NOVA IMS
Information Management School

MGI

Mestrado em Gestão da Informação
Master Program in Information Management

The Impact of Attribution Modelling in Luxury E-Commerce

Attribution Model simulation

Filipe Andre Borges Chaves

Project work submitted in partial fulfillment of the requirements for the degree of

Master of Science in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão da Informação
Universidade Nova de Lisboa

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

THE IMPACT OF ATTRIBUTION MODELLING IN LUXURY E-COMMERCE

by

Filipe Chaves

Project work submitted in partial fulfilment of the requirements for the degree of Master of Science in Information Management

Advisor: Professor Diego Costa Pinto

November, 2018

Acknowledgements

I want to thank the company that allowed me to study and to do my project work with their data, specially to the manager and product owner that gave so many ideas to choose from.
Thanks to my advisor that gave me the tools, possibility and motivation to do a better work.
Thanks to my parents and brothers that were always there for me and gave the necessary strength to continue my studies.
Diana that heard me in the darkest hours and for all the motivation to finish this work.
And finally to all my friends who understood my lack of time and never let me quit.

Abstract

Luxurious goods started to be sold in online stores at a slow pace, but today the market has grown and it is working at high speed and has great potential. In fashion world exclusivity is the main word and along with the creation of online stores by luxurious brands, it is necessary to advertise them to the right audience and at the right time. So, companies tend to apply their marketing budget in multiple channels (search, email, display, affiliate and social networks) and because there are multiple ways to reward each channel, companies need to decide which Attribution Models they will use when rewarding affiliates. Using a well-known and multi-brand seller as an allied to better understand advertise in ecommerce websites, this project will be based in a real database in order to create accurate simulations and to get the best model for similar companies. The simulations created for this project were based in almost two hundred and fifty thousand conversions and more than two million interactions with the e-commerce website.

Keywords: Online advertisement; attribution modelling; e-commerce; online choices; personalization; online marketing.

Resumo

Inicialmente, os bens de luxo começaram por ser vendidos em lojas online a um ritmo consideravelmente lento, sendo que, atualmente, já é possível ver um rápido crescimento do mercado e o seu grande potencial. Neste mundo em que "exclusividade"é palavra chave e com, cada vez mais, marcas a lançar as suas próprias lojas online, é imperativo atrair o público certo, no momento certo. Desta forma, as empresas tendem a aplicar o orçamento de marketing em vários canais (pesquisa, e-mail, display, redes de afiliados, ou até em redes sociais). Como é possível utilizar diferentes Modelos de Atribuição para definir recompensas, as empresas necessitam de selecionar qual o que deve ser utilizado para premiar cada um desses canais. Através da parceria com uma empresa do setor de luxo, cujas vendas se baseiam em produtos de diversas marcas conhecidas mundialmente, foi possível utilizar neste projecto, desde início, uma base de dados real. Desta forma, foi possível criar simulações precisas de forma a obter o melhor modelo para empresas como esta, tendo sido baseadas em quase duzentas e cinquenta mil conversões e mais de dois milhões de interações geradas na loja eletrónica da empresa.

Palavras-chave: Anúncios na internet; modelos de atribuição; lojas eletrónicas; escolhas na internet; personalização; marketing na internet.

Contents

List of Figures

List of Tables

Acronyms

AM Attribution Model.
API Application Programming Interface.

GDPR General Data Protection Regulation.

PPC Pay per Click.

ROI Return of Investment.

SEA Search Engine Advertisement.
SEO Search Engine Optimization.
SPSS Statistical Package for Social Sciences.

URL Uniform Resource Locator.
USD United States Dollar.

1 Introduction

During the last few years, mankind has faced a huge change in their lives, from daily life always offline to always online (Baron, 2008; Stephen, 2016), leading to more than two billion people using smartphones with internet access on 2021 (Marinucci, 2018). This change began with personal computers and now in developed countries there are more than eighty people in one hundred that have the capability to browse the internet (Smith, McGeeney, Duggan, Rainie, and Keeter, 2015), and in the whole world more than fifty five people (ninety seven in developed countries) in one hundred have mobile-broadband subscriptions (ITU World Communications; ICT Indicators Database, 2017), being able to browse the internet anytime and anywhere.

With this change, physical stores changed their business model and created points of sale in the online world, creating e-commerce websites and connecting to much more customers than ever. With this gap, there were some companies that understood the potential of the market, with wider audiences and visibility, and decided to create marketplaces to sell online what others could only sell inside their physical stores (Schmidt, Dörner, Berg, Schumacher, and Bockholdt, 2015). At start, this companies are not known by the public in general, so they need to advertise what they are selling and try to be better than the others, so that they have a chance and survive.

For this to happen, they need an investment in marketing channels to gain market share, customers and with time, the return of the investment. One of this companies, in pursue of a better attribution model, offered their facilities and their data to turn this project a reality. In this way it was possible to gather data from millions of clients and their touch points, using it to simulate multiple attribution models, and see which one is the better for their reality, using this information to help other companies as well.

1.1 Background and problem identification

Companies that rely totally in their sales from a website face one challenge that the others don't: not having a physical store to interact with their audience. This leads to the need of seeking potential new clients that will be engaged by an advertisement and this can come in the form of an e-mail, display, paid search and affiliation. This last form of advertisement can be seen on social networks, youtube, blogger websites and even comparison online tools. This interactions are managed by affiliation networks like Rakuten Marketing (previously

known as LinkShare - EDELMAN and BRANDI, 2015 -).
The major difference between the first three advertisement forms and the affiliate one, is the moment when the company need to pay the advertiser. In the first three cases when someone clicks to enter the website, the company is already paying for this potential costumer. In the fourth case, the advertiser only receives its commission when the costumer makes a purchase (EDELMAN and BRANDI, 2015) but normally the commission paid is higher than it is in the first three cases. For each one of this entrances in the website (and also for direct entrances) a touch point is recorded so that when a purchase is made, there is enough data to understand which channels the costumer used. When a company uses more than one channel, there is the need to examine in which channel the client interacted (Stephen, 2016)in order to redirect the investment to the most important channel as well as to grow the Return of Investment (ROI) value. For this, companies need to adopt an Attribution Model that fits them better, but this can be difficult to choose with little data, which normally leads to a first-click or a last-click attributions (Olson, 2016).With the evolution of the business and with more costumers accessing the website and converting, it can be possible to get a new and better model to take into account more touch points.
The evolution of this model is important to understand if the investment in each of the channels is being accurately spent, and if not, leads to changes in the percentage of marketing budgets for all of this channels and can help increasing sales and attracting even more costumers to the website.

1.1.1 Interaction

An interaction can start with a user that goes directly to a website (writing the website Uniform Resource Locator (URL) in a browser) or from a range of other channels. The most common scenario is that the interaction is saved in some kind of database, with information that comes in the URL that identifies from where the user entered the website, so that a company can understand from where the interaction came (Bucklin and Sismeiro, 2009). In the company were the data were analyzed, this information can be obtain with the combination of four values, designated as referral, utm_source, utm_medium and $utm\_campaign$. With the values analyzed, the company have the exact source from where the user started the interaction, and if this values are not present, the user entered the website directly.
Normally, an interaction occurs when a user have a shopping decision in mind, and with this decision a user can visit multiple websites (Park, 2017) to compare prices, find a specific product size/color or even to find different brands/models. The multiple visits to an website for this or other reasons can be a carryover effect, if the channel of entrance is the same than the previous visit, or a spillover effect, if the channel of entrance is different (Anderl, Becker, von Wangenheim, and Schumann, 2016).
The most common ways to enter a website is through Affiliates, Social Media, Search Engine Advertisement (Search Engine Advertisement (SEA) or Pay per Click (PPC)), Display, Email,

Search Engines and directly through the type in of the URL (Anderl, Schumann, and Kunz, 2016).

1.1.2 Conversion

There are multiple ways to consider a conversion in websites, in certain cases (e.g. blogs, YouTube videos) a conversion can be made if a user clicks in advertiser links or the number of downloads of a certain document (Clifton, 2010, Monetizing a Non-E-commerce Website), and in e-commerce websites cases a conversion occurs when a user makes a purchase, but, in the end, a conversion always stands for the same reason, a user that achieved a pre-determined objective (Clifton, 2010, p.55).

