 1500 in SAT Exam. Why? Was it because he practiced all model exams or the preparation with this friend or for “X” days of practice or a specific youtube video.
- A Car Agency sells a car. John is the salesman. Did he sell it alone? Did the buyer interact with multiple people before buying, like Juan from Customer Service, Lisa at the Front Desk?

All the examples stated above share some similar traits. First, the outcome is specific and known. It can be measured precisely, and it can be both qualitative or quantitative. Second, there has to be some intuition or some initial variables that can be known or not that affect the outcome variable.

As it is starting to look like a Supervised Learning problem from the Machine Learning Field, this sounds familiar. The attribution problem belongs to this family of questions, only with some twists, but the idea remains the same. Like linear regression, there are independent variables, one or more dependent variables, and we need to understand the correlation between them. This gives much insight into the dependent variable. It can now know what percentage of the variables explain the outcome, the weights, future results, estimate new data points, and more. To see the matter, if a coach is shown the following sequence, can he/she evaluate the outcome?

Puyol Recovers Ball > Pass to Iniesta > Pass to Suarez > Pass to Xavi > Pass to Messi > ?

What is the probability that this sequence ends in a goal? If the actual chance is > 60%, the coach can adjust his strategy to look more for this sequence to happen. In an E-Commerce:

User googles Beauty Products > User Watches a Youtube Video on how to make up > Users sees company ad in the subway > User searches for the Brand > User Buys ($100 sale)

It is difficult to know precisely what occasioned the purchase since there are multiple factors involved. Many unconscious processes are working in parallel in the customer's mind that they are not aware of. If the customer is asked why they bought the product and assigned a weight (%) to each of the paths stated they would not know. It makes sense, as there are multiple drivers. The company can understand that if they scale all their customers' paths, they can explore the data and see some patterns. What they will certainly observe is that some paths have lots of conversions while others have low conversions. They could find that some of the drivers they thought were useful are not present in any converting path.

In Digital Marketing, expressly, the problem is stated as follows:
"When a user buys or subscribes from a website, do we know what contributed to that action?"
The interest is not in the psychological or motivational drivers (both studied in marketing) but in its actions to incentivize the purchase. There are multiple factors involved like Branding, Social Media, Ads, Product Quality, Pricing, and many more. There are no models that contemplate all variables as the real world is way too complex to simulate. Despite this, using some Machine Learning Methods and optimizing only the Ads part can bring outstanding results and performance for companies using only heuristic methods.

Thesis Objective

There are over 50 papers that have approached this problem in excellent ways. Multiple experts from Machine Learning, Statistics, Economics (game theory) stated plausible theoretical solutions for solving real-life company scenarios.

Thus, this thesis's objective is to provide a more precise understanding of the applications of multi-touch attribution models and how to evaluate and compare the models while also getting a better grasp of the data work that is needed. By this I mean, what does the data look like and how to obtain it. Clear examples of what it would look like to deploy a model in a Company and stress-test the model to improve performance.

This thesis will not aim to improve any of the existing methods but only apply them to actual data. By using state-of-the-art models, the idea is to encourage other Data Scientists or Marketers to apply their models to their companies.

In short, use state-of-the-art attribution models to generate a real impact on a Business and explore several plausible applications and methodologies to deploy, test, and improve it over time to create lifts in ROI (Return over investments).

Literature Review

The attribution problem has been vastly studied in the scene, with great accomplishments. Starting in the early 2000s, with the rise of Machine Learning and computing power, the problem was classified as supervised learning. In Data Science, supervised learning is the task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples.

One of the most cited papers behind attribution is the solution proposed by [Shao and Li, 2011] using a bagged logistic regression. The first approach was to model the data to look like a **logistic regression** (supervised learning). Each path contains a series of steps that finish or not in a conversion. To transform this data into a regression, the idea is to label each of the paths with a **Binary Variable (0/1) Y to state if it converted or not**. Each path in the customer journey to conversion could be thought of an independent variable that contributes to the final state 0 or 1. It uses two comparable models, bagged logistic regression and a probabilistic model. Bivariate metric uses the average misclassification rate and average variability of the estimate to assess the performance of the model over n iterations. By doing this, the model can estimate approximately how much contribution is made by each independent variable to the final outcome. This research also uses **bootstrapping bagging that is an ensemble meta-algorithm** focused on improving the stability and accuracy of the models used in classification and regression. It also helps reduce the variance and overfitting.

Other researchers have also used Regressions methods to solve the attribution problem and approached it mixing Economic Theory and Game Theory. Zhao et al [2018] used dominance analysis and relative weight analysis, focused at finding the coefficient of determination of the regression model as attribution values. Dominance analysis is specially important as it ensures that all touch sources are considered for calculating the attribution values. By using Shapley Value, the dataset is now thinked of as a cooperative game where every contributes to the final outcome which is Conversion (0/1). Every player in the game has a specific contribution or weight, thus, Shapley Values method can allocate the corresponding weights. As the authors state, their solution has an increased efficiency at scale and is also able to correctly model the order of the channels visited by the users. It ensures efficiency and fairness.

The third approach usually observed in research is **Markovian Graph-Based modelling**. The most cited paper using Markov is Anderl et al.[2013] that follows the work of Abhishek, Fader, and Hosanagar (2012). They propose a dynamic hidden Markov Model, based on individual consumer behaviour. It is often difficult to compare and evaluate the results of different models because they are executed on different datasets which could affect the results. Generally speaking, digital tracking and analytics's data is often messy. Nevertheless, the hidden markov model proposed by Abhishek et al outperforms logistic and shapley value models on its root mean squared error and log-likelihood. Six different criteria were used to assess the attribution model in practice, which was Objectivity, Predictive Accuracy, Robustness, Interpretability, Versatility and Algorithmic efficiency. Since the data is highly imbalanced, the authors use ROC AUC score for predictive accuracy.

Overall, this review of existing literature on attribution modeling indicates that there has been a rich cross-pollination between fields such as Computer Science, Economics, Statistics and Marketing. From an academic standpoint, progress is observed over the last decades. But, in the business world, it has been difficult for marketers to apply these algorithms into real life scenarios to optimize their budgets and decisions. Some of the big channels such as Facebook or Google are starting to add Data Driven models into their platforms, but there is still a lot of room for progress, applications and academic studies to improve the Multi Touch Attribution Landscape.

For this reason, this paper aims to continue the exploration of business applications of multi touch attribution so that marketers can have a richer set of tools to expand and optimize their budgets.

Attribution Modelling

Attribution modeling is the set of rules that determines how the credit for conversions is allocated to the touchpoints in conversion paths. Conversion paths are those paths along which conversion has taken place (IAB, Attribution, 2012). A touchpoint can be defined as user interaction with a business through a website or any other application. In marketing, it describes the ways information is displayed to the prospective user. Attribution modeling maps the user journey from conversion to the user considering every touchpoint along the way. It helps marketers to see the impact of the different touchpoints in user conversion, and it can be used to optimize their marketing spend. The idea is to obtain the highest rate of investment (ROI) and conversions possible. E.g., 80% of the conversions come from Display Ads while the rest come from Social Media. This would help marketers to invest more in Display Ads as compared to Social Media marketing which would help in increasing their ROI

Heuristics

The simple decision framework is usually used in Google Analytics (K.Bill Attribution Playbook GA, 2012) as the de-facto solution for web tracking. Other channels such as Facebook, Pinterest and Google Ads use one of these simple heuristics as the default attribution.

First touch

The First Touch model allocates 100% of the attribution or credit to the channel that drove a visitor to the website or application for the first time. It is the first channel that the user interacted with before converting to the site. It is also known as first-interaction. It belongs to the family of single-touch models.

Last touch

The Last Touch attribution model assigns all the credits to the last touchpoint in the journey. As the first touch, it is simple to implement but only considers the last touchpoint leading up to conversion and thus just gives an idea of what happens at the end of the journey. This model is the easiest to apply since the only needed information is whenever the user converts. In the first touch, we should register both the first landing and the conversion one.

Linear

The linear model assigns credits equally between touchpoints along the journey. It is better than the last touch and the first touch as it explains the whole journey and not just one touchpoint. It is the simplest of Multi-Touch Heuristics, considering an equal weight to all the paths involved. However, it is not valid in all cases that each touchpoint contributes equally to a conversion.

Time Decay

The time decay attribution model assumes that touch points closer to conversion along a journey contribute more and should receive more credits. It considers multiple touchpoints as opposed to single touchpoints, and it makes sense that the ones closer to conversion would have more impact on it. However, in some cases, initial touchpoints might have a more significant effect which this model tends to ignore.

U Shaped

The position-based attribution model, also called the U-shaped attribution model, assigns 40% of the credits to the first touchpoint, 40% credits to the last touchpoint, and 20% of the credits are evenly distributed to the rest of the touchpoints. It addresses several flaws from the previous model by emphasizing the middle part of the journey and still considers the first and last touchpoint's contribution.

Marketing Channels

For the MTA Model, every touchpoint is a marketing channel involved in sending traffic into the site. A Marketing Channel can be potentially any website on the web that contains any link forwarding to the website analyzed. As there are millions of websites, the idea is to group them into several categories for simplicity in practice. The main drivers of traffic into any website come primarily from Google or other search engines like Yahoo or Bing. This makes sense, as it is the initial place where every visitor starts from when opening the browser.

Every link on the web can contain relevant information on the precedence of that click. For example, it is commonly used worldwide to use UTM parameters in the URLs. As defined by Crazy Egg (a tracking solution):

UTM stands for Urchin tracking parameters. They are strings of data that we add to our URLs to see where other traffic comes from.
www.yoursite.com/pricing?utm_source=active%20users&utm_medium=email&utm_campaign=feature%20launch&utm_content=bottom%20cta%20button

The code above extending the URL contains different attributes that can store the data in a server for every user that clicks a link. The idea is that any marketer using the UTM parameters can track efficient traffic coming from different sources and campaigns. If 500 campaigns are running on Facebook sending traffic to the same website, the UTM helps identify every campaign and every ad.

In this case, the dataset presented is summarized into the top 20 sources, containing around 99% of the traffic incoming to the site. The most important generally for all businesses are:

- Facebook: Traffic coming from the Social Network only from Paid Campaigns. Organic posts and Fan Page links are directed to Social Source.
- Social: Social Networks organic traffic from the feed or profile page. Includes Pinterest, Tiktok, Instagram, Facebook.
- Google: Google Paid Campaigns. Contains Search, Display, and Programmatic from Google.
- Organic: Organic landings coming from search engines (not paid)
- Email: Traffic coming from Email Campaigns sent through the CRM.

It is important to highlight that around 70-80% of the marketing budget is spent on Facebook for this particular business, making Facebook the primary traffic source.

Understanding the channels also means understanding that every channel serves a purpose in the Customer Journey's purchase. For example, Pinterest and Youtube are focused on Brand Awareness, meaning they are top-of-funnel channels, maximizing views and reach. On the other hand, channels like Google or Facebook (retargeting) focus on finishing conversions for people with higher intent to buy (end of the funnel).

By nature, attribution models tend to overestimate the end of funnel channels and minimize awareness channels that are top of the funnel. For this reason, marketers that know their channels should be involved when building the MTA model.

Machine Learning

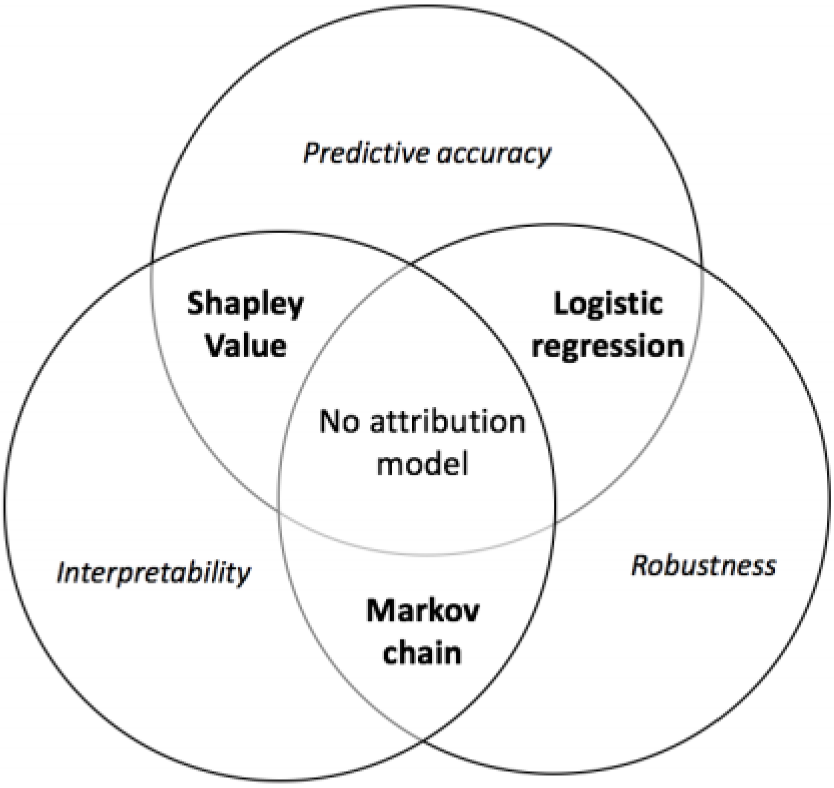
Algorithmic attribution is another approach that can be used to solve the attribution problem. In this sense, the heuristics (simple rules), are changed into algorithmic decisions that Machine learning models back up.

To summarize, machine learning is a field of computer science that uses statistical techniques that allow the models to improve their learning over time, always based on the quality of data in the input. Broadly, there are two significant types of categories, supervised learning and Unsupervised learning. Multi-Touch Attribution relies on the Supervised kind of problems since the outcome is known. The Conversions are known, so for a given sequence of paths, the data contains if it finished or not in a conversion. The data is said to be labeled. The model will optimize the levers and relationships between the dependent variable (convert or not) against the different independent variables (in this case, the paths) to predict accurately whether the sequence ends or not in a conversion (1).

As the model ingests more and more information, it should predict more efficiently the different paths and learn the “true” probabilities for each. The ML experts’ question is what model to use to “learn” the relationships and predictive approaches. Since 2006 many approaches have been developed, and the state-of-the-art has been shifting from Logistic Regression Approaches to Markov Chains and Neural Networks. It is well known that Markov Chains and NNETs have been shown in multiple papers as the best solutions to model this type of problem. As stated before, the idea is not to enhance or modify the state-of-the-art models but focus on the actual applications that can be explored by using them.

