Multi-Touch Marketing Attribution Methods on Online Customer Journey

Empirical Comparison between Different Attribution Models on Real Life Data of a E-commerce Website

Student name: Jiayue Liu

Student ID number: 557540jl

Supervisor: Kathrin Gruber

Second assessor: M van Combrugge

Date final version: July 2022

Abstract

This thesis applies the traditional methods (First click, the Last click, and the Linear attribution models) and the Markov chain model to online customer journeys on the individual level. Empirical results show a significant difference between traditional and algorithmic models and give a practical insight into how the Markov chain model can better assist multi-channel online attribution. Empirical results also differ in varying lengths of customer journeys. The conversion funnel and customer decision process help interpret such differences. In conclusion, the traditional methods mostly reflect the channel occurrences and positions, and such methods understate Instagram. Combined with Instagram's higher attribution in a longer path and observed spillover, the underestimated value lies in incremental value in moving customers down the decision process and passing them to other channels. Longer paths allow the channels that foster states like brand awareness to show importance. Facebook and Online video are also such channels. Paid search oppositely fosters stages like alternative evaluation, explaining its higher credit in shorter paths. Online video shows high carryovers. Anticipating such effects help better targeting. The transition matrix can assist automation in anticipating and pushing efficient channels to customers. The empirical results mostly confirm extant literature. Significant differences in empirical results alert marketers to be careful to adopt generalization as heterogeneity among industry and customers may cause substantial differences in channel attribution.

List of Figures

List of Tables

Contents

1 Introduction

1.1 Research question

Decades ago, prints media and broadcast media were major means for an average person to know the outside world. In recent years, digital media (or new media, online media) has become the main source of information and has gained massive influence on people as a result of the development of the Internet and mobile technologies (Turow 2010). In E-commerce, online media has been largely used to impress an audience and produce the desired effect. In essence, it works as such: Separately, each channel impacts in a different fashion, while conjointly, a fine orchestration between each should be pursued to achieve a larger impact than the sum of its parts (Hennig-Thurau et al. 2010; Abhishek, P. Fader, and Hosanagar 2012). This thesis follows the common definition of '**channel**' - the medium of interactions between the companies and their potential customers, and the definition of '**touchpoint**' - the interaction instances between these companies and customers, according to Anderl et al. (2016). Moreover, the **online customer journey** of a particular customer is also defined as the collection of all the touchpoints he or she has visited on online channels before he or she headed to the webstore, for a potential purchase (Anderl et al. 2016; Salazar 2016).

Naturally, the pursuit of creating a profound marketing effect remarks the importance of knowing the actual channel contribution. However, in today's climate, consumers often get impressions by dazzling arrays of touchpoints on multiple channels. Not exclusively mentioned, it can be search ads, display ads, affiliated or non-affiliated ads, placed on a search engine, emails, social media or web shops (OrbitalAds. 2021). **Figure** 1 shows a typical online customer journey. The **customer Gina** wanted to buy some dumbbells. She began her journey by viewing a promoting post of dumbbells on her Instagram Story. After this, she viewed ads for dumbbells in Display, Paid search, and Video forms chronologically throughout the next week. The series of touchpoints impressions made her conduct an Organic search on her own. She finally found one website which provided discounts if she opted in the Email list. Finally, after reading a promotion Email and viewing another Display and Social media ads, she went to the webstore and made a purchase. And the same story could happen to another consumer Jim who did not purchase at the end of the journey. In the real world, complexity is exhibited in investigating the real contribution of channels behind the many touchpoints that make up a long customer journey (Abhishek, P. Fader, and Hosanagar 2012). It is similar to asking a conductor of a large-sized orchestra to tell which artist of his band has contributed the most to an impressive performance. These artists are all playing a distinct instrument in their time to play, while only the collection of the sounds they make will determine the joy of the performance.

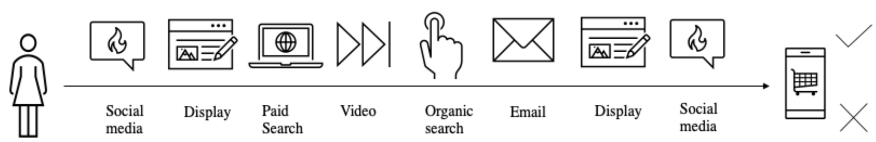


Figure 1: An example of a multi-touch online customer journey

The importance and complexity we attach to knowing the actual channel contribution has naturally led to the topic 'marketing attribution', which is often defined as the objective of assigning credit to one or more advertising channels for their impacts on a conversion (Shao and L. Li 2011). In this line, empirical research is conducted on a particular dataset where different methods (which will be further introduced in Chapter 2) were applied, in order to:

- Compare the results of several commonly used marketing attribution methods,

And

- Test the appropriateness for businesses to use generalized marketing attribution insights as their budget figures.