1.1.3 Data Source

There are multiple ways to gather interactions and conversion data from a website, in this case all data is gathered through a pixel implemented in the company's website that send all the information to a database. This approach allows a very detailed information, known as "site-centric" data source (Bucklin and Sismeiro, 2009). This can be a problem, because the company only knows what the user do inside the website, but it gives no information of what the user is doing in other websites (Bucklin and Sismeiro, 2009). In that way, this analysis will only provide inside information, being at this state impossible to determine what the other websites of the competition that an individual user is browsing.

1.1.4 Sale and Non-sale seasons

Just like in physical stores, e-commerce websites tend to have sale and non-sale seasons thought the year. The most important sale seasons in the year are Singles day (with more impact in China), Black Friday and Cyber Monday, three key moments that usually generates more revenue to online sellers (Kumar, 2017; Zakkour, 2017). In this particular e-commerce website, the sales season began in 10th of November of 2017, one day before Singles Day, and finished after Cyber Monday (27th November of 2017).

1.2 Attribution model

Attribution models give credit to each interaction based in an algorithm that tries to see which interactions had more importance in the decision of making a purchase (Con et al., 2016; Olson, 2016). Credit is given to an interaction if it was generated on a thirty day period before the conversion was made.

For single touch attributions, only one interaction is rewarded and many studies already did find that companies should stop using them (Fain, 2018). For multi touch models, all interactions on the user's journey are rewarded (Con et al., 2016; Geyik, Saxena, and Dasdan, 2015).

Table 1.1: Difference between Single and Multiple Touch Attribution and examples of some models. Table adaptation from B2B Marketing – Attribut101n (Con et al., 2016).

Attribution Type	**Attribution Model**	**Reward for interaction**			
		#1(Direct)	#2	#3	#4(Direct)
Single Touch	First Non-Direct	0%	100%	0%	0%
	Last Non-Direct	0%	0%	100%	0%
Multi Touch	Linear	25%	25%	25%	25%
	Time Decay First	40%	30%	20%	10%
	Time Decay Last	10%	20%	30%	40%
	U-Shape	40%	10%	10%	40%

On the next sections the different attribution models will be explained, using Figure 1.1 as an example, having in mind that for all this models, a thirty day period is used for all conversions.

Figure 1.1: Example of a user journey in a website.

1.2.1 First Non-Direct Click

First Non-Direct Click is normally used when a company wants to reward the referrers or ads that gives the website a larger number of audience. In this way, and as shown in Figure 1.2, this attribution model gives the first non-direct entrance one hundred percent of its rewards (Con et al., 2016).

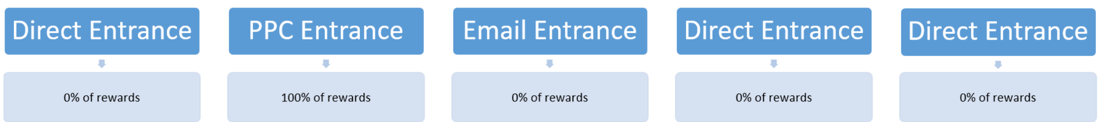

Figure 1.2: Rewards given for each one of the example interactions using First Non-Direct Click.

1.2.2 Last Non-Direct Click

Last Non-Direct Click is normally the default attribution for almost every attribution modelling analyzer (e.g. Google Analytics) and it gives all the rewards to the last interaction that was not a direct one (Con et al., 2016; Fain, 2018). In this way, and as shown in Figure 1.3, this attribution model gives the first non-direct entrance one hundred percent of its rewards.

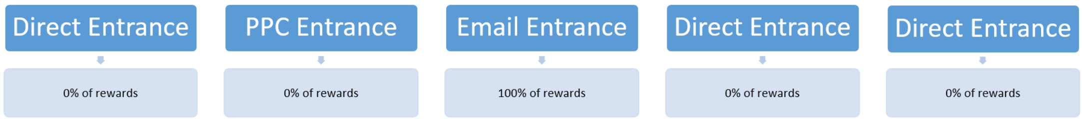

Figure 1.3: Rewards given for each one of the example interactions using Last Non-Direct Click.

1.2.3 Linear

In the Linear model all entrances are treated as same, being all channels rewarded in the same level (Con et al., 2016). This is a fairly simple model and it is explained in the Equation 1.1. The variable x_0 makes reference to the first interaction and $xn - 1$ to the most recent one, where n is the number of interactions in the last thirty days.

$$i_x = \frac{1}{\frac{x_{n-1}}{x_0} 1} \tag{1.1}$$

With this equation in mind, each interaction in Figure 1.4 takes 20% of rewards, having the Direct channel sixty percent of the rewards because three of the entrances in the website had origin in this channel.

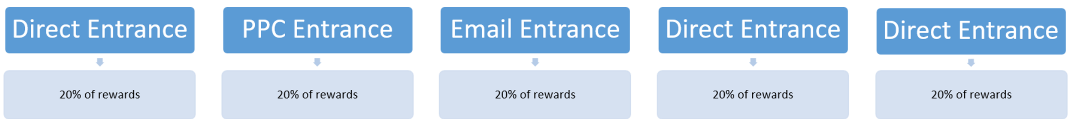

Figure 1.4: Rewards given for each one of the example interactions using Linear model.

1.2.4 Time Decay

In literature, it's more common to have only one Time Decay model, but since this project tries to aim the perfect attribution model to a company, it was decided to include two different ways to measure interaction rewards through this model. It is possible to reward more the most recent interaction, Time Decay Last Click, or the opposite, Time Decay First Click. In the next two sections, it is possible to understand the difference of both of this models.

1.2.4.1 Time Decay First Click

In the Time Decay First Click, it is possible to give more rewards to the First entrance to the website and less to the last. This is an evolution of Linear and First Click attribution models, having a mix of the two in this model. The Equation 2 creates this possibility, with n being the number of interactions in the thirty days before the conversion, x_0 being the first interaction, $xn - 1$ the interaction immediately before the conversion and t the positive number of days before the conversion.

$$i_x = \frac{2^{\frac{t+1}{30}}}{\sum_{x_0}^{x_{n-1}} 2^{\frac{t+1}{30}}} \tag{1.2}$$

With Equation 1.2 the distribution of rewards for this example is as follows in Figure 1.5.

Figure 1.5: Distribution of rewards for example interactions in the Time Decay First Click model.

1.2.4.2 Time Decay Last Click

The algorithm to create the rewards for Time Decay Last Click is the same as the First Click with the difference that in this model, an interaction that occurs at the same day that the conversion is the most rewarded one (Con et al., 2016). The Equation 3 creates this possibility, with n being the number of interactions in the thirty days before the conversion, x_0 the first interaction, $xn - 1$ the interaction immediately before the conversion and t the positive number of days before the conversion.

$$i_x = \frac{2^{-\frac{t+1}{30}}}{\sum_{x_0}^{x_{n-1}} 2^{-\frac{t+1}{30}}} \tag{1.3}$$

Equation 1.3 creates the opposite of the one shown in the last section, and the rewards for each one of the interactions are shown in Figure 1.6.

Figure 1.6: Rewards for each one of the example entrances using Time Decay Last Click.

1.2.5 U-Shape

In the U-Shape attribution model, there is a mixing of the two Time Decay models shown in the previous sections, having the more rewards being given to the most recent or the more distant interactions from the conversion made. Rewards for each channel can be obtain using the following equation, with n being the number of interactions in the thirty days before the conversion, t as the positive number of days preceding the conversion, x_0 as the first entrance of the journey and last interaction as $xn - 1$.

$$i_x = \frac{\frac{t-15}{30}_2 + 0.1}{{}^{x_{n-1}}_{x_0}\ \frac{t-15}{30}_2 + 0.1} \tag{1.4}$$

Figure 1.7: Rewards for each example interaction using U-Shape model.

1.3 Project relevance and importance

There is no Attribution Model that fits all companies, so there's the need to understand which is the best for a specific company. The only way to choose one, is to run a series of simulations with the company data, run them through the equations shown on section above, and then see which is the most effective.
Because the company that asked for this study didn't have any mechanism to capture and store interactions, and to cross reference with the conversion generated, no simulations were done since the beginning of this website, the company have no benchmark from previous simulations. Outputs from this project are the ones that will define future investments on the different channels.