Nevertheless, it is worth analyzing the whole set of models applied to MTA to understand from the simplest model to the latest. As seen in the industry, many companies tend to deploy models using simple solutions that are easy to maintain and adapt. The output should look similar between models, and the real purpose of this analysis is to extract value from them. The Venn diagram below shows no perfect solution as every model has its advantages and disadvantages.

Figure 1. Venn diagram stating the different attributes of Attribution models based on the algorithm



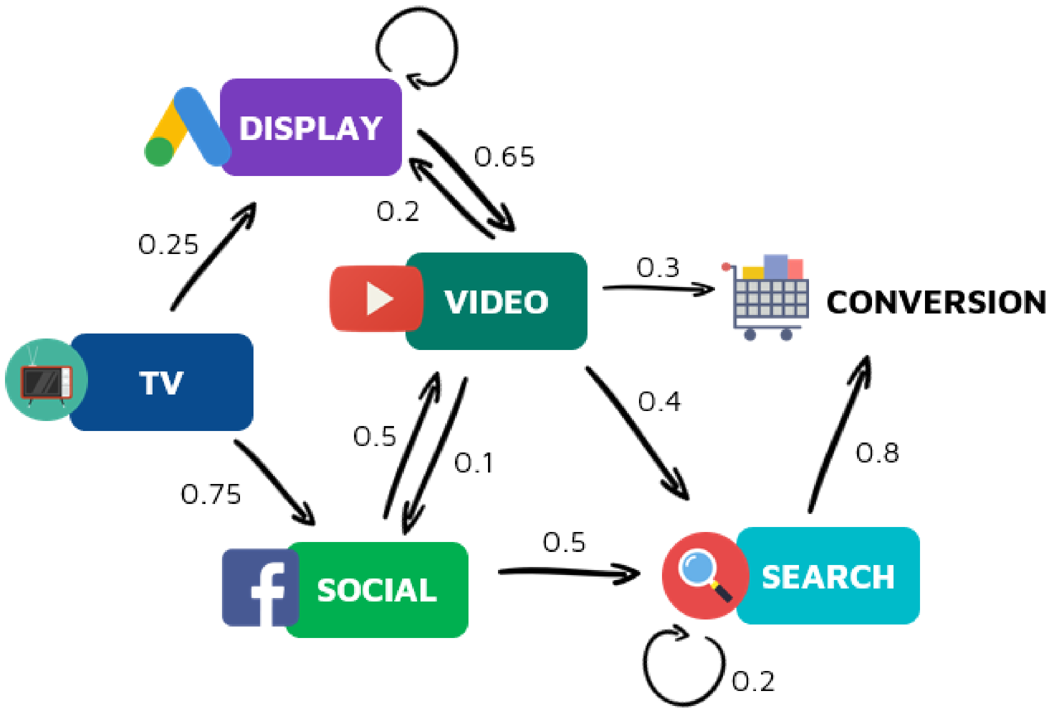
For this case study, Markov Chains are used as the model due to its interpretability and robustness while sacrificing predictive accuracy. It provides many of the benefits of other models and contemplates the sequence of events (especially when higher-order states). Also, it can be easily modified and prepared according to new variables or touchpoints.

Markov Chains

As the experts tried to approach the MTA problem with diverse techniques, Markov Chains (Wangenheim, F., & Schumann, J. H. 2016) stood out. It maps transitions of events from one state to another according to specific probabilistic rules (Keilson, 2012). One attractive attribute is that the probability of transitioning to another state depends only on the current state and time. It is frequently employed in Economics, Game Theory, NLP, Finance, Genetics, amongst others.

The attribution touch-points look similar to a sequence of states that begin at a start point (first touch-point) and end in the last state that can be considered a conversion (0/1) (Anderl et al., 2016; Norris, 1998). It is important to denote that the timeframe needs to be constrained and limited to a specific date range for each point sequence. Due to this property, the starting points and endpoints for every user are known and certain as the business can know the first time a user landed on a website and at a specific date if it converted or not. A collection of the possible states is the state set: $S = \{s1, ..., sn\}$. The first-order Markov assumption states that the information captured at time t is fully explained by the feature at time t-1 implying that observations before t-1 do not matter (Keilson, 2012). The transition probabilities are calculated with the following formula: $wij = P(Xt = sj \mid Xt{-}1 = si)$, $0 \le wij \le 1, \sum wij\ N\ j{=}1 = 1 \forall i$ where wij is the transition probability of hopping from state i to state j.

Figure 2. Example of a Path to conversion shown as a Markovian graph with its transition probabilities



A simplified example of four customer journeys is depicted as follows:

Customer journey 1 → TV -> DISPLAY -> CONVERSION

Customer journey 2 → SOCIAL -> VIDEO -> CONVERSION

Customer journey 3 → SOCIAL -> VIDEO 1 -> SOCIAL > SEARCH > CONVERSION

Customer journey 4 → SEARCH -> CONVERSION

A list of four customer journeys and their graphical representation is provided. The nodes represent the states, the arrows indicate the direction, and the probability of hopping from one state to the next is given. The transition probabilities are commonly presented in a transition matrix, representing a map of customers' paths. These transition probabilities express the sequential nature of the customer journey rather than an aggregated collection of touchpoints. The transition matrix helps discover rarely or frequently walked paths that drive conversions. The transition matrix allows for identifying structural correlations between touchpoints to construct an attribution model. More specifically, attribution is estimated as the change in probability to reach the conversion state from t = 0 when removing si from the matrix. Anderl et al. (2016) refer to this as the removal effect. The formula of the removal effect is given below: $A(xi\ ) = 1 - (\ Converion\ rate\ without\ touchpoint\ i)$. In other words, the removal effect provides the change in conversion if state i is wholly removed, enabling the performance of a counterfactual analysis for computing attribution. After estimating the removal effect, the touchpoints' contribution will be normalized to assist interpretation and comparison.

First-order Markov chains imply that the current state solely depends on the previous touchpoint and not on earlier touchpoints. An extension of the Markov chain is to relax the first-order assumption to higher-order assumptions. Anderl et al. (2016) adopt this approach and take the customer journey's t latest touchpoints into consideration. They estimate the first-, second-, third-, and fourth-order Markov chains. By relaxing this assumption, the state becomes a sequence of touchpoints. A generalization of the provided formulas for the Markov chain is applied. Higher-order Markov chains are generated to incorporate more extended temporal dynamics, which may lead to better performance.

Case Study

Dataset

The dataset was provided by a Direct to Consumer company (DTC) located in California, United States. With over 4 million active subscriptions, they are leaders in the Subscription Make-up business category. Log-level user data is generated from the website and mobile application with around 10MM records per day. It is also known as clickstream data in the digital ecosystem.

The analysis is conducted using the following information:

1) Clickstream events: Actions or events set up in the web/app
2) Page views: Page views of users that interacted with the web/app.
3) Marketing spend: Daily spend at a campaign level.
4) Conversions: Specific events based on conversion goals like signups or subscriptions

The data consists of 1.542.555 different paths (unique users) and a total of $3.968.670 spent in Digital Marketing in the period analyzed (30 days). It is worth mentioning that the datasets provided lack impression-level information which is usually restricted by large media corporations such as Google or Facebook. The majority of papers analyzing multi-touch attribution typically do not have this information critical in evaluating the Customer Journey. To get a better sense of the impression data, in the ads industry, a common Click Through Rate on Ads is about 2-4%. This means that without impression log-level data, MTA analysis loses around 90-98% of the information. Nevertheless, click-only information provides a good starting point for modeling this kind of problem. Data is extracted using Databricks, PySpark, SQL, Hive.

Data Preparation

The data needs to be transformed into the correct format for a Machine Learning problem. In this case, the starting point is the clickstream data generated in the back end of the application or website. It is essential to have a unique identifier for every visitor that lands on the website, the most common ID used is the SessionID.

Every SessionID contains a unique tracking id for a specific visitor. There are some properties about the sessionid that will not be discussed, but the most crucial element is that each one contains information such as the Source and Campaign where the visitor came from. It also adds more information like timestamp, IP, device information, and more. We can think of this as the log-level data of the server.

The problem with sessionid is that they are unique for each session, so if visitor A lands on the site on ten different occasions, we would have ten other session ids. We create a new tracking id called TID which will reflect a Visitor and group multiple sessions. This is the most critical cookie-based id we have to track people correctly.

Whenever this TID lands on the site and converts (signup, subscribe), the server receives the email and successfully creates the account. In this case, we can use the userid (id or email) to identify that user. This is important since the tracking id is cookie-based. Whenever the cookie expires, or the visitor clears navigation information in the browser, a new TID will be generated, which will look in our analysis as another visitor.

In terms of Data preparation, the most important thing is to clean up the Signup and Subscribe Events to obtain the Timestamp, the session, and the metadata (campaign, source) from the conversion event. Once we have that for each day for each user that converted, we can prepare the historical events that lead to that moment. Based on the clickstream data, we search for every TID that matches the one that converted and register all the sources and campaigns they touched before converting. By doing this, we can create the history for each converted as a timeline that should look something like this:

Table 1. Dataset used for the Attribution model dimensioned by userid

DT	User_id	Touch_ts	sources
2021-03-01	125200	[2020-11-04 07:22:53]	[Email]
2021-03-01	125200	[2020-11-04 07:23:50]	[Facebook]
2021-03-01	125200	[2020-11-04 08:22:53]	[Organic]
2021-03-01	125200	[2020-11-05 11:22:44]	[Conversion]
2021-03-01	233201	[2020-11-11 07:22:53]	[Facebook]
2021-03-01	233201	[2020-11-12 07:22:53]	[Facebook]
2021-03-01	233201	[2020-11-13 07:22:53]	[Google]

The table contains all the sources where the user came from, ordered in time. It is the sequence of all the touch sources before converting or not converting.

The idea is to prepare the dataset as a succession of events or chains in Markov, where the visitor transitions from state 0 to conversion. This means that all paths should be grouped in one cell using the tid. In short, one row per converting user. By doing this, some additional valuable metrics can be obtained, such as:

- Number of Paths: Count(number of sources/touches)
- Time-Lag of conversion: Converting Timestamp - First Landing Timestamp

These two should provide good insights for the business, understanding the whole cycle of conversions to analyze if it is short (low amount of days) or an extensive process like, for example, in the Car Industry or B2B. The Multi-Touch Attribution problem arises and expands whenever the number of paths increases (> 2), as the complexity of attribution and weighting also grows.

The exact process is repeated for Non-Converting Paths. Every day, we filter for the TID's that did not convert and are non-users of the site. Then we combine all their previous landings on the site to generate a complete view of the Non-Converting Paths. It is vital to feed the model the converting paths and add those that did not since it will help optimize the Customers' Journey.

It is essential to highlight that most E-Commerce businesses containing high impulse buying products tend to have shorter paths. If most of the conversion paths (> 80%) include only one path, meaning that the visitor landed once and bought instantly, it does not make sense to develop a multi-touch attribution model. These types of businesses usually have a low AOV (Average order value) with cheap products. There is a high positive correlation between the product's price and the number of paths. Thus, a one-touch heuristic model should work perfectly fine.

In the case of Google Analytics, which most companies use, there are some ways to generate individual user-level data. Nevertheless, a section contains all the paths that lead to conversion and paths that did not end in conversion at an aggregate level that could also be used to generate an MTA model.

Exploratory Data Analysis

The sample table generated in previous steps:

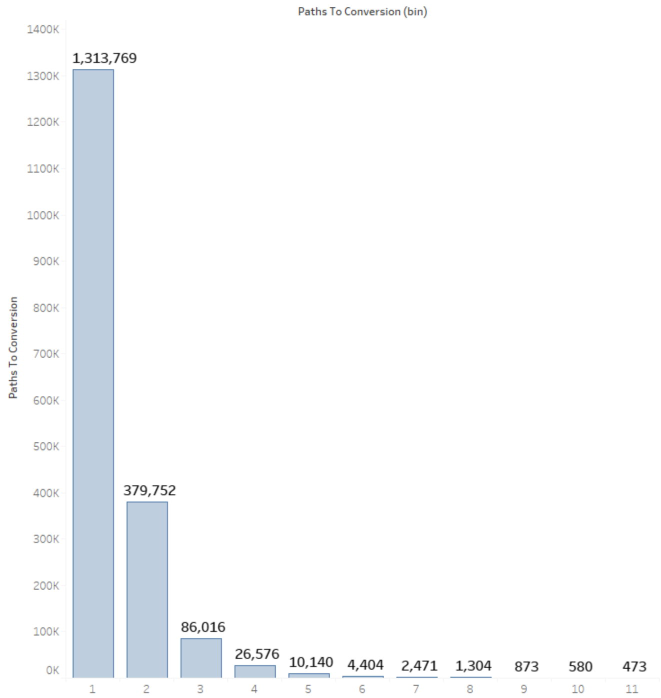
Table 2. Final dataset used for Markov model. Denotes each of the touchpoints for each user

Dt	Signup_ts	Userid	Touch_timestamps	Touch_sources	Paths_To_Conversion	Days_To_Convesion
11/1/2020	2020-11-01T10:59:18.200+0000	u-kgzhahfuj0tn118m	[2020-11-01 10:55:26.536]	[organic]	1	0.002685185
11/1/2020	2020-11-01T10:59:23.885+0000	u-kgzhalu44gv4118v	[2020-11-01 10:56:12.194]	[organic]	1	0.002210648
11/1/2020	2020-11-01T10:59:24.017+0000	u-kgzhalxq7zie10zd	[2020-11-01 10:51:30.265]	[Google]	1	0.005486111
11/1/2020	2020-11-01T10:59:26.035+0000	u-kgzhanhilcxr10f1	[2020-11-01 10:52:47.615]	[Google]	1	0.004618056
11/1/2020	2020-11-01T10:59:26.643+0000	u-kgzhanxj5f7n1128	[2020-11-01 10:52:52.637, 2020-11-01 10:59:30.064]	[Facebook, organic]	2	0.004560185

- Dt → The Date on which the conversion event occurred
- Signup_ts → Timestamp UTC from the conversion event
- Userid → Unique ID for each user that signups
- Touch_Timestamps → The different timestamps in which the user landed at the site
- Touch_Sources → Registering the source that the user interacted with before landing
- Paths_to_Conversion → Count the numbers of sources present
- Days_to_conversion → Date Difference in days between first and conversion landing.