The motivation behind these is explained in the next chapter.

1.2 Motivation

My motivation mainly consists of two points. One is to enrich the empirical records of the application of algorithmic models. Two points trigger this, the first being the **lack** of knowledge and fewer adoption of algorithmic attribution models in the business. During the lecture on New Media Analytics at ESE Erasmus University Rotterdam, the founder of PR agency Mr. Oskam has shown that monopolies (e.g., Google and Facebook) are keen to take advantage of this inadequacy to make profits from it. They do so by providing their platform for marketers to advertise and charge them for it. For example, Google provides bidding plans for words, which are the words marketers want to include in their advertisements. In today's climate, it is impossible for marketers using these similar services to see if this charge is **fair** or not (Oskam 2021). Perhaps an even more crucial question for marketers is - by paying passively the cost per Google platform, do they have a chance to see their actual returns from the channels and to optimize their budget allocation (Bryl' 2016). According to Oskam's experience, the chance is rather slim.

Besides the point above, the aim for enriching empirical records is also triggered by the wide **gap** in applications between simpler and algorithmic models. Researchers often deem the algorithmic models as more accurate than the others. However, many companies have yet to taste the fruits of it, as today, largely used are still the simpler ones (Lukmani 2021). Thus, this research aims to present empirical evidence to those managers who are not familiar with algorithmic models to be better informed and make wiser decisions.

In addition to providing empirical data for **widening** the acceptance of algorithmic models, this research also tries to discover if **generalized** business insights on marketing attribution are a decent mid-way solution for companies without their own budget solution. As we know in this era, information is largely accessible in all fields. Should this be the case in E-commerce, then following a commonly used strategy is way more encouraged than fabricating one's own solution, for the latter pursuit is mostly considered costly and tedious, especially for companies with limited capacity and experience such as small-to-medium enterprises (SME). In fact, in E-commerce, this **insider information** might be accessible and sounding, as the one Oskam (2021) has shared during the class New Media Analysis at ESE EUR. According to Oskam, the percentage of traffic per channel for a general E-commerce business is as in **Figure** 2.

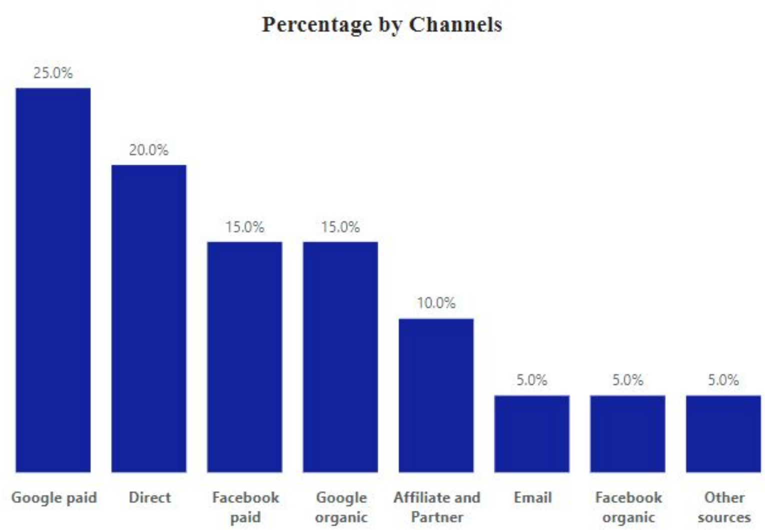


Figure 2: Percentage by channel

Useful this common knowledge may seem, however, the **difference** from business to business is still distinct enough for marketers to beware of, especially when the lengths of customer journeys vary largely across businesses (Oskam 2021; Abhishek, P. Fader, and Hosanagar 2012). Several studies confirm that ignoring the difference per industry, and blindly following a common sense budgeting can be detrimental for a company. One of the reasons is that with various lengths of customer journeys, the impact of the same channel can be vastly **distinctive** (Wolny and Charoensuksai 2014; Bryl' 2016; Abhishek, P. Fader, and Hosanagar 2012).