1.4 Research objectives

In this project all interactions in a thirty day period before a conversion will be analysed, using the outputs to compare the most used attribution models and understand which one is the better for the company that the data was gathered. With the analysis of the data collected, there will be run a series of simulation with various scenarios of attribution models, comparing the output of the simulation and understand what would be the best model. With this outputs and with the comparison of results, it is expected to encounter the best attribution model for this company and, if the results found are enough to change the current attribution model, three secondary goals.

1.4.1 Track user journey and conversions

Before this project, the company didn't had the capability to analyse user journey's and to understand which combined channels can increase user retention. With the data generated by the mechanism created for this project, the company can analyse other data that will be ignored in this study, but that are still stored on companies database, available for future studies.

1.4.2 Increase the effectiveness of some channels

With the data collected, it could be possible to understand which of the channels is more capable of getting new costumers that make purchases, increasing the budget for those channels.
Increasing budgets of some channels can lead to future simulations, being capable of creating a snowball effect, always trying to get a better model, and being constantly improving the effectiveness of all the channels (Talbot, 2018).

1.4.3 Find the best attribution model for specific sales season

With the data collected, it could be possible to understand which of the channels is more capable of getting new costumers that make purchases, increasing the budget for those channels.
Increasing budgets of some channels can lead to future simulations, being capable of creating a snowball effect, always trying to get a better model, and being constantly improving the effectiveness of all the channels.

2 Methodology

Through the gathering of all touch points and purchases made in the company's website, this project will analyse conversions from 4th of October and 27th of November of 2017, all simulations will be made with the same inputs so the final result can be trustful. As the default time range to analyse interactions in a conversion is thirty days before until the date of the conversion, all interactions from the 4th of September to 27th of November of 2017 were also gathered.
The company chose the six Attribution Models presented on Table 1.1 to be analysed with the same input values, so they could have a better understand of their user's behaviour and make a better use of their marketing budget.

2.1 Target

All interactions made on the same device than a specific order are accounted by the simulator. This data is divided in two, because of the need to run the simulations independently for sale and non-sale seasons. With the divided output, Attribution Model (AM) can be analysed to understand which one is the better for each of the seasons on this study. For each simulation two variables are accounted, value and number of products for a conversion that are multiplied by the reward value for each interaction.
The data gathered analysed consists in more than two million interactions and almost two hundred and fifty conversions. To be possible to have meaningful data, Flaiz, 2018 says that more than one month should be analysed. In this way, interactions are divided in two different groups, each one with more than forty five days of data. The first one, for conversions made between 4th of October and 9th of November of 2017, will be categorized as Non-sale season, and the second one will be tested in conversions occurred between 10th and 27th of November will be categorized as Sale season.

2.2 Data preparation

On the beginning of January, this company had no internal mechanism to categorize and store interactions or conversions. This was a process made by an external partner, that when receiving a conversion, returned only the First and Last Non-Direct interaction for that conversion. Because many interactions were being lost, the company decided to implement

an in-house mechanism to store this information.
There were three topics that were key on this implementation. The first one was to get a service to store and categorize all interactions from a user or device, the second topic was to create an identifier that could connect one interaction to a purchase, and the last one was to collect and store some properties of all purchases made through the website.
The first topic was achieved by implementing a trigger on the website, that for each page view there were send five different properties for a service: URL for the current page; the URL from which the user entered on the company's website; the browser used to access the page; in which operating system the browser was running; and finally the unique device identifier to be able to cross reference with the future purchase. All the information stored in the database are saved also with the date in which the action was made to be possible to compare between sale and non-sale seasons, and also to use only the thirty day period stated on section 1.2.
All the data gathered were carefully selected due to the incoming General Data Protection Regulation (GDPR) law. In this way the information stored isn't personal data, since the only values used are: price in United States Dollar (USD); currency used to pay; number of products; and the unique device identifier where the purchase was made.

2.3 Data collection

After implementing the mechanism to categorize interactions, it was possible to start collecting all the data needed to start this project. During the first three months (sixty days of interaction in non-sale, thirty days in sale and one month of purchases for each season) there were a replication of data from the production database to a local one, creating the a scenario possible to made all simulations needed.
For this case, the local database had to contain certain properties so that the various simulations could cover all channels in study. For each interaction the database had an URL and a *document.referrer* (a property capable to be returned from a browser to get the previous webpage from where the user got to the current page). The URL is needed to be able to access a series of information from the *query string*, for instance, it can have more meta-data that the page could not be obtained from any other way. Giving an example for one of the most used websites in the world, when someone enters *www.google.com* and searches for "water", the user is redirected to a page with the URL *https://www.google.com/search?q=water*, the query string for this page is "*q=water*". This happens to pass information from the first page (where the used typed "water") to where the search is displayed.
On Table 2.1, the columns marked with an asterisk are some of the fields that can be present on the query string of an URL, if so, the first comparison is made on the *utm_medium* value, if it is equal to the one on the table and the other utm's are present (or not present in the case of being a direct entrance), and the referrer as shown on the table, the categorization can be made. If there is no match, the interaction is considered as *Other*, being not so

Table 2.1: Example of the rule set used to categorise an interaction.

Category	***referrer***	***utm_source****	***utm_medium****	***utm_campaign****
Affiliate	*<not empty>*	blog	affiliate	network2018
Email	*<empty or not>*	website	email	campaign2018
PPC	*<not empty>*	google	cpc	ppc2018
Display	*<not empty>*	advertiser	display	display2018
SEO	*<not empty>*	*<empty>*	*<empty>*	*<empty>*
Social Media	*<not empty>*	facebook	facebook	social2018
Direct	*<empty>*	*<empty>*	*<empty>*	*<empty>*

relevant for the study.

So an entrance on the website with an URL like *www.website.com?utm_source=blog &utm_medium=affiliate&utm_campaign=network2018* and with a *document.referrer* as *network.com* is categorised as Affiliate. If the entrance URL were $www.website.com$ and the *referrer* present in the web page were empty, it would be a Direct interaction, beacuse it as no *utm* or *referrer* fields.

2.3.1 Single touch models

For Last and First Non-Direct models, each conversion only have one associated touch point. The first part is to discard all Direct categories and then link the conversion with the first interaction that occurred prior to the conversion (the last click that was not Direct) and the oldest one (always inside the thirty day period) that will be defined as the First Non-Direct Click. If in the user journey there weren't any interaction different from Direct, the two of them are considered as Direct entrances.

2.3.2 Multi touch models

For multi touch models, there is the need to analyse all interactions in an user journey. This way, and because a journey can contain more than one interaction of the same category, for every touch point with the categorization than another, the results for the equations showed in sections 1.2.3, 1.2.4 and 1.2.5 are summed so that can be possible to analyse the interactions by category.

2.4 Statistical analysis

Statistical analysis could only start after the gathering of all data, separated in number of products, Total value and if the conversion was placed inside of outside a Sale season. Because all data is classified and could not be showed to general public, some modifications were done, but maintaining structural relations between all values so that they could be

compared in the right way.
With the data aggregated in two datasets, one for sale and another for non-sale season, a Linear Regression was run in Statistical Package for Social Sciences (SPSS) for each one of the models in both seasons. SPSS generates a table for each one of the regressions being ANOVA and Coefficients output available in Annex I and in each one of the next sections, summary results are also available for analysis.
On the output generated significance is important to know if the results can be reliable, so only p values lower than $.050$ will be important to this study. A better result will be the larger Unstandardized Coefficient B on the table, being the channel that most impact brings to the studied dependent value.

3 Results

3.1 First Non-Direct click

As stated in section 1.2 this model and Last Non-Direct are the simpler models to implement, because they only needs to consider the first or last interaction to reward a channel for a given conversion, however, at least for this company, it is not the best model to use. For some channels it is better than Last Non-Direct, but it is worst than the most complex models. Looking to ANOVA results on tables I.1, I.3, I.2 and I.4, it is possible to see that

Table 3.1: First Click model summaries for all four regressions.