In the dataset there are over 1.5 million users that signed up. The distribution of the number of paths for each signup look as follows:

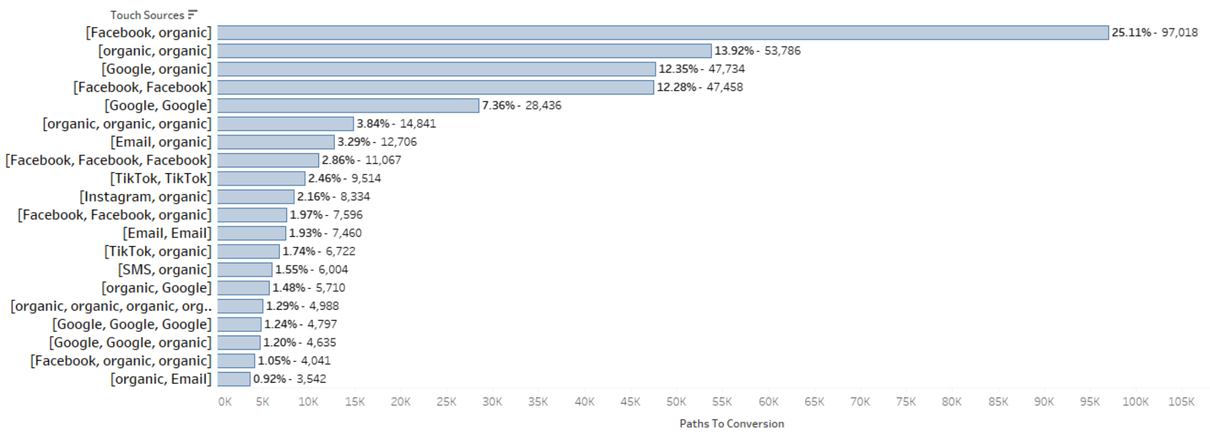
Figure 3. Number of paths at a user level. On the right, the distribution of users by source.



Sources	
Facebook	581,766
Google	424,320
Organic	352,386
TikTok	64,930
Instagram	28,496
SMS	23,061
Snapchat	17,673
Bing	12,172
Link	9,751
Pinterest	9,706
MSA	7,190
ImpactRadius	5,507
Twitter	2,476
Youtube	1,939
Geistm	835
TV	333
Social	14

As seen in previous research, around 70-80% of conversions account for only one path. In the case of this Beauty Retailer, the same rule applies. The biggest problem is that in the circumstances that just one touch is observed, there is no multi-touch attribution model that can be applied since all the Heuristics and models would give the same result. Thus, all research methodology will be used on the other 228,786 remaining conversions.

Figure 4. The distribution of occurrences at a path level for each of the multi-touch paths in the dataset.

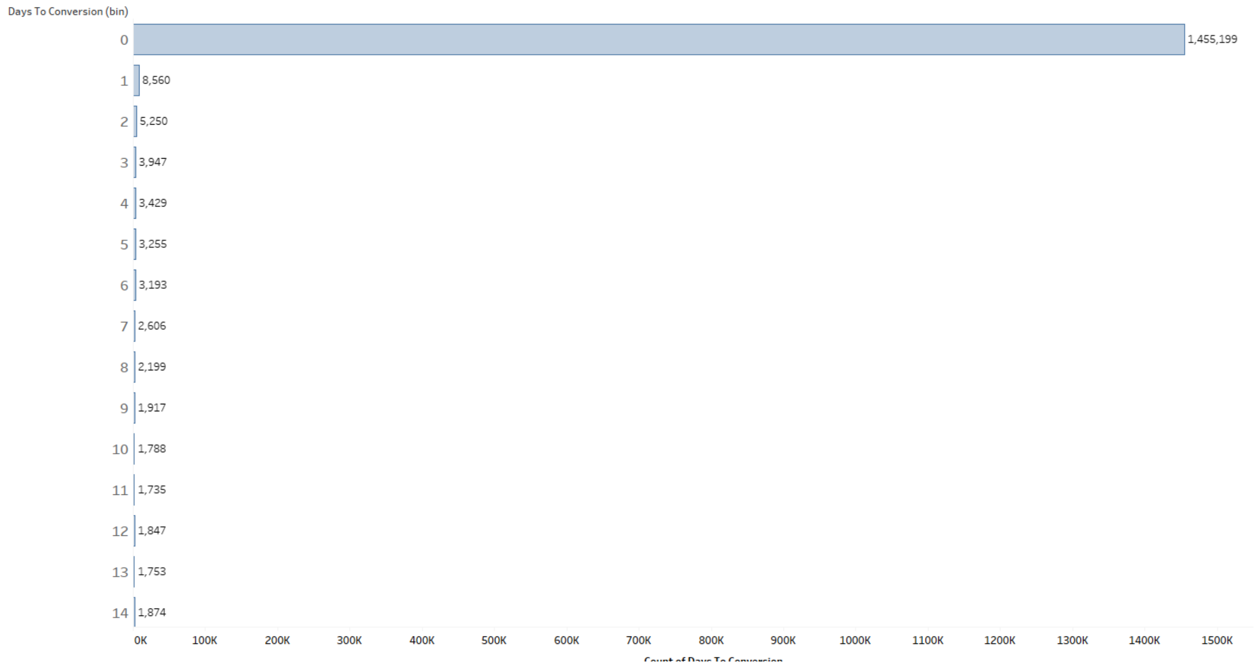


For paths containing two or more touchpoints, we can observe that there are around 380,000 paths that have two and 86,000 that have three. Nevertheless, approximately 25% of multi-touch paths contain the same source repeated, which would not count for the analysis and modeling. For the path [Organic, Organic, Organic], any model would not change the outcome since it is always the same.

The chart above shows the Ranking of sources by the number of appearances in the paths. Logically, Facebook and Google are the main drivers of conversions and are present in over 60% of the paths. It is worth mentioning that given the scale of both Facebook and Google, it would be interesting to open them up in smaller and smaller categories. A good idea for further research would be to analyze everything at the Campaign Level and not at Source Level. The company studied contains 20-30 Facebook campaigns running in parallel, while Google includes a similar amount.

Another basic graph that must be analyzed is the Days to Conversion for every conversion. In this case, 1.3M conversions have only one path, which means they converted in their first landing. Thus, 0 days passed from the point they landed to the fact they converted. We can observe that only ~ 60,000 conversions (4%) have delayed signups in the Distribution below. Beauty E-commerce companies generally follow the same trend; product-intent is impulsive and fast. In the analysis case, as we consider signups mainly, this makes sense as signing up does not require any monetary effort or compromise. In this sense, it does need to complete several steps and beauty quizzes that take between 5-10 minutes approximately.

Figure 5. The number of days that pass from the landing timestamp to the conversion timestamp aggregated at a user level.



The chart denotes how many days passed between the first Landing Timestamp and their final timestamp. By counting how many conversions occur (Y-axis) and how many days have passed since Landing (X-Axis), we can get a better sense of how much time customers usually take to convert. The marketing team needs to push hard into the signup conversion within the same day as the probability of converting after 24 hours drops substantially. In practical terms, it would not make much sense to push remarketing campaigns in the posterior 24-72hours if the user did not convert on the day of landing on the site. Other companies and industries present completely different Customers Journeys and Days to conversion.

In terms of analyzing the channels, Organic seems to be one of the worst-performing paths. Nevertheless conversion rate increases substantially when combined organic with other organic or paid media touches. For example, [Organic, Organic] increases +24.74% and [Organic, Organic, Organic] +14.75% the conversion rate. Second, paid sources like Facebook seem to be sending highly relevant traffic that converts with a higher probability than other channels.

Any paid channel, finished by a new touch point coming from Organic, is of great value. All marketers, in general, try to minimize the attribution into organic as much as possible since it is not a channel they can put money in. Nevertheless, this point can be discussed since there is a lot of work in SEO, Branding, and Internet Presence that directly affects this channel. As seen in the industry, many companies set custom rules into their attribution pipelines, stating that whenever the converting touchpoint ends as organic, they choose another paid channel. In this way, they can enhance and increase the value of their marketing spend on other channels. In the last part of the paper, we will explore some of these custom business decisions for attribution modeling.

Model

As discussed, the Markov Chains Model was selected due to its robustness, adaptability, and interpretability. The library used is called ChannelAttribution by David Altomare and David Loris, available in Python and R. The input of the model:

Table 3. Markov's output dataset

	touch_sources	total_conversions	total_conversion_value	total_null	total	conversion_rate
0	[Facebook]	494261	494261	4689689.0	5183950.0	0.095344
1	[Google]	372178	372178	2127847.0	2500025.0	0.148870
2	[organic]	220471	220471	7068110.0	7288581.0	0.030249
3	[TikTok]	54632	54632	349598.0	404230.0	0.135151
4	[Email]	52737	52737	281988.0	334725.0	0.157553
...	...	...	...	...	...	...
2567	[organic, FP, FP]	1	1	25.0	26.0	0.038462
2568	[TikTok, organic, TikTok, TikTok, TikTok, TikT...	1	1	NaN	NaN	NaN
2569	[SMS, SMS, organic, organic, SMS, organic]	1	1	2.0	3.0	0.333333
2570	[organic, organic, Google, Google, organic, or...	1	1	NaN	NaN	NaN
2571	[organic, organic, Other, organic, Google, Goo...	1	1	1.0	2.0	0.500000

Some transformations need to be done to the data to look like this. We need to group by every existing path and aggregate all the metrics. It is essential to use the conversions generated for the algorithm to work, and all the paths observed where there was no conversion. This gives the model way richer data as we can calculate the Conversion Rate for each path at an aggregate level, which states the path's efficiency. This is of great importance; if we only observe nominal conversions, we might state that some paths are of higher value simply because they provide higher traffic, but it can be highly inefficient to convert visitors.

- Convert NaN to 0.
- Remove '[' and ']' from the paths
- Change ',' into '>' in touch_sources

In this dataset, as each conversion is a Signup event, we don't have a Value associated with it (it could be calculated, though, but it would be a different study). Different from an Ecommerce where there would be 1 Purchase Event and a total value of the order, for example, $50.99. The model can also take into account which of the paths generate the higher value conversions. In this case, we do not use the value for the algorithm.

The model output:

Table 4. Aggregating the Markov output at a channel level

	channel_name	total_conversions
0	Facebook	529996.703898
1	Google	398303.743005
2	organic	352849.480235
3	TikTok	64200.557668
4	Email	65004.380420
5	Instagram	25980.699667
6	SMS	18105.150560
7	Snapchat	14028.620889
8	FP	16487.935737
9	Bing	13464.031098
10	Pinterest	8612.386630
11	Link	10382.710549
12	MSA	7703.301375
13	ImpactRadius	6918.617260
14	Twitter	2535.869397
15	Youtube	1961.710288
16	Other	2516.730760
17	Organic	2047.834154
18	Geistm	1349.273905
19	TV	105.262503
20	Social	0.000000

The model calculated the transition probabilities for each channel to move into the next one, and implicitly the last channel in every path is the "conversion." For each of the 20 channels in the dataset, the model outputs the final Conversions in numbers.
Based on the probabilities, the model can calculate accurately for every conversion that contains multiple paths, the contribution that each "player"/path made to acquire that conversion. Logically, it considers all of that channel's participation in all the dataset with simple probabilities.

The model then starts calculating what is denominated as the "removal effects" to calculate the channel's importance. We can simulate what occurs after removing a channel from the system and observing what happens to the conversions given the system's probabilities (transition matrix). Therefore, after n simulations, we reach convergence, which in our case is < 5%. The convergence number is a variable that the algorithm uses to stop the iterative process once the percentage of variation of the results over different repetitions is less than convergence parameter, as defined by the authors of the package. The model uses simulations to reach Markov's convergence. There are several papers showing algorithms to reach a Markov's convergence depending if the stationary matrix has finite convergence time or not.

Figure 6. Console information when running the Markov model in R, using ChannelAttribution library.

```
Suggested order: 2
Number of simulations: 100000 - Reaching convergence (wait...): 11.04% > 5.00%
Number of simulations: 150000 - Reaching convergence (wait...): 7.93% > 5.00%
Number of simulations: 225000 - Reaching convergence (wait...): 6.35% > 5.00%
Number of simulations: 337500 - Reaching convergence (wait...): 5.51% > 5.00%
Number of simulations: 506250 - Convergence reached: 4.36% < 5.00%
Percentage of simulated paths that successfully end before maximum number of steps (24) is reached: 100.00%
```

The definition of the removal effect is the percentage of conversions that would disappear if a channel or player was removed from the system. In other words, creating a new model where the specific channel is set to 0 conversions highlights the effect of removing that channel on the overall system. Mathematically speaking:

$$Removal\ effect\ of\ channel\ k = \frac{p(conversion\ in\ absence\ of\ channel\ k)}{p(conversion\ in\ presence\ of\ channel\ k)},$$

where k = 1…N, N is the number of channels.

$$p(conversion) = \sum_{n=1}^{N} \prod p_{ij},$$

$$Credits(k) = Removal\ effect\ of\ channel\ k * total\ conversions,$$

The Removal Effect is a way to measure the contribution of individual channels in generating conversions. This is done by completely removing the channel from the path. The larger the impact, the higher the value attributed to the channel. We can do this iteratively for each channel to see the impact on conversions and ultimately quantify the value of each channel.

First of all we calculate the probability of that path to conversion by counting how many equal paths ended in conversion or not conversion. Then we calculate the transition probabilities from channels like this:

Table 5. Example of the Markovian transition probabilities to calculate the removal effect.

Path	Probability
Start > Email > Conversion (50% x 70%)	0.35
Start > Email > Display > Social > Conversion(70% x 50% x 50% x 90%)	0.1575
Start > Display > Social > Conversion (30% x 50% x 90%)	0.135
All Paths	0.6425

By removing Email, the only converting path remaining is Start > Display > Social > Conversion with 13.5% conversion probability. The Removal Effect formula is:
1 – (.135 / .6425) = 0.79.

By using this formula, it can effectively calculate the overall attribution of each channel in the system. The transition matrix states the probabilities of moving into the next step. For simplicity, we are only considering order one that just looks at the step before. Some papers try to optimize the best order to work with, increasing the model's complexity and providing higher accuracy.