2 Literature review

2.1 Types of attribution methods

"**Marketing attribution**" is often defined as the objective of assigning credit to one or multiple advertising channels for their influence on a desirable action of customers (e.g., buying, subscribing) (Shao and L. Li 2011). A good marketing attribution strategy is equipped with accurate information on the efficacy of specific channels in influencing the desirable action, i.e., the knowledge of a channel's added value. The research consensus is that a **good** attribution model should closely mirror a company's online marketing paradigms and help understand the customer journey, in contrast to a **bad** attribution model, which twists the reality and leads to suboptimal budgeting (Singal et al. 2022). However, as customer journeys often range across multiple channels (Facebook, Instagram, Email, Google, etc.) and are much longer than before, this **complexity** makes it difficult to discover how different channels impact the conversions (Lukmani 2021). Thus, the multi-channel attribution problem, under today's climate, calls for more adequate attribution methods.

2.1.1 Simple heuristic methods

Marketing attribution models are mostly classified into types as simplistic, heuristic, and algorithmic models (Moffett 2012). Simplistic models represent a set of rudimentary attribution methods (Lee 2010). These models give complete conversion credit to a single touchpoint without looking at each channel's value. The most frequently used **simplistic** models are First touch and Last touch models. First touch model assigns all conversion credit to the first touchpoint, under the premise that the first one is solely credible for the final conversion. This premise is not as logical since a customer often interacts with more than one channel before purchasing the product, which makes the nonsense of this method felt. Such simple attribution therefore may prompt managers to discard the valuable subsequent channels, same as the Last touch model, which assigns all credit to the last touchpoint, leading up to overseeing all the precedent channels' value, which also undermines the conversion and revenue.

Heuristic models are slightly more complex than the simplistic ones, yet also have some ingrained limitations. Heuristic models assign credit to multiple touchpoints based on fixed rules. These fixed rules are made by heuristic rules-of-thumbs. Linear, Position-Based, Time Decay, U-shaped and W-shaped models are the typical heuristic models. In **Linear** attribution, credit is assigned evenly among all touchpoints. Hence, this over-simplicity neglects real contribution and leads to indifferent and unjust budget allocation. In Position-Based attribution, more credit is given to the **first** and the **last** touchpoint, whereas the left credit is evenly assigned to the others. In **Time Decay** attribution, more credit is assigned to touchpoints closer to a conversion, as these channels are assumed by this method to have higher efficacy in attaining and converting customers. The Time Decay model seems more logical and frequently used than the other ones, as it attempts to give "fairer" credits to the channels rather than evenly assigned. In the same line, **U-Shaped** and **W-shaped** models also frequently used due to better soundness. With the U-shaped model, most credits were assigned on the first and the last touchpoints, which are the two tops of the U, whereas gradually decreased credits are assigned to the other channels in between, where the lowest credit finds the most middle touchpoint. This model is based on the idea that the first and the last channel have the largest impact on the conversion, as the first channel attracts customers and the last channel makes them convert, whereas the more a touchpoint approaches to the middle, the less the conversion has to rely on it, and hence the most middle-positioned touchpoint gets the least credit (Brown 2020). With a slightly different assumption, the W-Shaped model assigns the most credits to the first, middle, and last touchpoint while the others get decreased credits according to their closeness to these three peaks, just as the W shape depicts.

2.1.2 Markov chain models in the marketing context

According to Anderl et al. (2016), a few **algorithmic** attribution models have been substantially developed, for example logit models (Shao and L. Li 2011), models based on game theory, mostly using Shappley value (Berman 2018; Dalessandro et al. 2012), and Bayesian models (H. Li and Kannan 2014). However, this thesis sets focus and hence reviews primarily about the **Markov chain models**, on how the researchers model the customer journeys through a Markovian model and what findings they have.

Anderl et al. (2016) provides a **Markovian graph-based** model, for three research objectives, which are to investigate the attribution of a channel, to discover the interplay among the channels on the same customer journey (e.g., carryover and spillover effects, explained in next paragraph) using the Markovian property (removal effect), and to generalize these answers. By analyzing the individual level of customer journeys in four real-life data from three businesses, they provide **generalizable** insights, including the cross-company and the company-specific ones. A base (first-order) Markov chain model and a **higher-order** (maximum four-order) Markov chain are applied to capture the sequential nature of customer journeys better. Comparing Markov chains against the heuristics (the First and the Last click method) and two logistic regression models shows that the higher-order Markov chain model has higher predictive accuracy and robustness than the other models.