Season	Dependent Variable	R	R^2	Adjusted R^2	Std Error
Sale	Total Value	.050	.002	.002	1.455
	Number of items	.080	.006	.006	1.752
Non-Sale	Total Value	.063	.004	.004	1.397
	Number of items	.062	.004	.004	1.406

this models have great significance, with $p < .001$ in all four regressions. On table 3.2 both

Table 3.2: Beta Unstandardized Coefficients for First Non-Direct Click simulations.

	Total Value		**Number of items**	
	Sale	**Non-Sale**	**Sale**	**Non-Sale**
(Constant)	**1.113***	**1.162***	**1.813***	**1.614***
Affiliates	**-.141***	**-.146***	**-.246***	**-.210***
Direct	**-.308***	**-.148***	**-.330***	**-.119***
SEO	**-.146***	**-.232***	**-.221***	**-.192***
Display	**-.160***	**-.309***	**-.113***	**-.108***
Social Media	**-.202***	**-.153***	**-.196***	**-.125***
PPC	**-.159***	**-.183***	**-.208***	**-.144***
Email	.005	.025	**.113***	**.067***

Values marked with * have good significance.

sale and non-sale season Email channel as a higher significance ($p = .808$ for sale season and $p = .319$ for non-sale), which shows that it is not a reliable variable to look at, and full coefficients output is available in annex on tables I.5, I.6, I.7 and I.8.

On sale season two regression equations were found ($F7, 141391 = 50.610, p < .001$), with

a $R^2 = .002$ for total value and ($F7, 141391 = 128.505, p < .001$), with a $R^2 = .006$ for the dependent variable number of items. Conversion predicted impact is equal to $1.113 + -.141$ (Affiliates interactions) total value and equal to $1.813 + .113$ (Email interactions) for number of items bought. The average total value decreased $.141$ points for each conversion that started with an Affiliate interaction, being the channel that had minor losses and on number of items Email interactions increased $.113$ points being the only channel were there results above zero.
For non-sale season the regression equation found was ($F7, 81834 = 47.260, p < .001$), with a $R^2 = .006$ for total value and ($F7, 81834 = 44.446, p < .001$), with a $R^2 = .004$. Conversion predicted impact is equals to $1.162 + -.146$ (Affiliates interactions) of total value and equal to $1.614 + .067$ (Email interactions) for number of items bought. The average total value decreased $.146$ for each conversion that started with an Affiliate interaction, being the channel that had minor losses and on number of items Email interactions increased $.113$ points being the only channel were the results are above zero.

3.2 Last Non-Direct click

ANOVA results for this model can be reviewed on tables I.9, I.11, I.10 and I.12, where just like in the previous model, all output results have significance, with $p < .001$ in all four regressions. Joining ANOVA and results from table 3.3, two equations for sale season

Table 3.3: Last Click model summaries for all four regressions.

Season	Dependent Variable	R	R^2	Adjusted R^2	Std Error
Sale	Total Value	.085	.007	.007	1.751
	Number of items	.053	.003	.003	1.752
Non-Sale	Total Value	.086	.003	.003	1.397
	Number of items	.063	.004	.004	1.405

can be found, ($F7, 141391 = 50.148, p < .001$), with a $R^2 = .007$ for total value and ($F7, 141391 = 146.927, p < .001$), with a $R^2 = .003$. For non-sale season the equation ($F7, 81834 = 40.237, p < .001$), with $R^2 = .003$ is the one found for total value, and for number of items has an equation of ($F7, 81834 = 46.699, p < .001$), with $R^2 = .004$. Analysing table 3.4, and for total value dependent variable results, Email channel cannot be accounted because it's significance value is $p > .150$, all other values have $p < .001$ giving them enough significance to be studied, full coefficients are available in annex on tables I.13, I.14, I.15 and I.16.
The channels that most impact create on a conversion are Affiliates with $1.114 + -.117$ of predicted impact on total value in sale season, and Direct channel with $1.157 + -.129$ of impact on the same dependent variable but in non-sale season. This two values are negative but on the study, this are the values that are closest to zero. On the other variable in study,

Table 3.4: Beta Unstandardized Coefficients for Last Non-Direct Click simulations.

	Total Value		**Number of items**	
	Sale	**Non-Sale**	**Sale**	**Non-Sale**
(Constant)	**1.114***	**1.157***	**1.813***	**1.612***
Affiliates	**-.117***	**-.154***	**-.223***	**-.214***
Direct	**-.333***	**-.129***	**-.346***	**-.116***
SEO	**-.158***	**-.219***	**-.242***	**-.181***
Display	**-.176***	**-.297***	**-.200***	**-.136***
Social Media	**-.292***	**-.212***	**-.276***	**-.117***
PPC	**-.168***	**-.174***	**-.229***	**-.146***
Email	.015	.006	**.103***	**.071***

Values marked with * have good significance.

the channel with best impact is Email with the only positive values in this regressions, $1.813 + .103$ on sale season, and $1.612 + .071$ impact on non-sale.

3.3 Linear

On Linear model, ANOVA present on tables I.17, I.19, I.18 and I.20 shows that for this model have significant results with $p < .001$. When looking also to table 3.5, the two

Table 3.5: Linear model summaries for all four regressions.

Season	**Dependent Variable**	**R**	R^2	**Adjusted R^2**	**Std Error**
Sale	Total Value	.064	.004	.004	1.454
	Number of items	.098	.010	.010	1.749
Non-Sale	Total Value	.086	.007	.007	1.395
	Number of items	.084	.007	.007	1.403

equations on sale season ($F7, 141391 = 84.111, p < .001$), with a $R^2 = .004$, ($F7, 141391 = 197.494, p < .001$), with a $R^2 = .010$ respectively for total value and number of items. On the second season, total value has an equation of ($F7, 81834 = 87.929, p < .001$), with a $R^2 = .007$, and the equation ($F7, 81834 = 83.692, p < .001$), with a $R^2 = .007$ was found for dependent variable number of items. On table 3.6 simulation outputs have no values with 3 significance, so all channels can be analysed. Full outputs are on annex tables I.21, I.22, I.23 and I.24. The channels that cause the most impact are Direct for total value with $.993 + .202$ (sale) and $1.093 + .170$ (non-sale season) and Email for number of items with $1.700 + .294$ and $1.570 + .171$ for value and number of items respectively. For non-sale season, and dependent variable total value, Email channel has only a difference of $.013$, being very close to the one that has the most impact.

Table 3.6: Beta Unstandardized Coefficients for Linear simulations.

	Total Value		**Number of items**	
	Sale	**Non-Sale**	**Sale**	**Non-Sale**
(Constant)	**.993 ***	**1.093***	**1.700***	**1.570***
Affiliates	**-.048***	**-.129***	**-.211***	**-.238***
Direct	**.202 ***	**.170 ***	**.221 ***	**.166 ***
SEO	**-.090***	**-.220***	**-.201***	**-.201***
Display	**-.138***	**-.392***	**-.116***	**-.127***
Social Media	**-.219***	**-.116***	**-.224***	**-.118***
PPC	**-.073***	**-.148***	**-.161***	**-.135***
Email	**.137***	**.157***	**.294***	**.171***

Values marked with * have good significance.

3.4 Time Decay First Click

When analysing this model's ANOVA results on tables I.25, I.27, I.26 and I.28 it is possible to see a $p < .001$, which indicates a great significance for the results taken. Table 3.7,

Table 3.7: Time Decay First Click model summaries for all four regressions.

Season	**Dependent Variable**	**R**	R^2	**Adjusted** R^2	**Std Error**
Sale	Total Value	.064	.004	.004	1.454
	Number of items	.098	.010	.009	1.749
Non-Sale	Total Value	.086	.007	.007	1.395
	Number of items	.084	.007	.007	1.403

shows the rest of the data needed to generate the equations on both seasons. For sale, ($F7, 141391 = 83.316, p < .001$), with a $R^2 = .004$ when analysing for total value and ($F7, 141391 = 193.910, p < .001$), with a $R^2 = .010$ for number of items dependent variable. On non-sale, total value has an equation of ($F7, 81834 = 87.953, p < .001$), with a $R^2 = .007$, an on number of items variable the equation ($F7, 81834 = 83.396, p < .001$), with a $R^2 = .007$ was found. With outputs visible on 3.8, it is possible to see that all values have low significance, being possible to analyse all channels on the four simulations. Full data is available on I.29, I.30, I.31 and I.32.