Table 6. Transition matrix output by ChannelAttribution library in R.

```
markov_common['transition_matrix'].tail(20)
```

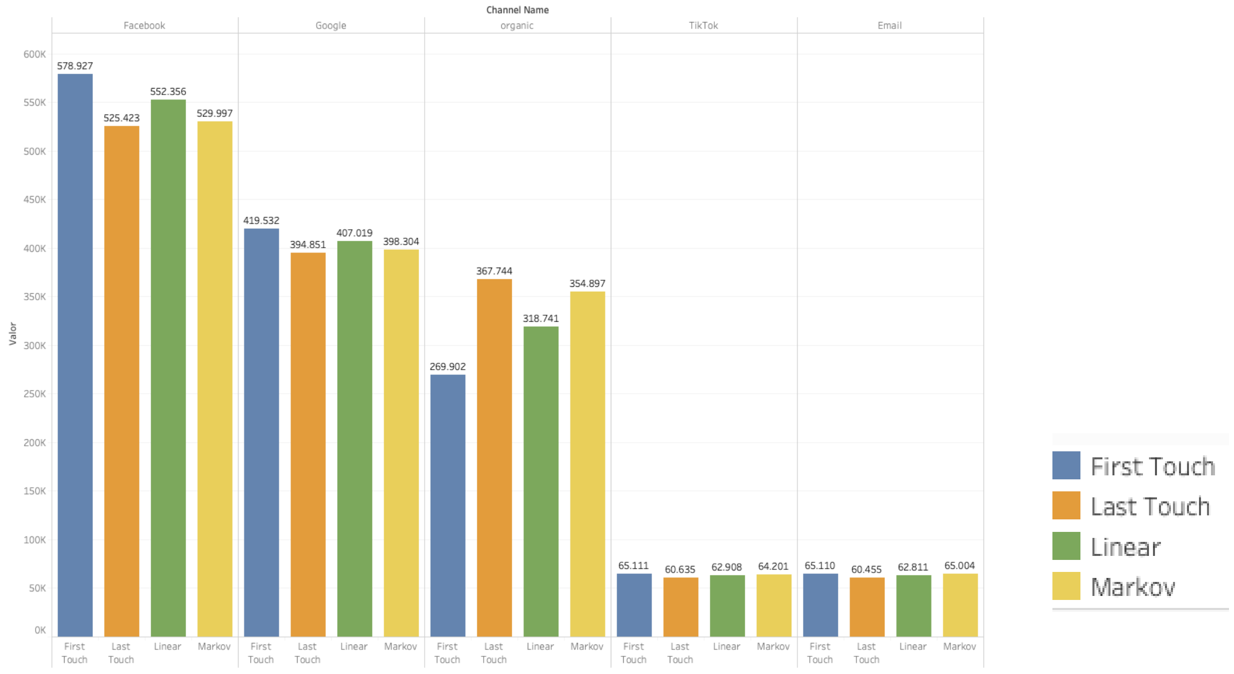
	channel_from	channel_to	transition_probability
1406	15 10	(conversion)	0.500000
1407	15 10	(null)	0.500000
1408	11 13	(conversion)	0.500000
1409	11 13	(null)	0.500000
1410	15 5	(conversion)	0.250000
1411	15 5	(null)	0.750000
1412	14 9	(conversion)	1.000000
1413	17 12	(conversion)	0.333333
1414	17 12	(null)	0.666667
1415	9 17	(conversion)	1.000000
1416	17 10	(conversion)	0.100000

(Note: The transition matrix is frequently visualized as a graph network)
Each channel is represented by a number in the model. Channel_from column represents if the transition from Channel i to Channel j ended in Conversion or Null. The model assigns the transition probability based on all the occurrences of the dataset for that path.

Results

Even though the dataset analyzed had only 20% of the full paths with two or more channels involved, there are key differences in the results.

Figure 7. Number of signups per channel dimensioned by the different attribution algorithms.



There are substantial differences between the heuristics and the algorithmic model for the top 5 channels representing almost all the dataset conversions (Markov).

Most companies usually use the last click heuristic approach. Facebook and Google are over-stated in first touch approaches, while Organic gets more conversions in the Last Click. A possible explanation is that whenever a user does not know the brand, the first impact is usually in Social Media or a Search Engine. After navigating the site, he or she exits and starts receiving retargeting ads. This dataset does not contemplate impression pixel information. The user then returns to the site by searching in Google, clicking on the Organic link into the site, and then signing up. Overall, there are more last-click sign-ups for organic than the first click. How can someone search for what is unknown at that moment?

The Markov Model understands the importance of the user searching for the brand and finding it quickly in the search results, giving almost the same significance as the last click. In this case, the model gives an output closer to Last Touch than any of the other heuristics. One of our hypotheses is that in this particular dataset more than 80% of the conversions have only one touch point so that could contribute to the similarity. Moreover, in the Paths not to conversion data, there is a high number of paths that do not end in conversion which may induce the model to give a low score to starting touch points (First-touch model). Another reason could be that the first touch points generally are high-awareness ads and not targeted ads for high conversion. In other words, the efficiency of the first touch points is lower than the last touch points that the marketers create to finish the conversion. We think that this is generally the case for Direct to Consumer companies, especially in the Digital ecosystem.

For Facebook, we see the largest disparity between the model and first-touch; as Facebook is present in many of the paths it underestimates 10-15% of the conversions vs. the first touch. This makes sense since Facebook tends to over-attribute many conversions, stating that it is the cheapest channel to acquire new customers. Furthermore, it is the biggest channel for the company for awareness ads. Thus, it sounds reasonable that most of the leads are starting their path to conversion from Facebook.

The idea is that after comparing the different attribution methods, the company can decide on the one that aligns better with their business. Several papers have discussed that algorithmic approaches outperform heuristic ones in almost all performance metrics.
As an example, if the company decides to go with the Final Output from the Markov Model, then they can re-calculate their True Cost Per Signup (CPS $) with a simple formula:

CPS = $ Money Spent in Channel X / Markov Conversions from Channel X
= $2.000.000 / 529.997
= $3,77

The first click CPS for Facebook is $3,46. This number is crucial for marketing investment. As observed, the true CPS changes according to the attribution approach chosen by the company. If $3,50 is the profit threshold for acquiring a customer, paying more than $3,50 will make the company lose money in the long run. In this example, if the company invests in Facebook thinking that they are at a $3,46 CPS, which is profitable, it could mean bankruptcy since the attribution logic is mistaken or simplified.

Observing this in a simplified example looks straightforward, but when results are extrapolated to n > 20 channels, which affect each other, the problem starts to grow exponentially. Also, the Money and Resources to be spent are limited, so companies that maximize their efforts in the correct channels can expect to maximize ROI. Understanding how channels interact is of great importance; in many cases, Display Ads or ads focused on

Awareness and Brand generation are often understated in Heuristic models. In practical terms, there is no incentive for companies to invest in Display Ads because they observe low ROI and High CPS. But, if improved models were to be used, generally, display ads would be valued higher as they are present in the conversion paths because they are cheap to show and work in mass media channels.
Another result to be observed is the overall importance of middle channels that contribute to conversions. In the dataset analyzed, mixing email with social media has the highest conversion rate. This is a great insight to be tested; the more emails recollected means that these people could be targeted through social media with offers. By mixing up these channels in the customers' journey, the company could decrease their CPS $ and increase their signups. Some strategies to be explored could be: Popup for Email Leads, Creating new targeted campaigns based on email users (Facebook, google), Landing Optimized Sites for email recollection. Some other channels mixing observed in the data set that could be explored are:

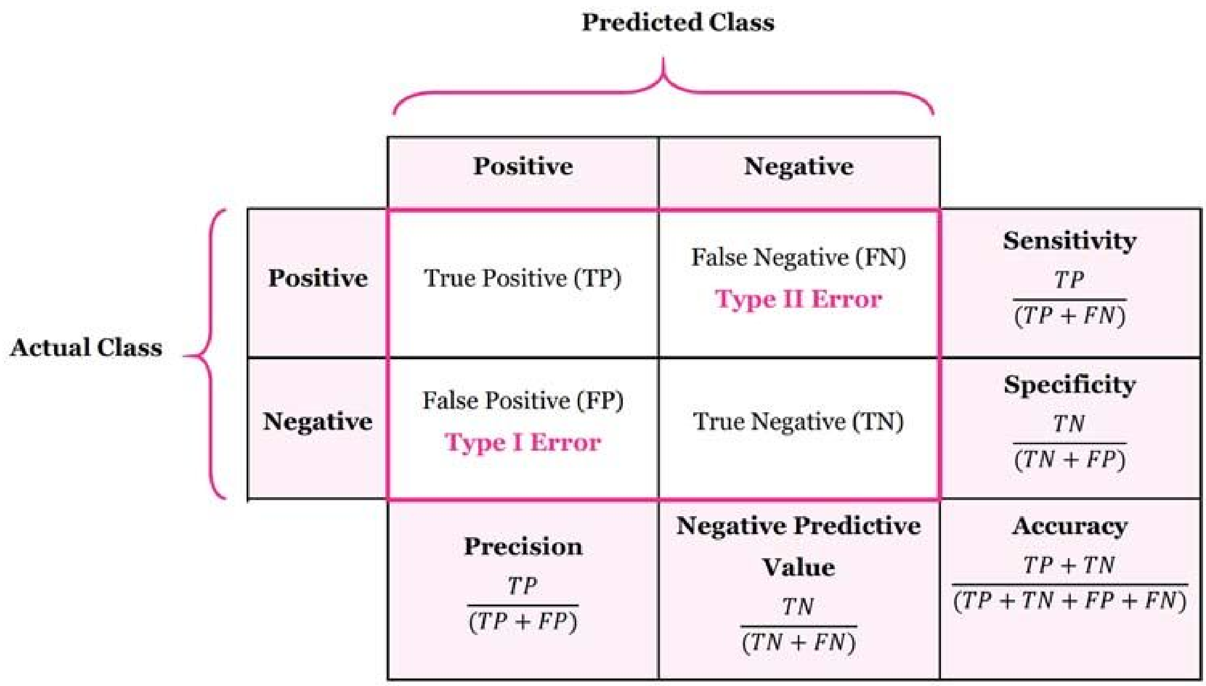
- SMS, Facebook
- Instagram, Facebook,
- Email, SMS
- Email, Organic

The organic channel cannot receive money since it is not a paid media channel. Nevertheless, it is an area called SEO that can be worked on and improved. Some further analysis could explore the improvement in keyword positioning in search engines and their contribution to increased conversions with paid media channels.

Predictive Accuracy (ROC)

The markov model can be evaluated in terms of its predictive performance. The output is the probability of each of the paths to actually convert or not (transition probabilities), so by setting different thresholds it's easy to observe how many paths are considered as conversions or not. Therefore, it can be analyzed how many paths are stated by the model as converted and compared against the real data. This leads to the Classification Matrix:

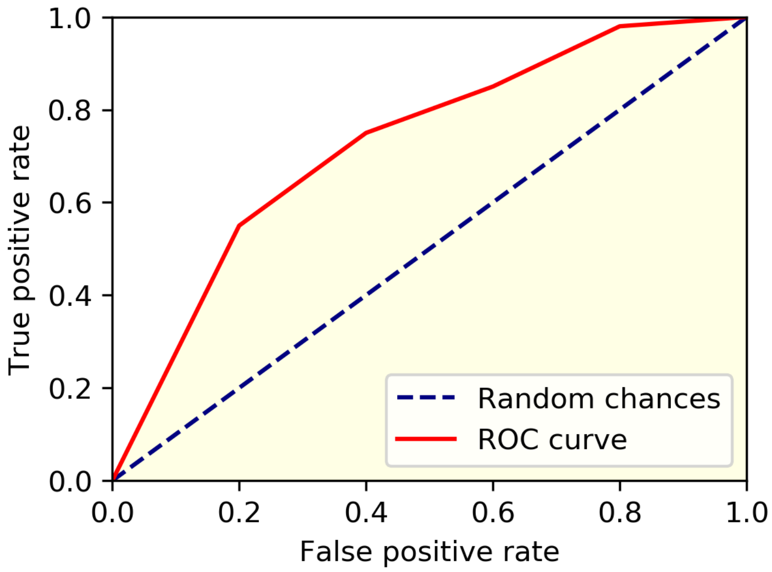
Figure 8. Example of a Classification matrix for machine learning.



With the above matrix we can then calculate a series of KPIs to actually measure correctly the performance of the model. The most frequent approach is to use the Receiving Operating Characteristics (ROC) within this framework. Ideally, we want to use this to compare between models and to detect if the model is working accordingly or not. ROC is a commonly applied metric in Machine learning and Data mining. The curve of the ROC graph can be used to visualize, organize and select different classifiers based on their performance.
In our case the four possible scenarios:

- Markov Model classifies as Conversion but in the dataset it's No Conversion. (Error -> False Positive)
- Markov Model classifies as No Conversion but the dataset its a Conversion (Error -> False Negative)
- Markov Model classifies as Conversion and its a Conversion (True Positive)
- Markov Model classifies as No Conversion and its a No Conversion (True Negative)

Figure 9. Example of a ROC Curve visualized for measuring the model.



The optimal point for the modal is in the point (0,1) based on the chart where the Probability of False alarm is 0 meaning that there are no mistakes. As well as where the True Detection is perfect as well, where 100% of the cases are correctly classified. Different models will present different ROC curves, the ones above logically are better. Once obtained the input for the ROC curve, the Area under the curve can be easily calculated and compared against different lines. As observed in the graph, if the area is closer to 1 then it tends to a better performance.

Table 7. R's output cell showing the AUC for each Markovian order hyperparameter

```
model[0]
```

	order	auc	pauc
0	1.0	0.708605	0.708566
1	2.0	0.708605	0.649099

The Markov model has been calculated for Order 1 and Order 2 but as observed the performance in terms of AUC it's almost the same. The Partial AUC (pAuc) considers only those regions of the ROC space where data have been observed.

Table 8. R's output cell showing the Classification Matrix input

	fpr	tpr
0	0.00000	0.00000
1	0.00000	0.00000
2	0.00000	0.00820
3	0.00000	0.00820
4	0.00000	0.00820
...	...	...
96	0.50317	0.82812
97	0.54809	0.84605
98	0.95275	0.99243
99	0.99300	0.99953
100	1.00000	1.00000

101 rows × 2 columns

To graph the ROC curve we need to define different thresholds and evaluate the true positives and false positives rates. The threshold cut defines if the probability of the model of classifying the event is positive or negative. As imagined, when the threshold is at 100% the model has a FPR of 1 and TPR of 1. For 100 for example, it considers everything as positive which means that it correctly classifies all real positive events, and also defines all the negative as positives. On the other hand, on 0, the model does not classify as positive any register which means that it misses all positive events and also has no mistakes (false positives).

ROC and AUC are the correct metrics to evaluate the model, in the case of this thesis the most important thing is that the model is actually better than heuristics. Predicting the probability of conversions for each path is not enough, the real application of the Attribution model is to define the real allocation of credit.

One of the classic ways of comparing the Data-Driven model performance to the traditional heuristics is to analyze the r2 of the model. Generally, the Logistic Regression models used for Attribution are easier to compare to heuristics, given that as a supervised problem, it is easier to compare with the traditional methods, and evaluate which one has more predictive power. With Markov, using the classification matrix shown above, we could compare the results to the other models. In practice, one possibility could be creating several Logistic Regressions for each of the heuristics using only the paths used for that heuristic as explanatory variables. An example:

1) Logistic Regression using only 1 column: First touch path.
2) Logistic Regression using only 1 column: Last touch path.
3) Markov model

Once obtained the classification matrix for each and the ROC Curve, the KPI for evaluation could be the AUC. Several papers have already shown that Data-Driven models (Markov and Shapley Values) outperform traditional heuristics.