Using sophisticated models, researchers including Anderl et al. (2016) can detect better interplay between the channels. The idiosyncratic channel preferences (i.e., the positive effect that a channel has for itself in the later interactions down the customer journey) are called the **carryover** effect (Anderl et al. 2016). The **spillover** effect occurs when the previous visit of a channel improves the performance of the other channels on the customer journey (H. Li and Kannan 2014). H. Li and Kannan (2014) discovered that most online channels present significant carryover effects. Anderl et al. (2016) identified that spillover effects are more present in customers-initiated channels than in firm-initiated channels. In their scope, **customers-initiated** channels mostly involve searching, such as Search Engine Advertising (SEA), Search Engine Optimization (SEO) search, Price comparison, and Type-ins (i.e., typing in the URL of the website). **Firm-initiated** channels are, for instance, Display, Social media, Retargeting and Newsletter (Anderl et al. 2016).

In Abhishek, P. Fader, and Hosanagar (2012)'s research, a dynamic **hidden Markov model** is applied to capture a customer's moving through different stages on his or her customer journey. The researchers describe this moving as deliberation process, which is captured in the concept of '**conversion funnel**'. The HMM they developed assigns attribution to channels based on their incremental values in increasing a customer's probability to convert, for instance, **Display** is credited relatively high, mostly for its ability to move a customer from a state of disengagement to a state of having product awareness. By analyzing large data of a car-launch online campaign, they provide empirical insights into channels' effectiveness in different customer journey stages. For instance, some channels are incredibly influential at an early stage (e.g., Display), and some are of impact across all phases (e.g., Search ads).

Conversion funnel is a central concept in the marketing literature for a long time. It describes the different stages of a buyer's journey before an eventual purchase. The funnel metaphor depicts the decreasing number of potential buyers moving down the conversion path. The exposure to channels along the buyer's journey is expected to help guide the customers moving down through the conversion funnel, approaching the purchase. Some researchers break it down to **AIDA 'Attention - Interest - Desire - Action'** as in **Figure** 3 (Abhishek, P. Fader, and Hosanagar 2012; Bruce, Peters, and Naik 2012; Parment, Kotler, and Armstrong 2011), or **'Awareness - Consideration - Purchase', or 'Need Recognition - Information Searches - Alternative Evaluation - Purchase - Post Purchase'** which are similar depending on the context (Abhishek, P. Fader, and Hosanagar 2012; Bruce, Peters, and Naik 2012; Jansen and Schuster 2011; Mulpuru 2011; Wolny and Charoensuksai 2014).

Figure 3: The conversion funnel with AIDA levels

Source: `https://www.abtasty.com/blog/ecommerce-conversion-funnel/`

However, in practice, the state of a customer is **latent**, and the progression is hidden, and these are only inferrable from the customer's webstore visits or conversions. The latent states are hence modelled by **HMMs** in a line of extant research (Abhishek, P. Fader, and Hosanagar 2012; Schwartz et al. 2011; Schweidel, Bradlow, and P. S. Fader 2011; Montoya, Netzer, and Jedidi 2010).

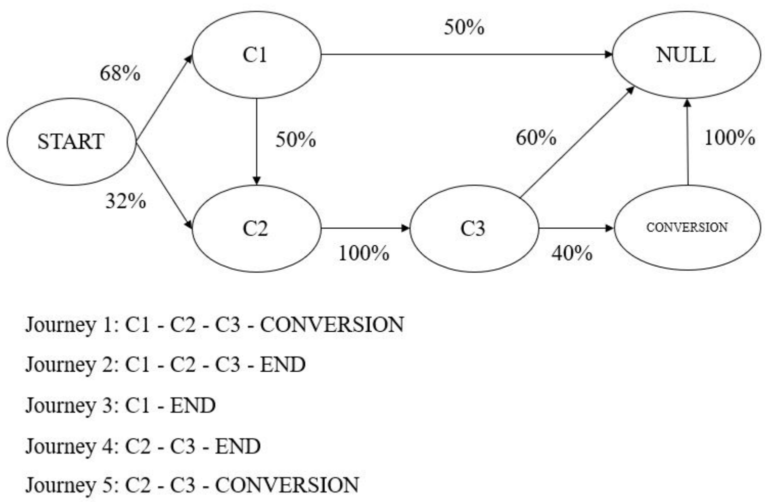
3 Methodology

3.1 Foundational theory & removal effect

In the last chapter, some traditional attribution methods have been introduced. In addition to them, also mainly used by this thesis is an algorithmic attribution method - the Markov chain model. It **describes** a sequence of possible events, among which the chance for each event to happen depends and only depends on the status of the previous event. This is one of the 'Markov properties', which makes it much easier to tackle **complicated** problems, such as online customer journeys.