On this simulations Direct channel has a greater impact when the dependent variable is total value, with $.993 + .204$ and $1.093 + .171$ for the two seasons on study. When looking for number of items, Email interactions are the ones that have better gains, with $1.700 + .293$ and $1.570 + .170$, but on the second value (non-sale season) Direct impact was almost the same, with a difference of only $.002$.

Table 3.8: Beta Unstandardized Coefficients for Time Decay First Click simulations.

	Total Value		**Number of items**	
	Sale	**Non-Sale**	**Sale**	**Non-Sale**
(Constant)	**.993 ***	**1.093***	**1.700***	**1.570***
Affiliates	**-.048***	**-.129***	**-.209***	**-.237***
Direct	**.204***	**.171***	**.221***	**.168***
SEO	**-.089***	**-.220***	**-.196***	**-.200***
Display	**-.135***	**-.392***	**-.107***	**-.125***
Social Media	**-.214***	**-.110***	**-.220***	**-.113***
PPC	**-.071***	**-.148***	**-.158***	**-.133***
Email	**.138***	**.157***	**.293***	**.170***

Values marked with * have good significance.

3.5 Time Decay Last Click

ANOVA results present on tables I.33, I.35, I.34 and I.36 show great significance on the results presented on this section, with $p < .001$. Being this model a mirror of Time Decay First

Table 3.9: Time Decay Last Click model summaries for all four regressions.

Season	**Dependent Variable**	**R**	R^2	**Adjusted** R^2	**Std Error**
Sale	Total Value	.065	.004	.004	1.454
	Number of items	.099	.010	.010	1.749
Non-Sale	Total Value	.086	.007	.007	1.395
	Number of items	.084	.007	.007	1.403

Click, being the summary values being the table 3.9 very similar to the ones showed on the previous section. With this five tables, sale equations are ($F7, 141391 = 84.510, p < .001$), with a $R^2 = .004$ for total value and ($F7, 141391 = 200.243, p < .001$), with a $R^2 = .010$ for number of items. On the other season in study equations are ($F7, 81834 = 87.309, p < .001$), with a $R^2 = .007$ to total value and ($F7, 81834 = 83.638, p < .001$), with a $R^2 = .007$ for the total of items present in a conversion. The results on table 3.10 shows that all values have great significance, having Email and Direct channels the only with positive impact on both variables in study, full information on the simulations are available on tables I.37, I.38, I.39 and I.40.

Email interactions have more impact ($1.700 + .293$ for sale and $1.571 + .170$ for non-sale) when number of items variable is used. For total value Direct has a greater result than the others with an impact of $.993 + .199$ for sale season and $1.092 + .166$ on non-sale. In both total value and number of products, Email and Direct values have little impact difference on non-sale season, with $.010$ and $.008$ respectively.

Table 3.10: Beta Unstandardized Coefficients for Time Decay Last Click simulations.

	Total Value		Number of items	
	Sale	Non-Sale	Sale	Non-Sale
(Constant)	**.993 ***	**1.092***	**1.700***	**1.571***
Affiliates	**-.047***	**-.128***	**-.212***	**-.239***
Direct	**.199***	**.166***	**.219***	**.162***
SEO	**-.092***	**-.220***	**-.204***	**-.200***
Display	**-.139***	**-.390***	**-.123***	**-.129***
Social Media	**-.223***	**-.121***	**-.227***	**-.123***
PPC	**-.075***	**-.148***	**-.163***	**-.137***
Email	**.135***	**.156***	**.293***	**.170***

Values marked with * have good significance.

3.6 U-Shape

For the last model in study, ANOVA results present on tables I.41, I.43, I.42 and I.44 show significance of $p < .001$ on the results presented on this section. Through table 3.11 and

Table 3.11: U Shape model summaries for all four regressions.

Season	Dependent Variable	R	R^2	Adjusted R^2	Std Error
Sale	Total Value	.063	.004	.004	1.454
	Number of items	.095	.009	.009	1.749
Non-Sale	Total Value	.086	.007	.007	1.395
	Number of items	.083	.007	.007	1.403

ANOVA results, sale season equations ($F7, 141391 = 80.342, p < .001$), with a $R^2 = .004$ for total value and ($F7, 141391 = 183.458, p < .001$), with a $R^2 = .009$ for number of items. For non-sale, $F7, 81834 = 86.675, p < .001$), with a $R^2 = .007$ was found for total value and ($F7, 81834 = 81.684, p < .001$ was the equation found for number of items. Table 3.12

Table 3.12: Beta Unstandardized Coefficients for U Shape simulations.

	Total Value		Number of items	
	Sale	Non-Sale	Sale	Non-Sale
(Constant)	**.993***	**1.094***	**1.700***	**1.568***
Affiliates	**-.049***	**-.129***	**-.204***	**-.231***
Direct	**.203***	**.172***	**.217***	**.169***
SEO	**-.082***	**-.216***	**-.183***	**-.197***
Display	**-.123***	**-.382***	**-.086***	**-.120***
Social Media	**-.199***	**-.096***	**-.209***	**-.102***
PPC	**-.065***	**-.147***	**-.150***	**-.128***
Email	**.139***	**.151***	**.288***	**.165***

Values marked with * have good significance.

shows part of the coefficient values for simulations made for this last model in study, that can be found in annex on tables I.45, I.46, I.47 and I.48.
U shape model has Direct and Email as the channels that most impact create on total value and items bought on a conversion. Direct gets more impact when looking to total value with $.993 + .203$ and $1.094 + .172$ for sale and non-sale, but in the last one with only a difference of 0.011 for Email, and for number of items in non-sale Direct also gets in front with more $.004$ than Email results, having a total of $1.568 + .169$. Analysing number of items on sale, Email channel has an impact of $1.700 + .288$.

4 Conclusions

After a year of study, being nine months to generate the data-gathering algorithm and more than one quarter of data collection, compiling order creation, user's interaction with the website and simulating results for the different models that the company wanted to be studied there is the need to present the conclusions of which was the best model.
On this chapter are present the main conclusions, possible challenges, limitations and future work needed if this or any other company would want to have this study reproduced.

4.1 Main conclusions

The main goal for this work was to find the best attribution model for one specific company that during months provided all the data needed for this study to be concluded. Through the use of linear regressions on the two datasets divided in seasons, as mentioned on 2.4, all models were statistically tested on SPSS program.
After analysing if all outputs had enough significance to be considered, all summaries, coefficients and ANOVA were examined and it was concluded that the channels with most impact on at least one model are Direct, Affiliates and Email, being the last one present in all models studied.

Table 4.1: Number of best results for each model studied.

	Affiliates	Direct	SEO	Display	Social Media	PPC	Email
First Click	0/1	0/0	0/0	0/0	0/0	0/0	0/0
Last Click	0/0	0/0	0/1	0/0	0/0	0/0	0/0
Linear	0/0	0/0	0/0	0/0	0/0	0/0	1/2
Time Decay First	0/0	2/0	0/0	0/0	0/0	0/0	0/1
Time Decay Last	1/1	0/0	0/0	0/0	0/0	0/0	0/0
U Shape	1/0	0/2	2/1	2/2	2/2	2/2	1/0

Values shown are divided in "Sale/Non-sale".

In table 4.1 is possible to see best results on more complex models, since First and Last non-direct have only two best results in this study. The last four models get all the rest of the best results, but U Shape model gets almost the double of the best results of the other complex models with fifteen versus eight. During section 1.2, it was stated that First

and Last click are simpler models due to the level of information that they need to process. Through this study it is possible to understand that this models have less capability to see the possibilities in each of the channels and that the company need to evolve to more complex ones. This is also what Forbes states, that "multi-touch combination of factors and exposures helps or detracts from each customer's likelihood to convert"(Fain, 2018).
Comparing Multi touch models, U Shape is clearly the best one, getting fifteen best results against only three in Linear and Time Decay First Click. U Shape model gets perfect score for Display, Social Media and PPC channels, having best results for both variables and seasons. Search Engine Optimization (SEO) channel also gets good output values, getting the best results for Sale, and one for Non-sale. Affiliates and Email channels get both one best result for sale season and Direct gets a perfect result for non-sale.
Just like Forbes states, "more and more companies are focused on including all marketing touchpoints in their analytics and attribution modeling"(Nichols, 2018), this company too should start focusing in all of its user's journey.
Other luxury companies should also do this work, start (or continue) to analyse all interactions in an user's journey. On this study two main variables were analysed, but results for both total value and number of products had almost the same results, so probably other companies can only look to one of this variables.
Using the results on this project, it is also possible that other ecommerce websites can benefit from this study, since all of the most complex models were whe ones with the best results of the study, being the only ones with two channels with positive rewards. Besides this, companies gain also more control of their user's journey and can perform other kinds of study, like doing A/B tests on different channels and partners, or even other kind of models like Market-Mix that can boost effectiveness and subsequently ROI of each channel (McDonald, 2018).