Applications

Most literature studying multi-touch attribution topics usually ends their analysis by stating the model's improvement using different machine learning methods like Shapley values, Markov, bagging. What is not clear is the applications of these kinds of models and how they can improve Paid Media's performance, which is the golden question. Other papers discuss the best evaluation methods for these models and lack their applications in actual Marketing departments, which are the customers of these models. The real question relies on if it is worth it for a Marketing Director to invest in a Multi-touch attribution model and how it could be used in a real-life business scenario.

The main problem with all existing models is the output they generate. Most frequently, as observed in previous chapters, the outcome is at an aggregated level for each channel stating the final number of conversions in a specific period. This means that when the business gets a new Conversion, they see:

John Doe Signup Event 30/01/2020

Paths: Email > Organic > Facebook > Instagram

In the heuristic world, this can be answered quickly as follows:

- First Click: Email 100%
- Last Click: Instagram 100%
- Linear: 25% to each
- Markov:?

Markov gives us an aggregate for each channel for the period (weekly, monthly, quarterly). How can we translate the aggregate view of Markov into the individual level, which is the most helpful view for businesses? Or how can we create a decision framework using an algorithmic model so that marketers can optimize their spending in real-time performance based on the model's output?

Another exciting application is how marketers can be sure that the data-driven model is outperforming the classic model. Can we think of some sort of experimentation or A/B testing to validate and test the model? Moreover, the channels or rules that the model learned change due to the market, new channels, new ways to advertise.

From Aggregated to Individual Level

Conversions flow every day into the business with all the necessary information like Username, email, Timestamp. Generally speaking, every hour-day-week-month, a sales report is generated, which states how many users were acquired. Depending on the business, this could be Signups, Leads, Purchases, Conversions, log in.

On the other side, Marketers are injecting money into different channels where they could potentially find users interested in their products/services. They work under certainty because the platform (FB, Google) states the exact amount of money spent and what they think are the conversions (through pixel information, cookies).

What most papers have been working on is the aggregated attribution solution. There are 20 channels; they output the number of conversions for each based on which should be credited for. This means that it does not matter if it was Joe, John, or Lucas who signed up, neither if this proved to be a loyal customer or not.

In practice, every Customer has a customer_id and a real value for the business. Every customer has unique attributes like age, economic data, marital, gender, location. Based on the aggregated data-driven model, if we wanted to know the accurate attribution for people in Georgia or Florida, we would not be able to do it. Other times, the business is running influencer campaigns with referral payments. In this case, every new conversion brought by those campaigns needs to be acknowledged, and in what %. If influencer @beauty_channel got 50 new signups, we need to know who they are and measure those 50 signups. According to what attribution model? Maybe they brought 50 and interacted in another 200 paths to conversion that would increase the value.

Given this problem, a new algorithm needs to be created and worked on. The Markov model output gives an aggregated view for each of the 20 channels, but the real output required is an individual view for each conversion opened up by how many paths the user interacted before converting. For Example:

- John Signups with the following path: Instagram > Organic > Facebook > SMS
- In the database, the output looks as follows:

Name	Email	Signup_date	Paths	Attribution
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	?

In a Heuristic Last Model, the dataset should look:

Name	Email	Signup_date	Paths	Attribution
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	SMS

Nevertheless, for a DataDriven model that contains weights, each of the four paths should have its %. Some data modeling can solve this problem:

Name	Email	Signup_date	Paths	Attr Channel	Attr Weight
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	Instagram	0.1
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	Organic	0.5
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	Facebook	0.3
John	John@gmail.com	01/01/2020	Instagram > Organic > Facebook > SMS	SMS	0.1

This example looks like the U-Shaped or Linear Heuristic model; the only difference is that the weights are not fixed but come from the algorithmic distribution.

Table 9. ChannelAttribution's output of Markov Attribution at an individual level

```
In [71]: markov['path_attribution'][19:30]
```

Out[71]:

	idpath	path	channel	total_conversions_weight	total_conversions_attribution	total_conversion_value_weight	total_conversion_value_attribution
19	17	Email > organic	organic	0.469087	2980.108976	0.469087	2980.108976
20	17	Email > organic	Email	0.530913	3372.891024	0.530913	3372.891024
21	18	Link	Link	1.000000	6147.000000	1.000000	6147.000000
22	19	organic > organic > organic	organic	1.000000	4947.000000	1.000000	4947.000000
23	20	MSA	MSA	1.000000	4834.000000	1.000000	4834.000000
24	21	TikTok > TikTok	TikTok	1.000000	4757.000000	1.000000	4757.000000
25	22	Instagram > organic	Instagram	0.521230	2171.965121	0.521230	2171.965121
26	22	Instagram > organic	organic	0.478770	1995.034879	0.478770	1995.034879
27	23	Email > Email	Email	1.000000	3730.000000	1.000000	3730.000000

In our data set, every idpath is one of the N paths that lead to a conversion. We had only the 20 channels with the data-driven attribution; thus, we have over +6000 different paths, each with their weights. The ChannelAttribution has a function called Markov_local that gives the table as output. Still, the idea behind it is that the Markov Model provides an underlying distribution at an aggregate level where each channel has its weight and total conversions.
For Example, Facebook accounted for 530.000 conversions in the dataset. This means that in the path level table, the sum of all rows that contain Facebook should aim to 530.000. Based on the overall distribution, an algorithm can be created that iterates N rounds for each of the paths to reach the model numbers' convergence. A simple algorithm could be:
For each pathid:

1. Open up the pathid into multiple rows for each channel involved (e.g., three paths, three rows).
2. Calculate each channel's importance based on the aggregated view of the model. (e.g., Facebook is present at 530.000 of 1.500.000 (33%).
3. Use the probabilities to define the path-level attribution.
   (e.g.: Path 1 - Organic > Facebook)
   Organic (15%) → 15/48 = 31%
   Facebook (33%) → 33/48 = 69%
4. Based on the weights, calculate each channel's sum and compare it with the model's aggregated view.
5. Define a % of error (1-2%) and repeat 1-4 until the desired % is achieved.

The real model provided by the ChannelAttribution library is more complex, but the underlying logic applies. It needs to translate the overall aggregated view into each of the paths observed. It is important to mention that if new data (conversions) arrive and paths are not attended before, the model needs to be re-trained. Once the attribution is represented at the path level, it becomes a simple task to attribute each of the conversions. Whenever a customer converts from a specific path journey, the path level table can be mirrored with those numbers and replicated in the customer table.

1. John Doe Converts with a Path → Organic > Facebook > Instagram
2. Look at that path in the table and assign the corresponding weight values.

As mentioned, the model should be re-trained every time a new channel is used or frequently (daily, weekly), depending on the amount of data, to adapt the weights to the dataset.

Evaluating Conversion Value

The current dataset is based on a subscription model business, specifically analyzing signups that have no value at first hand. The Markov model can be adapted to consider each of the paths/channels' Conversion Value to represent a more accurate reality view.
It makes sense; 10 conversions that sum up $50 are not the same as one accounting for $100. In Attribution logic, it makes sense to give a higher credit to those channels that bring more valuable customers that provide more profit. The formula would not be complete unless we also involve the Spend for each Marketing Channel. As a thought experiment, maybe the channel that brings high-value customers is costly, bringing profit down.
A plausible solution for this is adding a couple of variables into the model:

- Cost per Signup: It can be based on the UI data or the aggregated Markov model. (e.g: Facebook CPS $20)
- Conversion Value: In signups, this is more difficult to calculate because it is free and has no monetary value at first. Nevertheless, Signups do convert into Paying Subscriptions, and the ratio can be calculated for each channel. Doing some simple math, if 25% of signups convert into a Subscription of $20, then we can establish that the expected value of a Signup is around $5. (This should vary per channel, but it can be an approximation).

Having only a single number for each signup makes calculation easier since every signup has a $5 value. This is different from an E-commerce scenario where each purchase has several items and a total value. Based on each channel's subscription/signup ratio, we add this variable to enhance the model further and consider the signups' future value. Use case: Suppose Tik Tok is bringing lots of new signups, but the actual Ratio to subscription is low. In that case, the Markov model could evaluate it, assigning less attribution credit to the channel. This means that marketers would know precisely how much money to invest in the channel based on the actual monetary performance and not in the single signup attribution measure.

Table 10. Example of a CPS calculation using UI data.

Channel	Conversions UI	Platform Spend	CPS	Signup To Sub Conversion Rate
Facebook	5153	$10.000,00	$1,94	0.8
Google	2022	$5.000,00	$2,47	0.65
Pinterest	288	$800,00	$2,78	0.6
Tik Tok	268	$600,00	$2,24	0.7
Bing	112	$500,00	$4,46	0.85

Based on the two columns added, CPS and the Signup to Sub Ratio, the model can ingest these new variables to weight the total contribution of the channel to conversions. Conversion value should be lower when the Ratio is lower, meaning that less signups will actually convert into paid subscriptions. The relevancy or the quality of the users acquired by that channel should then be decreased. The same happens if the CPS is extremely cheap, generating lots of signups or conversions, but the quality of the users is not adequate.

To sum up, adding new features to the model to evaluate more accordingly the Conversion Value could bring better results and higher efficiency to the model.

Impression Level Data

Every analysis of Data-Driven Attribution models for Marketing avoids some crucial concepts about marketing campaigns. It is necessary to add new features to the model based on marketing expertise so that accurate attribution can be contemplated or as close to reality as possible.

Not only is every channel different, but every channel serves a purpose in the Marketing Journey. Some are focused on awareness, some on intent, some on product discovery, and some are the final "punch" to conversion. Some would say that the model should be able to learn all the relationships on its own. The truth is that the Customer's Journey is a complex problem and how the channels interact with each other to generate conversions is also hard to understand.

As mentioned before, without Impression level data, the model is half-blinded. Some companies are starting to add into the model impression data and offline data. Given localization in mobile phones, companies can predict whether a specific user views a street ad or walks in front of the business front. In digital marketing, impressions are vast, so adding each channel's impressions into the customer's journey could make the model super robust. This would also help awareness channels that do not account for many clicks, assigning them a higher contribution to sales. Nowadays, marketers are data-savvy, investing only in channels that have high ROI's. The problem with this is that by observing the incorrect metrics or biased heuristic models, channels that help but have low numbers of clicks are not provided with budgets. This is the case of Display Ads or Youtube Video ads that generate plenty of views but do not have high-value metrics. This is because in the Last click world, generally, the click-based channel gets all the credit. Companies that are starting to understand the actual value of ads using impression data are seeing increased results.

Figure 10. Example Google Ads display network types of ads

Some channels are starting to provide the impression level data information to their customers. The real challenge is to create an algorithm that can proficiently link each customer to its own unique id so that they can be tracked through the internet. The new markov model with the impression level data should look like this:

Table 11. Example of the Attribution input dataset when adding impression information

DT	Touch_ts	sources
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Email View, Facebook View, Facebook Click, Organic]
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Email View, Instagram View, Facebook Click, Direct]
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Organic, Facebook Click, Email Click]
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Facebook Click]
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Email Click]
2021-03-01	[2020-11-04 07:22:53.668, 2020-11-04 07:28:16.14	[Google View, Organic]

Every touch source can be opened up into a Click or a View (impression) which would expand the number of sources/paths for every customer.

Campaign Level Information

Up to now, we have discussed only the Channels that were involved in the dataset. The 20 channels are an easy framework to start playing with Data-Driven Models. In real-life scenarios, having Facebook as a big channel does not make much sense. Why? Facebook has over 10-15 different ads with different placements (mobile, desktop) and lots of different creatives to be tested. Aggregating all the different ads into a "Facebook" label makes a considerable loss of rich information.

Marketers optimize budgets not at a channel level but at a Campaign/Ad level. In our dataset, the numbers of Campaigns would translate into 100-200 different campaigns at any given moment. There were around 6000 different possible paths; for the 200 campaigns, we could be talking of 60.000-100.000 different possibilities. The model would be more challenging to run daily and extract value from cost and time.

Nevertheless, research should be applied to campaign-level attribution, not only for better results but also for a better understanding of how each campaign interacts with others. This would help marketing teams optimize their bidding in the Channels based on the data-driven model running daily. Even more, the process could be automated using Channels' API's contemplating in real-time the spend being used in each campaign and the real conversions in the dataset. If the ROI is positive, then the bot could add more budget to scale it up.