The Markov chain model is a stochastic model that can be represented either by a **transition matrix** or a **graph**. An exemplary graph is shown by **Figure** 4. Note that the START represents the start of a customer journey, C_i denotes channels along the path, CONVERSION remarks a customer conversion, and NULL represents the end of the customer journey. When the C_i directly transits to the NULL without a CONVERSION beforehand, it denotes an ending journey without a conversion; otherwise, there will be a CONVERSION before the NULL.

Figure 4: An exemplary graph of Markov chain



The transition matrix between the states are as below, each column represents the starting states, denoted by s_i $(i = 1, 2, 3, 4, 5)$, each row represents the arriving states, denoted by s_j, and w_{ij} represents the probabilities from state i to state j accordingly. Note apart from $s_{1...3}$ denoting C1, C2, C3 on the graph, the states CONVERSION and NULL have also been included as s_4 and

s_5 respectively.

$$\begin{bmatrix} w_{11} & w_{21} & w_{31} & w_{41} & w_{51} \\ w_{12} & w_{22} & w_{32} & w_{42} & w_{52} \\ w_{13} & w_{23} & w_{33} & w_{43} & w_{53} \\ w_{14} & w_{24} & w_{34} & w_{44} & w_{54} \\ w_{15} & w_{25} & w_{35} & w_{45} & w_{55} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.4 & 0 & 0 \\ 0.5 & 0 & 0.6 & 1 & 1 \end{bmatrix} \tag{1}$$

Another 'Markov property' is that the sum of the edge weights of the outgoing arrows on any state in a Markov graph is always **one**. Correspondingly, the sum of each column of the transition matrix equals one, as these are transition **probabilities** from the same state assumed to sum up to one. For example, on **Figure** 4, the sum of weights on the outgoing arrows from C3 to NULL (60%) and from C3 to CONVERSION (40%) is one, same as the sum of w_{35} and w_{34} on the transition matrix on **Expression** 1.

To summarize, mathematically, any customer journey can be represented by a Markov graph $M = \langle S, W \rangle$ defined by a set of states

$$S = \left\{ s_1, ..., s_n \right\}, \tag{2}$$

and a transition matrix W with edge weights

$$w_{ij} = P\left(X_t = s_j | X_{t-1} = s_i \right), 0 \leq w_{ij} \leq 1, \sum_{j-1}^{N} w_{ij} = 1 \forall i, \tag{3}$$

where

- w_{ij} is the transition probability of the previous state (X_{t-1}) transiting to the current state (X_t),
- the transition probability (w_{ij}) must be between 0 and 1, and
- the sum of all transition probabilities stemming from a state must equal to 1.

If a Markov model has **absorbing** states, we can easily calculate the long-term equilibria of the model. A state is an absorbing state if and only if it is impossible to leave this state, for example, the NULL state on **Figure** 4. Any state which is not absorbing is called a transient state.

Attribution using the concept of Markov chains is calculated by **Removal effect**, which in essence, is the change of the probability of converting when a state s_i is removed. By removing this state, the sequences after this state is canceled out, turning to a direct linage to the absorbing state NULL. This way, counterfactual data based on the real data is simulated (Anderl et al. 2016). According to Anderl et al. (2016), the Removal effect(s_i) can be efficiently derived by matrix multiplication, more specifically, the **multiplication** of Visits(s_i) and Eventual conversions(s_i) or by applying local algorithms invented by Archak et al. (2010). For example, for **Figure** 4, the removal effects is presented in **Table** 1. Five customer journeys are illustrated on the graph. For example, the removal effect of C1 is calculated by multiplying the probability of Visits of 0.68 and the **eventual conversion probability** of 0.2, which gives a removal effect of 0.14, taking up to 21.21% of the sum of removal effects of all the channels. This is lower than the removal effects of C2 and C3, 0.26 (39.39%), due to the lower eventual conversion probability, which is 0.2 led by C1. C2 and C3 have the same removal effects due to the **transition probability** between these two being 100%, inferring that removing one means removing them both. Removal effects(s_i) can range from zero to the whole model's total conversion rate. In this thesis, for the ease of comparing with traditional models, following (Anderl et al. 2016), I chose to represent removal effects as the **percentages** they take up among the sum of removal effects of all states, where the context-based states START, CONVERSION and NULL are not included. Noteworthy is that when it comes to the two-order model, in this thesis, the removal effect of channel n is calculated by the **average** removal effect of all states having channel n as the last observed one, following the practice of research including Anderl et al. (2016).