4.2 Main challenges

One of the main challenges of doing this work was the need to create all the mechanism to get and store all informations related to conversions and page views. This kind of project had the need to create public Application Programming Interface (API) to handle all the information that needed to be stored (more than fifteen thousand page views every minute) and correlate those to all purchases made on the website, and a service that received all of this data, analysed all interactions and stored only the ones that were the first page on a users journey.
Three months after all the mechanism to correlate interactions with purchases were made and all the data were being saved correctly on the database, started the time to begin the simulations. This was the time were the company defined what they wanted to see analysed, which models to run against and how the categorization of interactions would be made. After almost one month of trial and error the categorization were ready and the simulations could start.

4.3 Study limitations

As stated on 4.2 this company didn't had detailed information on user journeys or page views on the website. In this way, and due to time restrictions, some features were left behind during the eight months of developing the mechanism to support this project. One of this features was the Universal user identifier, that could identify all page views of a user that visits the website from multiple devices. Because this feature was not built, all the conversions used on this project have a journey limited to the device where the purchase occurred which can impact the data and simulations made on this project.
On the simulations side, much more information could have been used for other purposes (country of the user, device and browser used by each user, currency in which the purchase was made, etc...) but because of the new GDPR and some more internal rules, the company didn't allow the use of this variables.
Due to confidential data from the business side, this project does not count with budget differences between channels. Results got from this simulations need to be reviewed by the company's business side to understand if some results might be discarded due to low budgeting or if they just need to be used in other ways and to different kinds of audience.
After all the analysis made and when writing this projects main conclusions, an issue on the Direct categorization was found by a team on the company that gave the access to the data. This could have led to a greater impact on models that gave better rewards to clicks on the initial part of the journey. This was measured internally and on the conversions studied on this project less than five percent had this issue.

4.4 Future work

This project had very specific objectives, that were centred mainly on increasing the investment done on the major channels that contribute to a larger number of sales. With the dataset available there could be done various types of other studies, many of them using the power of Neural Networks or other types of deep learning. With this new searching types it is possible to learn much more from the user side such as behaviour or even costumes for each kind of country or continent.
With the conclusions taken from this project the Attribution Model with the best results will be monitored from the business team and can be implemented when the company makes that decision, but because marketing budgets are already defined for the next year, this changes can only enter on the budget for 2020.
Even if the company decides not to implement this new model right after publishing this project, data analysis will continue so that in the future better models can be applied if necessary.

Bibliography

Anderl, E., Becker, I., von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing*, *33*(3), 457–474. doi:10.1016/j.ijresmar.2016.03.001

Anderl, E., Schumann, J. H., & Kunz, W. (2016). Helping Firms Reduce Complexity in Multichannel Online Data: A New Taxonomy-Based Approach for Customer Journeys. *Journal of Retailing*, *92*(2), 185–203. doi:10.1016/j.jretai.2015.10.001

Baron, N. (2008). *Always On: Language in an Online and Mobile World* (1st). Oxford University Press.

Bucklin, R. E., & Sismeiro, C. (2009). Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing. *Journal of Interactive Marketing*, *23*(1), 35–48. doi:10.1016/j.intmar.2008.10.004

Clifton, B. (2010). *Advanced Web Metrics with Google Analytics* (3rd ed.). doi:10.1017/CBO9781107415324.004. arXiv: arXiv:1011.1669v3

Con, J., Rigotti, D., Nguyen, A., Frye, L., Getscher, A., Hylbak, L., & Sharf, E. (2016). *B2B marketing - Attribut101n*. Bizible, Inc. Retrieved from http://www.bizible.com/hubfs/B2B-Marketing-Attribution-101-ebook.pdf

EDELMAN, B., & BRANDI, W. (2015). Risk, Information, and Incentives in Online Affiliate Marketing. *Journal of Marketing Research (JMR)*, *52*(1), 1–12. doi:10.1509/jmr.13.0472

Fain, J. (2018). Why Marketers Need To Stop Focusing On Last-Touch Attribution. Retrieved September 25, 2018, from https://www.forbes.com/sites/forbesagencycouncil/2018/06/28/why-marketers-need-to-stop-focusing-on-last-touch-attribution/%7B%5C#%7D3b93f5d846ad

Flaiz, W. (2018). Do You Know What Is Really Driving Your Sales? Retrieved September 25, 2018, from https://www.forbes.com/sites/forbescommunicationscouncil/2018/09/20/do-you-know-what-is-really-driving-your-sales/%7B%5C#%7D37216eed6e43

Geyik, S. C., Saxena, A., & Dasdan, A. (2015). Multi-Touch Attribution Based Budget Allocation in Online Advertising. *Multi-Touch Attribution Based Budget Allocation in Online Advertising*. doi:10.1145/2648584.2648586. arXiv: 1502.06657

ITU World Communications; ICT Indicators Database. (2017). *Global ICT Developments*. International Telecommunications Union (ITU). Retrieved from https://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2017/ITU%7B%5C_%7DKey%7B%5C_%7D2005-2017%7B%5C_%7DICT%7B%5C_%7Ddata.xls

Kumar, B. (2017). Black Friday Cyber Monday 2017: An Analysis of Over $1 Billion in Sales. Retrieved January 20, 2018, from https://www.shopify.com/blog/black-friday-cyber-monday-2017-recap

Marinucci, J. (2018). Four Realities That Will Rock The World Of Marketing In 2018. Retrieved September 25, 2018, from https://www.forbes.com/sites/forbesagencycouncil/2018/04/12/four-realities-that-will-rock-the-world-of-marketing-in-2018/%7B%5C#%7D55f6e51067ba

McDonald, S. (2018). Measuring The ROI Of Marketing: A/B Tests Vs. Market-Mix Models Vs. Multi-Touch Attribution. Retrieved September 25, 2018, from https://www.forbes.com/sites/scottmcdonald1/2018/01/23/measuring-the-roi-of-marketing-ab-tests-vs-market-mix-models-vs-multi-touch-attribution/%7B%5C#%7D410acbe62576

Nichols, J. (2018). How To Prove Your Partner Marketing Drives True Incremental Sales. Retrieved September 25, 2018, from https://www.forbes.com/sites/forbescommunicationscouncil/2018/04/02/how-to-prove-your-partner-marketing-drives-true-incremental-sales/%7B%5C#%7D2a9554ca6203

Olson, C. (2016). Rethinking today's attribution problem in digital marketing. Retrieved October 21, 2017, from https://searchengineland.com/rethinking-todays-attribution-problem-260767

Park, C. H. (2017). Online Purchase Paths and Conversion Dynamics across Multiple Websites. *Journal of Retailing*, *93*(3), 253–265. doi:10.1016/j.jretai.2017.04.001

Schmidt, J., Dörner, K., Berg, A., Schumacher, T., & Bockholdt, K. (2015). The opportunity in online luxury fashion. Retrieved July 12, 2017, from http://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/the-opportunity-in-online-luxury-fashion

Smith, A., McGeeney, K., Duggan, M., Rainie, L., & Keeter, S. (2015). The Smartphone Difference. In *Pew research center* (pp. 1–60). Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Stephen, A. T. (2016). The role of digital and social media marketing in consumer behavior. *Current Opinion in Psychology*, *10*, 17–21. doi:10.1016/j.copsyc.2015.10.016. arXiv: arXiv:1011.1669v3