Figure 11. Facebook UI adset information

Nombre del anuncio	Entrega ↑
GBPlus_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-CelebMarchCreator-KhloeVID_13-65_F_Bid-BidCap-Purch_GBP-VID-0321-KhloeVF2-US-9x16-59s-CR-000001-8_Home_WL_3_18_25Price	Activa
GBPlus_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-CelebMarchCreator-KhloeVID_13-65_F_Bid-BidCap-Purch_GBP-VID-0321-KhloeVC2-US-4x5-91s-CR-000001-3_Home_WL_3_18_25Price	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_100_263...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_99_262_...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_100_262...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_98_262_...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_ 98_263...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-iSeriesMarchSaleCopyTest_13-65_F_Auto-Autobid-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Home_WL_ 99_263...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMinimalExclusions-MarchBestPeformers2-Creator_13-65_F_Bid-CostCap-Purch_GB-VID-0221-AngelicaHaulVA1-US-9x16-29s-CR-000011-1...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMinimalExclusions-MarchBestPeformers2_13-65_F_Bid-CostCap-Purch_GB-GIF-1020-iSeriesGIFVB-US-4x5-08s-12Price-IH-000007-1_Hom...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_tsBroadMinimalExclusions-MarchBestPeformers2_13-65_F_Bid-CostCap-Purch_GB-VID-0221-A649v2BVideoTSJ615v2BxSTRx-Has12Price-US-9...	Activa
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMinimalExclusions-MarchBestPeformers2_13-65_F_Bid-CostCap-Purch_GB-IMG-0221-IGStyleQuizVB-US-1x1-00s-NoPrice-IH-000002-16_H...	Activa
GBPlus_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMinimalExclusions-MarchBestPeformers2-Creator_13-65_F_Bid-CostCap-Purch_GBP-VID-0321-EstherHaulVA2-US-4x5-64s-CR-000019...	Activa

Resultados de 9.943 anuncios

The above table shows an example of nine thousand ads running in parallel on Facebook at a given date. In the internal database, marketers could see something like this:

Table 12. Dataset containing campaign information from internal Measuring tool

```
SELECT distinct payload.campaign FROM events.raw_event_cs WHERE dt = '2021-02-15'
```

(3) Spark Jobs

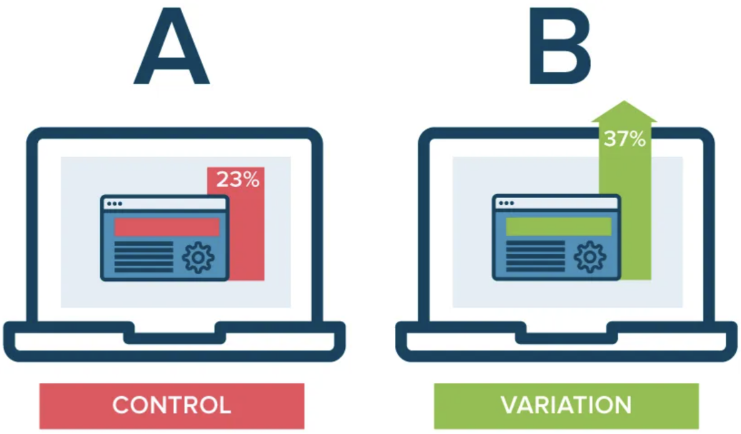
	campaign
1	ad_476224537862_keyword_monthly products_000717_20201029
2	ad_496638543257_keyword_
3	ad_485085682967_keyword_ipsy glam bag x_000915_20201208
4	ad_390104149092_keyword_glam bags_000451_20200806
5	202101_GBP_Tracking_1835559
6	ad_485427750181_keyword_box subscription_000924_20201209
7	20210215_Offer_Act_TEAMI_1995800
8	ad_390180579509_keyword_ipsy glam bag_000451_20200806
9	ad_494866870494_001058_20210127
10	ad_493968399724_keyword_like birchbox_000366_20200624
11	20200201_Shopper_Act_MysteryBag_ValentinesDay_1954874
12	ad_488295703199_keyword_beauty subscription boxes_001001_20201223

Hundreds of ads at a campaign or ad level are sending people to the site daily. In this thesis, we applied the Markov model to a set of 20 different channels, but in reality, every channel contains hundreds of different ads which need to be optimized, scaled or deactivated. Based on current research, there is no exploration of the multi touch attribution at a campaign level, which is most useful for marketers. The reason for this is that the complexity increases exponentially as well as the resource consumption to run the model, which is something not all companies can afford to do.

A/B Testing Attribution Models

A/B testing (also known as split testing) is a process of showing two variants of the same web page to different segments of website visitors at the same time and comparing which variant drives more conversions [VWO]

Figure 12. A/B Testing example



A/B Testing can be used to determine whether changing the UI leads to higher conversions. Source.

In the Machine Learning world, A/B testing can be accomplished in a different way. The idea is simple, how can the company be certain that the Data Driven Attribution model built is the best option out there. As discussed before, there are several KPI's and methodologies for measuring the performance of machine learning models.

A good approach for constantly iterating and improving a Multi touch attribution model is to test it against other variants on a daily basis. By configuring different parameters or even weights to the channels companies can create multiple models running in parallel. Each model's output will present the Channel's performance of Cost Per Subscriber (conversion). For example, Model 'A' could say the ROAS of Facebook is +50%, stating that for every dollar spent , $1.50 return as revenue from that ad. But Model 'B' could say that the channel's ROAS is 0% which means that the channel is generating the same exact money being spent in revenue. When adding COGs and other types of costs then the channel could be unprofitable. A big challenge for marketers, since both models are stating different outcomes and actions to be taken. For this reason, it is vital for companies to measure and monitor several models in parallel.

A good approach to stress test models on a daily basis would be to generate real A/B testings. For example, given the targeting possibilities in platforms the marketing team could optimize some campaigns as the Model A states and other campaigns or channels as the Model B states. Always thinking that the experiment should be scientifically determinant as much as possible. In other words, the same budget should be used for both, both samples (targeting) should not overlap, and the behaviour of both samples should look alike.

If a company wants to A/B test two ways of investing in ads and they choose two different similar states, there are many things that could go wrong. Every state could have their own particularities like California's wildfires, or an election, or sudden changes . Fortunately, platforms have some amazing tools to divide users into several groups for testing. Moreover, the time period for analyzing the experiment should also be thought of. Depending on the budget allocated, the time period, and the amount of users impacted then the experiment can become valid and the results significant.

At the end of the day, marketers could measure the total performance of Budget A vs Budget B using the different models and the one that has a higher ROI should be the winner. The model can also be experimented with the traditional heuristics like last click or first click. The scientific proposition of Data Driven Attribution models is that as they are more sophisticated they can allocate better budget and thus drive increased ROI. But, based on actual business experience, the best idea is always to experiment, test and compare the models against current or variant scenarios.

ROAS Modifiers

ROAS, defined as Return over Ad Spend, is the primary KPI marketers analyze when optimizing campaigns and choosing which campaigns to inject more money and resources. Marketing departments usually start their day by checking the performance of all their campaigns in the Channel's UI. This means that, for example, Facebook that contains and shows their attribution model will show based on last-click which are the Best Performing campaigns and what is the ROAS for each.
As we discussed earlier, these results are often biased for a couple of reasons:

1. Facebook has an incentive to show the best possible results using its data. In attribution language, they only use Paths containing Facebook Campaigns. So the view is not only limited, but they are overestimating their conversions.
2. Heuristic Models only containing Facebook Data.
3. Facebook Pixel.

This applies to every marketing channel that uses only their information to attribute conversions and conversion value. If marketers need to keep optimizing their budgets in the

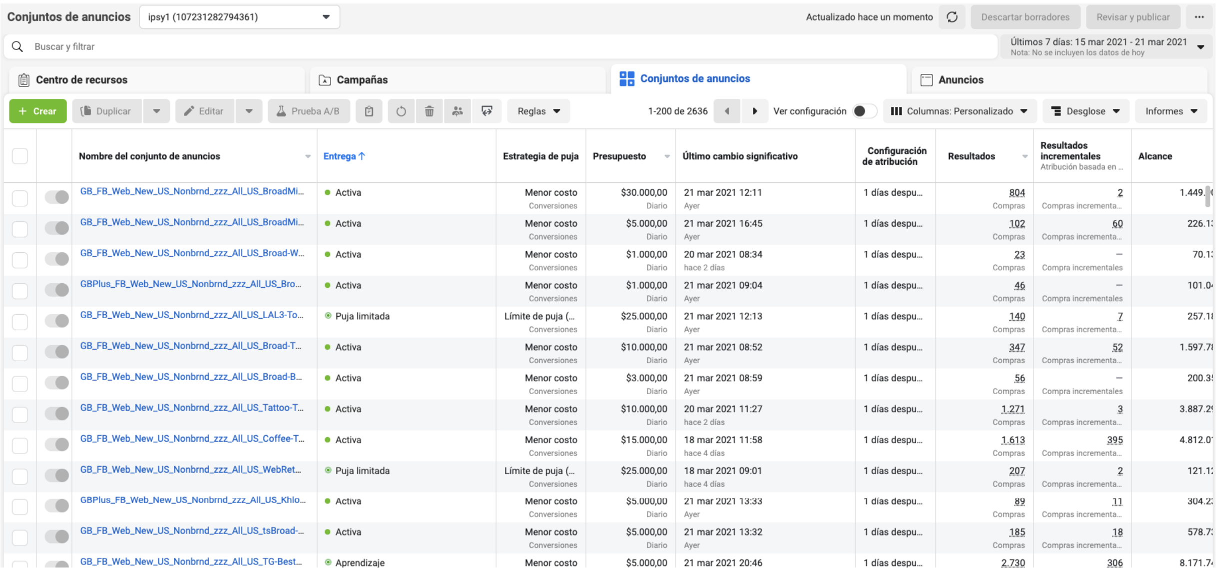
marketing channels and the business only accounts with a data-driven aggregated model, then ROAS modifiers could be used.
The Conversions generated by a Channel can be tracked over time while contemplating the Conversions being attributed by the data-driven model daily. For example:
Facebook UI - Last 7 Days - 100 Conversions
Data-Driven Model (local) Facebook - Last 7 Days - 80 Conversions
In practice, this means that Facebook is over-attribution around +25% more conversions than it should. By calculating the CPS, let us imagine the $1000 were spent on Facebook in that timeframe. The CPS is $10. Nevertheless, as noticed, the real CPS is different based on the model. The Real CPS is $1000/80 = $13. A simple application for marketers is that they should always apply a factor of 80% to the conversion value observed in the Channel's UI.
To enhance this model further, a distribution could be analyzed, and the ROAS Modifier (0.8) could be dynamically calculated based on different factors like:

- % of Channel Spend of overall Marketing Budget: This is straightforward; as the channel has more weight in the marketing mix, it will over attribute more conversions.
- Amount of Money Spent in the Channel: Every Business has a Customer Profile to be targeted. When the Channel saturates the potential customer audience, the efficiency tends to drop. As more money is spent on a Target Audience, more impressions are covered, increasing over-attribution since every potential customer is impacted.
- Type of Channel (Video, ad, programmatic, search): Generally, Social Networks tend to over attribute more than others.
- Platforms UI Attribution Method: It is common in the industry that the attribution times are one day, seven days, 28 days. This is both for Click Attribution and Impression Attribution. As the timeframe increases, it means that if an ad was shown to a user 27 days ago, and they signed in that day, it would be credited for that ad shown 27 days ago. The longer the attribution timeframe, the more the channel over attributes itself.
- Time Dimensions: Day of Week, Month, Season, Special Days

The ROAS modifiers could be enhanced as much as the company needs for daily operations. In practice, this could mean a table that contains the Modifier for every channel, campaign, and day of the month that should be used. (e.g., Based on the ROAS model, last seven days marketing mix of spend, day of the week, month, channel and campaign → Apply 0.85 factor to Facebook Video in News Feed for this week.).

Using these modifiers, marketers could approximate the UI data, the data attribution model, assign better allocation of resources, and hopefully, improve the performance indicators. In other words, they are paying less to acquire customers and bringing more conversions to the business.

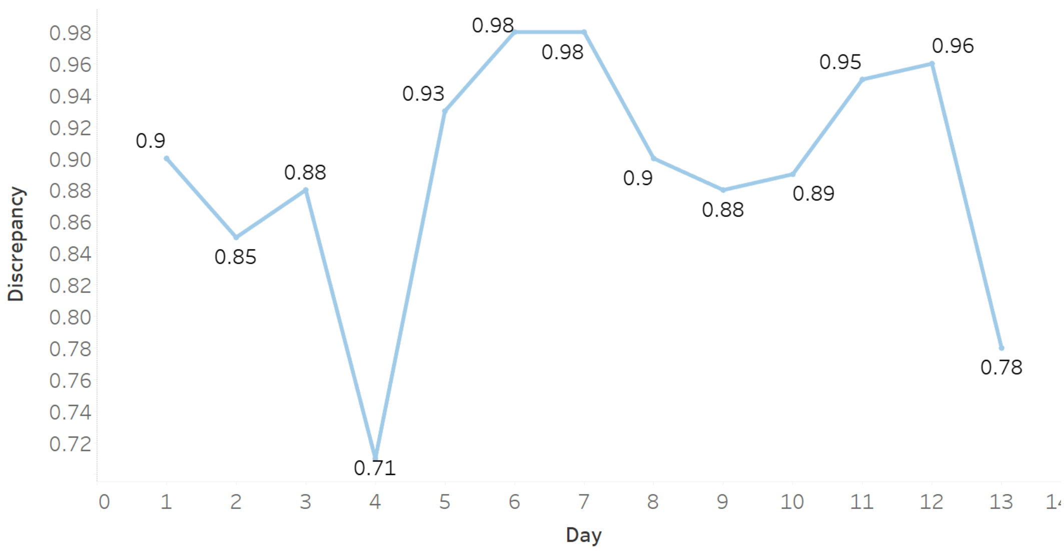
Figure 13. Facebook Ads manager for optimizing campaigns



Nombre del conjunto de anuncios	Entrega	Estrategia de puja	Presupuesto	Último cambio significativo	Configuración de atribución	Resultados	Resultados incrementales Atribución basada en ...	Alcance
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMi...	Activa	Menor costo Conversiones	$30.000,00 Diario	21 mar 2021 12:11 Ayer	1 días despu...	804 Compras	2 Compras incrementa...	1.449.
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_BroadMi...	Activa	Menor costo Conversiones	$5.000,00 Diario	21 mar 2021 16:45 Ayer	1 días despu...	102 Compras	60 Compras incrementa...	226.1
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-W...	Activa	Menor costo Conversiones	$1.000,00 Diario	20 mar 2021 08:34 hace 2 días	1 días despu...	23 Compras	— Compra incrementales	70.1
GBPlus_FB_Web_New_US_Nonbrnd_zzz_All_US_Bro...	Activa	Menor costo Conversiones	$1.000,00 Diario	21 mar 2021 09:04 Ayer	1 días despu...	46 Compras	— Compra incrementales	101.0
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_LAL3-To...	Puja limitada	Límite de puja (... Conversiones	$25.000,00 Diario	21 mar 2021 12:13 Ayer	1 días despu...	140 Compras	7 Compras incrementa...	257.1
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-T...	Activa	Menor costo Conversiones	$10.000,00 Diario	21 mar 2021 08:52 Ayer	1 días despu...	347 Compras	52 Compras incrementa...	1.597.7
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Broad-B...	Activa	Menor costo Conversiones	$3.000,00 Diario	21 mar 2021 08:59 Ayer	1 días despu...	56 Compras	— Compra incrementales	200.3
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Tattoo-T...	Activa	Menor costo Conversiones	$10.000,00 Diario	20 mar 2021 11:27 hace 2 días	1 días despu...	1.271 Compras	3 Compras incrementa...	3.887.2
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_Coffee-T...	Activa	Menor costo Conversiones	$15.000,00 Diario	18 mar 2021 11:58 hace 4 días	1 días despu...	1.613 Compras	395 Compras incrementa...	4.812.0
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_WebRet...	Puja limitada	Límite de puja (... Conversiones	$25.000,00 Diario	18 mar 2021 09:01 hace 4 días	1 días despu...	207 Compras	2 Compras incrementa...	121.1
GBPlus_FB_Web_New_US_Nonbrnd_zzz_All_US_Khlo...	Activa	Menor costo Conversiones	$5.000,00 Diario	21 mar 2021 13:33 Ayer	1 días despu...	89 Compras	11 Compras incrementa...	304.2
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_tsBroad-...	Activa	Menor costo Conversiones	$5.000,00 Diario	21 mar 2021 13:32 Ayer	1 días despu...	185 Compras	18 Compras incrementa...	578.7
GB_FB_Web_New_US_Nonbrnd_zzz_All_US_TG-Best...	Aprendizaje	Menor costo	$5.000,00	21 mar 2021 20:46	1 días despu...	2.730	306	8.171.7

The results shown in the Facebook UI, as discussed previously, show the narrowed attributions based only on Facebook Ads. For the first adset, 804 conversions are stated by the UI. The Data Driven Attribution model gives the “real” number of conversions provided by a given channel or campaign. In practice, running the model every day and comparing the model value vs the UI’s value will give a % of discrepancy as a time series.

Figure 14. Discrepancy between Facebook UI vs Markov on a daily basis



The above chart is created by running the Markov model on an adset level (instead of channel level as done in this paper). By doing a quick prototype at an adset level, we can evaluate how many conversions are stated by the Markov Model for the adset: GB_FB_NEW_US_NONBRND. The Facebook UI shows how many conversions the adset had every day based on their model. We can compare these numbers against our model. Using Day 1 as an example, the number around 90% represents that for every 10 conversions that the Facebook UI shows, the Markov model only outputs 9. This extra 1 conversion could be considered as an over-attribution from Facebook.

On average, the model states that Facebook Conversions are 92% of what the Facebook UI shows. This means that the marketers should check Markov's model output before scaling or descaling ads on the UI.

LTV

Most of the companies that spend money on ads analyze the performance of their budget based on the conversions generated. In short, the focus is in the short term of the investment, Day 1 $100 is spent and the KPIs analyzed are how much money the ads brought in Day 2 or 3 or 4. In the most well known channels, the attribution analyzed is frequently 1 day click or 7 days clicks. Some use 30 days posterior to the click but it is the least used.
How many days have passed since the user viewed or clicked the ad and what were the actions or conversions done by that user. For example:

1. Day 0 User A views Ad
2. Day 1 User A clicks Ad
3. Day 5 User A purchases product

In a 1 Day Click the ad that the user viewed and clicked would not be attributed to the conversion. But in a 7 day click setup , then that purchase would correctly be attributed to that ad. Marketers usually try to narrow down what is known as the "attribution window" to minimize the impact of other channels in the customer journey. This equation simplifies decision taking as the immediate impact of the ad can be seen in the 24 hour window following the view or click.

As observed, this is a limited view of the world that is ingested into the Multi Touch Attribution model created. Furthermore, if channel A generated a lot of converting users that were following an offer and never bought again on the site, these users are of "less value". For example:

1 day after clicking the ad

Channel A - 100 users purchase at $1000 → ROAS = $1000/ 100 = 10

Channel B - 100 users purchase at $500 → ROAS = $500 / 100 = 5

Any marketer would state: "Scale Channel A" , meaning to allocate more budget into channel A.

But what would happen if in 6 months those 100 users demonstrated to be repeated purchasers or high value customers. If we followed that cohort of users through time:

Channel A - 100 users purchase at $2500 → ROAS = $2500/ 100 = 25

Channel B - 100 users purchase at $5000 → ROAS = $5000 / 100 = 50

Basically, customers value change over time. All this information could be provided to the algorithmic model, logically, with time. Each channel's performance will be changing over time as customers advance into more mature flows. Some channels might provide high intent customers associated with "offers purchase" while others could be more loyal to the brand in the long term.

To introduce this new concept to the model, we would then need to create a LTV model for each user, correctly identifying which Channel brought the user for the first time. The scope of this thesis is not about an algorithmic LTV model, but it's definitely worth how we could extrapolate those results into the MTA model to improve its performance in the long run. This thesis business case is about a Subscription Business, so the LTV model would need to be adapted into the subscription scenario. In this case, the output of the LTV Model would look something like this:

User	Channel Acquirer	Value LTV	Number of transactions/Months Active	DT
User A	Facebook	$100	10	3/1/2020
User B	Google	$50	5	3/1/2020

With this table, we can create a time series of the average LTV value generated by each channel in time.

1. Group by table by channel
2. Average Value LTV by channel
3. DT used in group by for time series

For each cohort of Date Acquired and Channel, a number representing the actual average LTV generated by each channel. The next step is to find a way to correctly introduce the new

information into the Markovian Model for example. As analyzed, the Conversion Value parameter states the real value of the user acquired, which is then used to assign the correct credit to the channel that brought the user. The final output of the model would look really similar to the classic one:
Markov's Model Output for the dataset. It is the same table from the Model section.

Table 13. Adding a new column Conversions Value to the model's output

Id	Channel Name	Total Conversions	Total Conversions Value
1	Facebook	529996	$7,949,520
2	Google	398303	$7,966,060
3	Organic	352849	$8,821,665

The only difference is that when using the Conversion Value, we would have a new column stating the actual value generated by the channel that could be monetary or not. In the above example, the Total Conversions Value would come from the model's output (the numbers used are purely an example). Analyzing the example, Facebook's conversions have a lower value of $15 each, while Google $20 and Organic $25. This improves the model as we include LTV for each of the channels.

For Subscription business models there are LTV models created that have good performance. Some extra tuning on the model like 6 months LTV or 12 months LTV should also contribute to the long term business KPIs. As commented, new variables could be inserted in the model to better estimate channel's performance in the long term.

Marketers could use this new KPI to calculate a projected ROAS, CPS and ROI to allocate more resources into long term channels when needed. They could also mix up both present and future views when needed to push more money into "fast acquirer" channels or more long term high value customers.

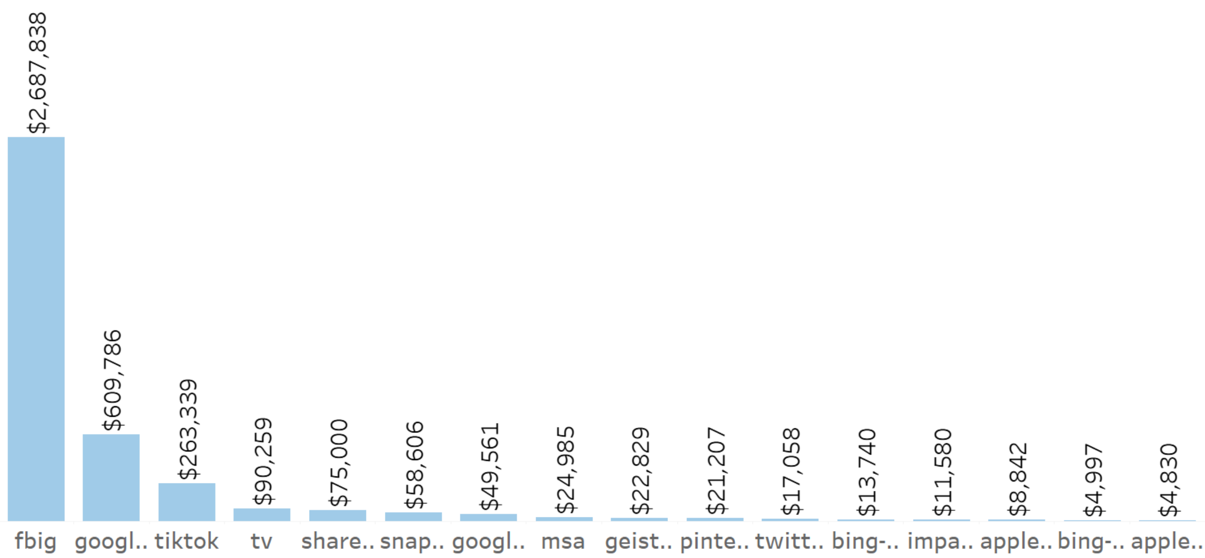
Business Recommendations

Most of the companies that are trying to expand their product or services allocate a great sum of resources to their Acquisitions teams. As discussed, the objective is to grow the user base

as maximum as possible while spending the minimum amount of money per user acquired. All our investigation and analysis is focused on finding ways for companies to evaluate more accurately the true results of their marketing efforts.
For the present case study, the biggest problem is: How to assess the real results of a channel like Facebook when we know that our reported Cost Per Signup(CPS) might not reflect the actual truth?

The initial analysis included 30 days of data with 1.542.555 different paths. The data contains a total of $3.968.670 spent in Digital Marketing in the period analyzed (30 days).

Figure 15. Amount of money spend by channel in the time period analyzed



Based on our Markov model we can now evaluate if the Marketing spend was distributed correctly amongst all the channels or if there could be improvements. Something important to note is that the company is willing to pay up to $10 for a signup which is its break-even profit point. This means that by paying that amount they can recover the marketing spend in new subscriptions. The projected LTV model for signups is responsible for getting the break even point.

First of all, the use of the ROAS modifiers could be highly beneficial for the company. As analyzed, Facebook UI is over-reporting signups at a 1.08 rate. This means that when optimizing campaigns, the CPS of the platform should be inflated by 1.08%. If the break-even is $10, the marketers should work with Facebook CPS divided by 1.08. The new true break-even CPS for Facebook should be $9.25 based on the attribution model, regardless of what the user interface says. This could be improved since 1.08 is just an average for the time period, by playing with the weekdays and differentiating special dates.

Secondly, we have seen that multi-touches have a higher probability of conversion than only one touch in the customers' journey. Marketers in this scenario, should try to create new ways for users to interact with multiple ads. The best way to do this is to enhance more Display or Awareness ads. This is a hot topic nowadays strictly related to the Attribution problem. In a last click world, many marketers often do not see the results of impression ads. In this paper, we have analyzed increased performance of up to 15-25%. For this company in particular most of the paths look like:

[Facebook] > Conversion
[Google] > Not Conversion

The ideal would be to reduce the 80% one-touch paths up to 50% playing more with awareness channels like Pinterest, TV, Offline ads and Youtube for example. More paths looking like:

[Facebook] > [Pinterest] > [Radio] > [TV] > [Google] > Conversion

Moreover, based on the current analysis and the transition matrix, there are some channels that can be considered as optimal finishers of the conversion. One clear example is Google. This makes sense, as people often search for the product before converting. The recommendation for marketers is to align channels like Google Search, Blog, Affiliate marketing at the end of the funnel being super focused in finishing the conversion with excellent call to actions. This company could be finishing conversions way better by understanding the multi-touch attribution and the results provided in this investigation. In practical terms, the recommendation would be to improve CTAs, drive traffic to google or others from awareness ads to create the multi-touch path which is more efficient.

One of the core enhancements to Markov provided in this thesis is the Individual Level Attribution. The dataset provided shows for every user, the actual conversion weight for each of the channels they interacted with.

Table 9. Reference to ChannelAttribution's output of Markov Attribution at an individual level

```
In [71]: markov['path_attribution'][19:30]
```

Out[71]:

	idpath	path	channel	total_conversions_weight	total_conversions_attribution	total_conversion_value_weight	total_conversion_value_attribution
19	17	Email > organic	organic	0.469087	2980.108976	0.469087	2980.108976
20	17	Email > organic	Email	0.530913	3372.891024	0.530913	3372.891024
21	18	Link	Link	1.000000	6147.000000	1.000000	6147.000000
22	19	organic > organic > organic	organic	1.000000	4947.000000	1.000000	4947.000000
23	20	MSA	MSA	1.000000	4834.000000	1.000000	4834.000000
24	21	TikTok > TikTok	TikTok	1.000000	4757.000000	1.000000	4757.000000
25	22	Instagram > organic	Instagram	0.521230	2171.965121	0.521230	2171.965121
26	22	Instagram > organic	organic	0.478770	1995.034879	0.478770	1995.034879
27	23	Email > Email	Email	1.000000	3730.000000	1.000000	3730.000000

This data can be mixed up with customer information with features like:

- Age
- Type of Customer
- Products acquired
- Demographics
- LTV
- Churn

With this information the company could calculate the optimal converting paths for specific clusters of customers. For example, the highest converting path for customers age 18-25 is:
[Pinterest > Social > Facebook > Conversion]
While for the range 45-65 is:
[TV > Radio > Google > Blog > Conversion]

In summary, we recommend the company starting using the ROAS modifiers provided for optimizing campaigns on a daily basis. Running the model on a weekly basis so that the modifiers can adapt to the most recent scenario. The true CPS for each of the channels is provided with the output Markov model, so that they can optimize budgets and scale or descale based on their break-even point, in this case $10.

We also recommend enhancing the Customer journey, where we concluded increased results on Multi-touch paths of up to 25%. By sharing some recommendations for creating the most efficient paths for conversion, both in days and in format, we can expect increased Returns in media spend by reducing the ratio of non-converting paths.

Conclusions

In addition to the academic contribution, this paper makes several managerial contributions. The idea was to install an easy way for marketers to explore the top performance Data Driven algorithm into a business scenario while exploring the different applications and impacts it could have. From Budget Allocation to other trivial problems like Campaign Optimization, Conversion Value and ROAS Modifiers. The industry is starting to shift their attribution models towards Data driven methods but there is still a long way to reach the final goal. Models should have not only more explainability, robustness and predictive accuracy but also real applications and decisions for the marketer or CMO.

The project was done in collaboration with a US Retailer company focused mainly in Digital Subscription products in the Beauty industry. Some of the data was anonymized to protect the naming conversions, real investment, and real conversions of the business. This thesis required expertise in the marketing domain which was given by several of the marketers in the company.

The main challenge was working with the datasets. The click-stream data is often messy and disorganized. There are millions of events in the log-level data of any website that has a high amount of traffic. Every campaign had to be identified to the correct channel and the overall conversion events needed to be prepared as well. The real work was to model the data accordingly for the attribution problem, meaning that for each user we needed to find all the different sources and timestamps by which they entered the website within a timeframe. Furthermore, another problem arises when trying to match effectively the internal data from the clickstream logs to the Channel's UI data.

For the attribution modelling, once the data was in the correct structure and loaded into Databricks/Jupyter Notebook it was easier to run the models since there are excellent libraries that do the work (e.i: ChannelAttribution by David Altomare). By preprocessing the data as a classification machine learning problem with the Binary Outcome 0/1, the model could calculate the transition probabilities and removal effects of each of the channels. The purpose of this thesis was to spend the majority of the time and effort in thinking and applying creative ways of using the Markov Models output into real life marketing scenarios.

Once the Markov Model was run and its results were compared with Heuristic Models and analyzed its predictive accuracy, seven different ways to apply the results were explored. This is the richer research and analysis that has not been explored in other papers. Strictly related to the Marketing business areas, and after having discussed the problem with the VP of Marketing in big companies, their first necessity was to have a level user attribution model. Nowadays Customer Data platforms are the most important tools that marketers have and being able to identify each of the paths for every user is extremely valuable. The main challenge was to translate Markov's aggregated output for each of the paths to every Customer's Journey. Luckily, there are a couple of people working on this problem and the ChannelAttribution library helped me output the model into a tangible format for marketers to work with.