Talbot, P. (2018). Multi-Touch Attribution Modeling: A More Revealing Look At The Customer Journey. Retrieved September 25, 2018, from https://www.forbes.com/sites/paultalbot/2018/06/08/multi-touch-attribution-modeling-a-more-revealing-look-at-the-customer-journey/%7B%5C#%7D57507c882191

Zakkour, M. (2017). Singles' Day 2017: Alibaba's Vision For The Future Of Retail. Retrieved January 20, 2018, from https://www.forbes.com/sites/helenwang/2017/11/12/alibabas-singles-day-by-the-numbers-a-record-25-billion-haul/%7B%5C#%7D6bac20581db1

I Annex 1 - SPSS Linear regression results

Table I.1: First Click ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	749.747	7	107.107	50.610	.000
Residual	299226.741	141391	2.116		
Total	299976.488	141398			

Table I.2: First Click ANOVA results with Number of items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	2760.461	7	394.352	128.505	.000
Residual	433896.188	141391	3.069		
Total	436656.649	141398			

Table I.3: First Click ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	645.661	7	92.237	47.260	.000
Residual	159716.544	81834	1.952		
Total	160362.205	81841			

Table I.4: First Click ANOVA results with Number of items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	614.677	7	87.811	44.446	.000
Residual	161677.033	81834	1.976		
Total	162291.710	81841			

Table I.5: First Click coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.113	.016		67.695	.000
Affiliates	-.141	.019	-.034	-7.282	.000
Direct	-.308	.037	-.025	-8.389	.000
SEO	-.146	.018	-.044	-8.052	.000
Display	-.160	.027	-.020	-5.960	.000
Social Media	-.202	.030	-.021	-6.673	.000
PPC	-.159	.018	-.048	-8.779	.000
Email	.005	.019	.001	.243	.808

Table I.6: First Click coefficient results with Number of items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.813	.020		91.564	.000
Affiliates	-.246	.023	-.049	-10.586	.000
Direct	-.330	.044	-.022	-7.456	.000
SEO	-.221	.022	-.055	-10.104	.000
Display	-.113	.032	-.012	-3.508	.000
Social Media	-.196	.037	-.017	-5.370	.000
PPC	-.208	.022	-.052	-9.548	.000
Email	.113	.022	.026	5.089	.000

Table I.7: First Click coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.162	.021		56.601	.000
Affiliates	-.146	.025	-.035	-5.951	.000
Direct	-.148	.050	-.011	-2.939	.003
SEO	-.232	.022	-.076	-10.363	.000
Display	-.309	.034	-.040	-9.218	.000
Social Media	-.153	.039	-.016	-3.966	.000
PPC	-.183	.022	-.061	-8.247	.000
Email	.025	.025	.006	.996	.319

Table I.8: First Click coefficient results with Number of items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.614	.021		78.150	.000
Affiliates	-.210	.025	-.051	-8.507	.000
Direct	-.119	.051	-.009	-2.344	.019
SEO	-.192	.023	-.063	-8.536	.000
Display	-.108	.034	-.014	-3.200	.001
Social Media	-.125	.039	-.013	-3.225	.001
PPC	-.144	.022	-.048	-6.417	.000
Email	.067	.025	.015	2.671	.008

Table I.9: Last Click ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	831.554	7	118.793	50.148	.000
Residual	299144.934	141391	2.116		
Total	299976.488	141398			

Table I.10: Last Click ANOVA results with Number of Items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	3153.329	7	450.476	146.927	.000
Residual	433503.320	141391	3.066		
Total	436656.649	141398			

Table I.11: Last Click ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	550.050	7	78.579	40.237	.000
Residual	159812.155	81834	1.953		
Total	160362.205	81841			

Table I.12: Last Click ANOVA results with Number of Items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	645.706	7	92.244	46.699	.000
Residual	161646.004	81834	1.975		
Total	162291.710	81841			

Table I.13: Last Click coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.114	.016		68.250	.000
Affiliates	-.117	.019	-.031	-6.289	.000
Direct	-.333	.038	-.026	-8.855	.000
SEO	-.158	.018	-.044	-8.565	.000
Display	-.176	.026	-.022	-6.695	.000
Social Media	-.292	.029	-.033	-10.206	.000
PPC	-.168	.018	-.048	-9.227	.000
Email	-.015	.018	-.004	-.825	.410

Table I.14: Last Click coefficient results with Number of Items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.813	.020		92.213	.000
Affiliates	-.223	.022	-.049	-9.959	.000
Direct	-.346	.045	-.022	-7.642	.000
SEO	-.242	.022	-.056	-10.914	.000
Display	-.200	.032	-.021	-6.341	.000
Social Media	-.276	.034	-.025	-7.994	.000
PPC	-.229	.022	-.055	-10.417	.000
Email	.103	.022	.025	4.732	.000

Table I.15: Last Click coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.157	.020		56.764	.000
Affiliates	-.154	.024	-.041	-6.494	.000
Direct	-.129	.052	-.009	-2.478	.013
SEO	-.219	.022	-.070	-9.778	.000
Display	-.297	.032	-.042	-9.427	.000
Social Media	-.212	.036	-.025	-5.914	.000
PPC	-.174	.022	-.057	-7.829	.000
Email	.006	.025	.001	.254	.799

Table I.16: Last Click coefficient results with Number of Items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.612	.021		78.638	.000
Affiliates	-.214	.024	-.056	-8.999	.000
Direct	-.116	.052	-.008	-2.222	.026
SEO	-.181	.022	-.058	-8.049	.000
Display	-.136	.032	-.019	-4.308	.000
Social Media	-.117	.036	-.014	-3.258	.001
PPC	-.146	.022	-.047	-6.527	.000
Email	.071	.025	.017	2.868	.004

Table I.17: Linear ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1243.982	7	117.712	84.111	.000
Residual	298732.506	141391	2.113		
Total	299976.488	141398			

Table I.18: Linear ANOVA results with Number of Items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	4228.094	7	604.013	197.494	.000
Residual	432428.556	141391	3.058		
Total	436656.649	141398			

Table I.19: Linear ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1197.131	7	171.019	87.929	.000
Residual	159165.074	81834	1.945		
Total	160362.205	81841			

Table I.20: Linear ANOVA results with Number of Items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1153.581	7	164.797	83.692	.000
Residual	161138.129	81834	1.969		
Total	162291.710	81841			

Table I.21: Linear coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	.993	.020		50.703	.000
Affiliates	-.048	.023	-.009	-2.072	.038
Direct	.202	.024	.035	8.428	.000
SEO	-.090	.022	-.019	-4.045	.000
Display	-.138	.035	-.012	-3.900	.000
Social Media	-.219	.037	-.018	-5.921	.000
PPC	-.073	.022	-.017	-3.334	.001
Email	.137	.022	.031	6.211	.000

Table I.22: Linear coefficient results with Number of Items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.700	.024		72.126	.000
Affiliates	-.211	.028	-.035	-7.615	.000
Direct	.221	.029	.032	7.672	.000
SEO	-.201	.027	-.035	-7.466	.000
Display	-.116	.043	-.008	-2.720	.007
Social Media	-.224	.045	-.015	-5.028	.000
PPC	-.161	.026	-.030	-6.074	.000
Email	.294	.027	.056	11.074	.000

Table I.23: Linear coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.093	.025		43.998	.000
Affiliates	-.129	.029	-.026	-4.364	.000
Direct	.170	.031	.030	5.524	.000
SEO	-.220	.027	-.055	-8.047	.000
Display	-.392	.043	-.039	-9.216	.000
Social Media	-.116	.047	-.010	-2.483	.013
PPC	-.148	.027	-.038	-5.444	.000
Email	.157	.031	.028	5.119	.000

Table I.24: Linear coefficient results with Number of Items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.570	.025		62.845	.000
Affiliates	-.238	.030	-.047	-8.036	.000
Direct	.166	.031	.029	5.353	.000
SEO	-.201	.027	-.050	-7.299	.000
Display	-.127	.043	-.012	-2.975	.003
Social Media	-.118	.047	-.010	-2.514	.012
PPC	-.135	.027	-.034	-4.946	.000
Email	.171	.031	.030	5.508	.000

Table I.25: Time Decay First Click ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1232.260	7	176.037	83.316	.000
Residual	298744.229	141391	2.113		
Total	299976.488	141398			