Other interesting topics and ideas were explored such as ROAS modifiers or LTV models stacked into the Data Driven Attribution model. At first, theoretically explored since it would require further research and new datasets, as well as new sponsors in the company. In terms of Attribution modelling, the main topic is still how to implement, improve and apply the models into real life situations and obviously extract value from them. I think that adding some of the ideas explored could be substantially beneficial to companies' acquisition teams, improving the ROI and other business metrics.

The attribution problem is frequently left up to the channels to solve or dismissed using simple heuristics. The truth is that it is a broader area that is still in its infancy as there is a lot to research and explore towards better solutions. In research, there is a trend of more investigation in marketing science topics which can be groundbreaking for the industry that is massive.

Finally, this study shows clear paths and indications of how to apply the algorithmic models into the business and how to approach it from different angles. We hope to encourage more companies to start using this kind of approach and start evaluating the impact. Some of the flaws of the heuristic models have been discussed and how this could lead to suboptimal allocation of marketing budgets.

Limitations of this study & Further Research

A couple of limitations in this study are presented both by the dataset and by the approach taken.

One of the biggest pitfalls of an MTA model is not using Impression Level Data currently owned by the most prominent publishers (Google, Facebook). It would be crucial to have that information as it would provide a richer model with sizable paths as it contains all the impressions generated for a person. In this sense, the model is partially blind-folded since many of the Organic sessions are generated by ad impressions. It is well known in the industry that channels like Youtube, Pinterest, or Display generate lots of views at cheap CPMs but are constantly under-weighted in the final models. The same happens with Offline ads. Further research could focus on integrating some of the Machine Learning Models explored by the current state and enhance them with this type of information. The idea is always to replicate reality as closely as possible, to understand precisely all the factors and drivers that contribute to conversations.

Secondly, an essential piece of information is not fed into the model in this analysis: the Timestamps. In some cases, the paths' length is diverse amongst all the conversions, which s a critical variable to consider. If conversions frequently occur one hour after a channel is present, it is decisive to acknowledge it as high value. If a Channel is far away from the conversion (in time), it should be weighted accordingly and less than channels that are incredibly close to the conversion.

Third, the paper does not solve the problem of how to create the perfect path. This means that given the correct attribution and understanding of all possible paths with their Transition Matrix, how can we create a mathematical framework that helps marketers custom their Customers' Journey for optimal performance. In our analysis, we do some exploratory data analysis and give some of the intuition. However, it would be interesting to have another model which uses the MTA model as input and generates an Optimal Budget Allocator model to generate the perfect scenario (involving linear programming).

Furthermore, cookie-based tracking technology presents some difficulties, which makes data sometimes unreliable and messy. Technology in the future might resolve this issue with a new type of tracking methodology that could dissipate some of the disadvantages of cookies that can be erased, modified, and more.

Even though some ways of validating the models in terms of accuracy, interpretability, and robustness, the industry lacks stress-testing methods in real life that the companies can use to observe whether their new MTA model is superior to previous heuristics. In this paper, we explore some ideas of A/B Testing and experimentation that could be done to add new channels and explore both models. However, it would be determined to generate something bigger and more certain to validate them.

Finally, most of the papers examine only the point of conversion and finish their analysis there. A further research idea that could add value to companies is finishing the model when the conversion occurs and evaluating it into the future. In short, what channels provide the best customer retention and LifeTime Value? This is an exciting topic since most businesses care about getting customers and acquiring them for the long term. It is well known in the industry that some channels provide low-quality customers that mainly buy once and are driven by offers. A model that contemplates the conversion into the future would bring many new insights and possibilities to marketers. A simple example to analyze this could be: The MTA model says that Channel A and Channel B generated 1000 converters. However, logically, these customers' quality differs substantially since the channel has a specific target of customers/users. If the future value of channel A is more significant than that of channel B, it makes sense to invest more money today in Channel A.

List of Abbreviations and Acronyms

AD	Advertisement
AOV	Average Order Value
AUC	Area Under Curve
FP	False positive
FN	False negative
GA	Google Analytics
IAB	Interactive Advertising Bureau
LR	Logistic Regression
CPS	Cost Per Subscription
ROC	Receiver Operator Characteristic curves
TP	True Positive
TN	True Negative
TPR	True Positive Rate
URL	HTTP URL
ROI	Return Over Investment
ROAS	Return Over Ad Spend
CPA	Cost per Acquisition
MTA	Multi Touch Attribution
ML	Machine Learning
CTR	Click Through Rate
CPM	Cost per mile impressions
CR	Conversion Rate

References

[1] Abhishek, V., Fader, P., & Hosanagar, K. (2015), "Media Exposure through the Funnel: A Model of Multi-stage Attribution," Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2158421

[2] ACM Hülsdau, M., & Teuteberg, F. (2018, March). Towards a taxonomy of algorithmic attribution models– Which is the right model to measure, manage and optimize multiple campaigns? Paper presented at MKWI 2018. Retrieved from http://mkwi2018.leuphana.de/programm/sessions/

[3] Aggarwal, A.G., & ., A. (2018). A Multi-attribute online advertising budget allocation under uncertain preferences. Ingeniería Solidaria, 14(25), 1-10.

[4] Altomare, D., & Loris., D, (2018). ChannelAttribution: Markov Model for the Online Multi-Channel Attribution Problem. R package version 1.12.

[5] E. Anderl, I. Becker, F. V. Wangenheim and J. H. Schumann, "Putting Attribution to Work: A Graph-Based Framework for Attribution Modeling in Managerial Practice " 23 October 2013 [Online]. Available:

[6] K. Bill, "Attribution Playbook Google Analytics " Google 2012 [Online] Available: http://services.google.com/fh/files/misc/attribution_playbook.pdf. [Accessed 17 February 2014].

[7] J. Brnwee , Article. Available at: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/

[8] Berman, R. (2017), "Beyond the Last Touch: Attribution in Online Advertising." Available at SSRN: http://ssrn.com/abstract=2384211

[9] J. Chandler-Pepelnjak, "Measuring ROI Beyond the Last Advertisement " [Online]

Available://atlassolutions.com/wwdocs/user/atlassolutions/en-us/insights/dmiMeasuringROIBeyondLastAd.pdf. [Accessed 17 February 2014].

[10] Chatterjee, P., Hoffman, D. L., & Novak, T. P. (2003). Modeling the clickstream: Implications for web-based advertising efforts. Marketing Science, 22(4), 520-541.

[11] Clifton, B. (2012). Advanced web metrics with Google Analytics. Indianapolis, IN: John Wiley & Sons.

[12] Constantinides, E. (2002). The 4S web-marketing mix model. Electronic commerce research and applications, 1(1), 57-76.

[13] C. H. W. Jayawardane, S. K. Halgamuge, U. Kayande. 2015. Attributing Conversion

[14] D'alessandro, B., Perlich, C., Stitelman, O., & Provost, F. (2012, August). Causally motivated attribution for online advertising. In Proceedings of the Sixth International Workshop on Data Mining for Online Advertising and Internet Economy (p. 7).

[15 Filippi, F. (2002). Daniel Kahneman: Psicologia e decisioni. Università, Dipartimento di scienze economiche.

[16] M. Fischer, S. Albers, and M. Frie, "Dynamic Marketing Budget Allocation Across Countries Products and Marketing Activities " Marketing Science vol. 30, nr. 4, pp. 568-585, 2011.

[17] A. Goldfarb and C. Tucker, "Online Display Advertising: Targeting and Obtrusiveness " Marketing Science vol. 30, nr. 3, pp. 389-404, 2011.

[18] G. Iyer, D. Soberman and J. M. Villas-Boas, "The Targetting of Advertising " Marketing Science, vol. 24, nr. 3, pp. 461-476, 2005.

[19] Keilson, J. (2012). Markov chain models—rarity and exponentiality (Vol. 28). Springer Science & Business Media.

[20] Kotler, P., & Armstrong, G. (2010). Principles of marketing. Pearson education

[21] P. Kireyev, K. Pauwels and S. Gupta, "Do Display Advertisements Influence Search? Attribution and Dynamics in Online Advertising, workin paper " Harvard Business School, 2013, February

[22] Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. Journal of Marketing, 80(6), 69-96. 55

[23] Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. Journal of Marketing

[24] Molnar, C. (2018). Interpretable machine learning. A guide for making black-box models explainable. Retrieved from https://christophm.github.io/interpretable-ml-book

[25] P. Manchanda, J.-P. Dubé, K. Y. Goh and P. K. Chintagunta, "The Effect of Banner Advertising on Internet Purchasing " Journal of Marketing Research vol. XLIII, pp. 98-108, 2006.

[26] F. Nottorf, "Modeling the Clickstream Across Multiple Online Advertising Channels Using a Binary Lo it With Bayesian Mixture of Normals " Electronic Commerce Research and Applications, vol. 13, nr. 1, pp. 45-55, 2014.

[27] N. Necsulescu. 2015. Multi-channel attribution and its role in marketing investment. Journal of Digital & Social Media Marketing 3(2), 125-134.

[28] Rentola, O. (2014), Analyses of Online Advertising Performance Using Attribution Modeling, MSc. Thesis, Aalto University, Helsinki.

[29] X. Shao and L. Li, "Data-Driven Multi-Touch Attribution Models " Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data mining, 2011.

[30] Windsor.AI (n.d.), Optimise marketing ROI with multi-touchpoint attribution modeling. Retrieved from https://www.windsor.ai/attribution-modelling/

[31] Wooff, D., & Anderson, J. (2013) Time-weighted attribution of revenue to multiple e-commerce marketing channels in the customer journey., Working Paper. Durham Research Online.

[32] M. M. Yadagiri, S. K. Saini, R. Sinha. 2015. A Non-parametric Approach to the Multichannel Attribution Problem. In: Proc. of the 16th International Conference on Web Information Systems Engineering, Springer, 338-352

[33] S. Yang and A. Ghose, "Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? " Marketing Science, vol. 29, nr. 4, pp. 602-623, 2010.

[34] Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. International Journal

Zhao, K., Mahboobi, S.J., & Bagheri, S.R. (2018). Shapley Value Methods for Attribution Modeling in Online Advertising. Papers arXiv:1804.05327 [econ.EM]. https://arxiv.org/abs/1804.05327

[35] K. Zhao, S. H. Mahboobi, S. R. Bagheri, 2018. Revenue-based attribution modeling for online advertising. International Journal of Market Research, 21 pages.

[36] Zhang, Y., Wei, Y., & Ren, J. (2014, December). Multi-touch attribution in online advertising with survival theory. In Data Mining (ICDM), 2014 IEEE International Conference (pp. 687- 696). IEEE

Articles

[36] IAB Attribution Working Group, "Attribution Primer " 21 June 2012 [Online]. Available: http://www.iab.net/media/file/AttributionPrimer.pdf. [Accessed 10 February 2014].

[37] Redefine the omnichannel approach: Focus on what truly matters Mckinsey https://www.mckinsey.com/business-functions/operations/our-insights/redefine-the-omnichannel-approach-focus-on-what-truly-matters#

Tables & Figures

Table 1. Dataset used for the Attribution model dimensioned by userid

Table 2. Final dataset used for Markov model. Denotes each of the touchpoints for each user

Table 3. Markov's output dataset

Table 4. Aggregating the Markov output at a channel level

Table 5. Example of the Markovian transition probabilities to calculate the removal effect.

Table 6. Transition matrix output by ChannelAttribution library in R.

Table 7. R's output cell showing the AUC for each Markovian order hyperparameter

Table 8. R's output cell showing the Classification Matrix input

Table 9. Reference to ChannelAttribution's output of Markov Attribution at an individual level

Table 10. Example of a CPS calculation using UI data.

Table 11. Example of the Attribution input dataset when adding impression information

Table 12. Dataset containing campaign information from internal Measuring tool

Table 13. Adding a new column Conversions Value to the model's output

Figure 1. Venn diagram stating the different attributes of Attribution models based on the algorithm

Figure 2. Example of a Path to conversion shown as a Markovian graph with its transition probabilities

Figure 3. Number of paths at a user level. On the right, the distribution of users by source.

Figure 4. The distribution of occurrences at a path level for each of the multi-touch paths in the dataset.

Figure 5. The number of days that pass from the landing timestamp to the conversion timestamp aggregated at a user level.

Figure 6. Console information when running the Markov model in R, using ChannelAttribution library.

Figure 7. Number of signups per channel dimensioned by the different attribution algorithms.

Figure 8. Example of a Classification matrix for machine learning.

Figure 9. Example of a ROC Curve visualized for measuring the model.

Figure 10. Example Google Ads display network types of ads

Figure 11. Facebook UI adset information

Figure 12. A/B Testing example

Figure 13. Facebook Ads manager for optimizing campaigns

Figure 14. Discrepancy between Facebook UI vs Markov on a daily basis

Figure 15. Amount of money spend by channel in the time period analyzed

Additional Figures

Figure 16. Comparative chart by Anderl et al [2013]

Study	Methodology	Evaluation Criteria					
		Objectivity	**Predictive accuracy[1]**	**Robustness**	**Interpretability**	**Versatility**	**Algorithmic efficiency**
Shao and Li (2011)	(1) Bagged logistic regression (2) Simple probabilistic model	No; frequency of contacts and positions not considered	Yes	Yes	Yes	Yes	Not available
Dalessandro et al. (2012)	Causally motivated methodology based on cooperative game theory (Shapley value) combined with logistic regression	No; frequency of contacts not considered	Yes	Not measured	Yes	Yes	Not available
Abhishek, Fader, and Hosanagar (2012)	Dynamic Hidden Markov Model	Yes	Yes	Not measured	Limited	Limited; assumptions on channels and structure of decision process	Not available
Li and Kannan (2013)	Bayes	Yes	Yes	Not measured	Limited	Limited; assumptions on channels and structure of decision process	Not available
Kireyev, Pauwels, and Gupta (2013)	Multivariate time-series model (persistence modeling)	No; not based on individual data	Yes	Not measured	Limited	Limited; application based on 2 channels (display and SEO)	Not available
Haan, Wiesel, and Pauwels (2013)	Structural vector autoregression	No; not based on individual data	Yes	Not measured	Limited	Limited; not suited for performance-based channels (e.g. affiliate)	Not available
Our study	Markov walks (first- and higher-order)	Yes	Yes	Yes	Yes	Yes	Yes

[1] This table only indicates if predictive accuracy is evaluated in the respective study. The data sets used and implementation details are not publically available, and the measures vary, so a comparison of predictive accuracy across studies is not possible.

Figure 17. Markovian graph in R for 9 channels and their transition probabilities