Table I.26: Time Decay First Click ANOVA results with Number of Items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	4152.089	7	593.156	193.910	.000
Residual	432504.560	141391	3.059		
Total	436656.649	141398			

Table I.27: Time Decay First Click ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1197.468	7	171.067	87.953	.000
Residual	159164.737	81834	1.945		
Total	160362.205	81841			

Table I.28: Time Decay First Click ANOVA results with Number of Items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1149.162	7	164.166	83.396	.000
Residual	161142.548	81834	1.969		
Total	162291.710	81841			

Table I.29: Time Decay First Click coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	.993	.020		50.720	.000
Affiliates	-.048	.023	-.010	-2.100	.036
Direct	.204	.024	.035	8.470	.000
SEO	-.089	.022	-.019	-3.973	.000
Display	-.135	.035	-.012	-3.840	.000
Social Media	-.214	.037	-.018	-5.783	.000
PPC	-.071	.022	-.016	-3.237	.001
Email	.138	.022	.031	6.263	.000

Table I.30: Time Decay First Click coefficient results with Number of items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.700	.024		72.161	.000
Affiliates	-.209	.028	-.034	-7.556	.000
Direct	.221	.029	.031	7.653	.000
SEO	-.196	.027	-.035	-7.317	.000
Display	-.107	.042	-.008	-2.523	.012
Social Media	-.220	.044	-.015	-4.942	.000
PPC	-.158	.026	-.030	-5.991	.000
Email	.293	.027	.055	11.054	.000

Table I.31: Time Decay First Click coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.093	.025		44.067	.000
Affiliates	-.129	.029	-.026	-4.382	.000
Direct	.171	.031	.030	5.565	.000
SEO	-.220	.027	-.055	-8.048	.000
Display	-.392	.043	-.039	-9.207	.000
Social Media	-.110	.047	-.010	-2.355	.019
PPC	-.148	.027	-.038	-5.457	.000
Email	.157	.031	.028	5.113	.000

Table I.32: Time Decay First Click coefficient results with Number of items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.570	.025		62.892	.000
Affiliates	-.237	.030	-.047	-7.984	.000
Direct	.168	.031	.029	5.414	.000
SEO	-.200	.027	-.050	-7.296	.000
Display	-.125	.043	-.012	-2.926	.003
Social Media	-.113	.047	-.010	-2.404	.016
PPC	-.133	.027	-.034	-4.879	.000
Email	.170	.031	.030	5.506	.000

Table I.33: Time Decay Last Click ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1249.856	7	178.551	84.510	.000
Residual	298726.632	141391	2.113		
Total	299976.488	141398			

Table I.34: Time Decay Last Click ANOVA results with Number of Items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	4286.365	7	612.338	200.243	.000
Residual	432370.285	141391	3.058		
Total	436656.649	141398			

Table I.35: Time Decay Last Click ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1188.756	7	169.822	87.309	.000
Residual	159173.449	81834	1.945		
Total	160362.205	81841			

Table I.36: Time Decay Last Click ANOVA results with Number of Items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1152.839	7	164.691	83.638	.000
Residual	161138.871	81834	1.969		
Total	162291.710	81841			

Table I.37: Time Decay Last Click coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Standard Error	Beta		
(Constant)	.993	.020		50.689	.000
Affiliates	-.047	.023	-.009	-2.026	.043
Direct	.199	.024	.035	8.330	.000
SEO	-.092	.022	-.019	-4.094	.000
Display	-.139	.035	-.012	-3.926	.000
Social Media	-.223	.037	-.019	-6.035	.000
PPC	-.075	.022	-.017	-3.418	.001
Email	.135	.022	.031	6.151	.000

Table I.38: Time Decay Last Click coefficient results with Number of Items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Standard Error	Beta		
(Constant)	1.700	.024		72.097	.000
Affiliates	-.212	.028	-.035	-7.652	.000
Direct	.219	.029	.032	7.636	.000
SEO	-.204	.027	-.036	-7.580	.000
Display	-.123	.043	-.009	-2.887	.004
Social Media	-.227	.044	-.016	-5.101	.000
PPC	-.163	.026	-.031	-6.136	.000
Email	.293	.027	.056	11.055	.000

Table I.39: Time Decay Last Click coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Standard Error	Beta		
(Constant)	1.092	.025		43.940	.000
Affiliates	-.128	.029	-.026	-4.334	.000
Direct	.166	.031	.029	5.427	.000
SEO	-.220	.027	-.055	-8.029	.000
Display	-.390	.042	-.038	-9.178	.000
Social Media	-.121	.046	-.011	-2.607	.009
PPC	-.148	.027	-.038	-5.425	.000
Email	.156	.031	.028	5.075	.000

Table I.40: Time Decay Last Click coefficient results with Number of Items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.571	.025		62.812	.000
Affiliates	-.239	.030	-.048	-8.071	.000
Direct	.162	.031	.028	5.244	.000
SEO	-.200	.028	-.050	-7.280	.000
Display	-.129	.043	-.013	-3.026	.002
Social Media	-.123	.047	-.011	-2.629	.009
PPC	-.137	.027	-.035	-5.006	.000
Email	.170	.031	.030	5.478	.000

Table I.41: U Shape ANOVA results with Total value as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1188.459	7	169.780	80.342	.000
Residual	298788.029	141391	2.113		
Total	299976.488	141398			

Table I.42: U Shape ANOVA results with Number of Items as dependent variable - Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	3930.032	7	561.472	183.458	.000
Residual	432726.348	141391	3.060		
Total	436656.649	141398			

Table I.43: U Shape ANOVA results with Total value as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1180.188	7	168.598	86.675	.000
Residual	159182.016	81834	1.945		
Total	160362.205	81841			

Table I.44: U Shape ANOVA results with Number of Items as dependent variable - Non-Sale season.

Model	Sum of squares	df	Mean Square	F	Sig.
Regression	1126.032	7	160.870	81.684	.000
Residual	161165.622	81834	1.969		
Total	162291.710	81841			

Table I.45: U Shape coefficient results with Total value as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	.993	.020		50.770	.000
Affiliates	-.049	.023	-.010	-2.117	.034
Direct	.203	.024	.035	8.440	.000
SEO	-.082	.022	-.018	-3.721	.000
Display	-.123	.034	-.011	-3.575	.000
Social Media	-.199	.037	-.017	-5.432	.000
PPC	-.065	.022	-.015	-2.981	.003
Email	.139	.022	.032	6.300	.000

Table I.46: U Shape coefficient results with Number of Items as dependent variable - Sale season.

	Unstandardized Coefficients		Standardized Coefficients		
	B	Standard Error	Beta	t	Sig.
(Constant)	1.700	.024		72.250	.000
Affiliates	-.204	.028	-.034	-7.365	.000
Direct	.217	.029	.030	7.466	.000
SEO	-.183	.027	-.033	-6.865	.000
Display	-.086	.041	-.007	-2.079	.038
Social Media	-.209	.044	-.015	-4.742	.000
PPC	-.150	.026	-.029	-5.705	.000
Email	.288	.027	.055	10.856	.000

Table I.47: U Shape coefficient results with Total value as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Standard Error	Beta		
(Constant)	1.094	.025		44.222	.000
Affiliates	-.129	.029	-.026	-4.381	.000
Direct	.172	.031	.030	5.583	.000
SEO	-.216	.027	-.056	-7.974	.000
Display	-.382	.042	-.038	-9.052	.000
Social Media	-.096	.046	-.008	-2.079	.038
PPC	-.147	.027	-.038	-5.444	.000
Email	.151	.030	.028	4.948	.000

Table I.48: U Shape coefficient results with Number of Items as dependent variable - Non-Sale season.

	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Standard Error	Beta		
(Constant)	1.568	.025		63.031	.000
Affiliates	-.231	.030	-.046	-7.817	.000
Direct	.169	.031	.029	5.459	.000
SEO	-.197	.027	-.050	-7.234	.000
Display	-.120	.042	-.012	-2.832	.005
Social Media	-.102	.047	-.009	-2.190	.029
PPC	-.128	.027	-.033	-4.724	.000
Email	.165	.031	.030	5.365	.000

2018 The Impact of Attribution Modelling in Luxury E-Commerce Filipe Chaves MGI