Table 1: Removal effects for the Markov graph **Figure** 2

Channel	Visit(s_i)	Ultimate conversion(s_i)	Removal effects(s_i)	in %
C1	0.68	0.2	0.14	21.21%
C2	0.66	0.4	0.26	39.39%
C3	0.66	0.4	0.26	39.39%

3.2 Higher order

Although a strict Markov chain (a first-order Markov chain) assumes that the chance of each event happening depends only on the status obtained in the previous event, prior research suggest that click-streams should **not** be considered as strictly Markovian (Anderl et al. 2016; Chierichetti et al. 2012; Montgomery et al. 2004). Thus, following the previous study by Anderl et al. (2016), I adopted a Markov chain model with higher orders, meaning that the transition probability of transiting from the present event depends on the **last k observations**. For example, if $k = 2$, as applied in this thesis, the transition probability of transiting from the state s_i into the state s_j can depend on the observation of its two previous states - s_i and s_h. This way, transition probabilities can be translated into such:

$$\begin{aligned} P\Bigg(X_t = s_t | X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, ..., X_1 = s_1\Bigg) \\ = P\Bigg(X_t = s_t | X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, ..., X_{t-k} = s_{t-k}\Bigg). \end{aligned} \tag{4}$$

Following Anderl et al. (2016), a k-order Markov chain with a set of states over some alphabet A can be simplified as a first-order Markov chain, ranging across the alphabet A_k of k-tuples, which enables us to employ the same algorithm. However, Anderl et al. (2016) confirmed that as the order increases, the independent parameters also increase **exponentially**, making the real-world data too cumbersome to estimate. Thus, I decided to apply a model with a maximum order of two in this thesis for efficiency. As mentioned in the last subsection, the removal effect of channel n is calculated by the average removal effect of all states where channel n is the last observed one, following the practice of Anderl et al. (2016).

4 Data

In order to give marketers more empirical evidence of using available attribution methods on different customer journeys, this research takes a large number of Internet cookies from customers who visited one or multiple online channels prior to a purchase or ending the journey without a purchase. In this project, data is captured by a data scientist who is also interested in a similar topic. It records the **cookies** of customers of a particular E-commerce website from **1st July 2018 to 31st July 2018**. The cookies carry important **elements** including the cookie codes, the time of cookie generation, the interaction, the channel, and the conversion status. In this dataset, these channels are Instagram, Facebook, Paid search, Online display, and Online videos. For analysis, some manipulation has been done to gather all interactions under one customer journey together, so that each row of the data contains information about all the channels a customer has visited chronologically, until ending the journey, where the conversion status is recorded.

252464 customer journeys are recorded in this dataset. The indicator of a customer journey is a unique cookie code. On each customer journey, one or more touchpoints find their places. If we use the number of touchpoints on a customer journey as the measure of its **length**, then the length of the customer journeys in this dataset varies from 1 to 135, among which 75% are under 3 channels as in **Table** 2. Thus, it has an **imbalanced** structure where shorter customer journeys are the majority.

Table 2: Quantile of customer journeys

Quantile	0% - 50%	75%	100%
Number of touchpoints	1	3	136

As mentioned, there are five channels in the data **- Instagram, Facebook, Paid search, Online display, and Online videos**, whose occurrences are summarized in **Table** 3.

Table 3: Percentage of customer journeys containing channel

Instagram	Facebook	Paid search	Online display	Online video
21.2%	37.3%	36.5%	17.6%	17.6%

According to the table above, in each customer journey, the most common touchpoints to visit are Facebook and Paid search, meaning that at least **one-third** of the customers **have seen** ads on Facebook or Paid search, where the third one is Instagram. When we count the number of visits, Facebook and Paid search are the **most frequently visited** channels, which is similar to the last metric, however, Online video this time takes the third place, as in **Table** 4.

Table 4: Percentage of visits on channel

Instagram	Facebook	Paid search	Online display	Online video
13.2%	31.3%	24%	11%	20.4%

In addition to the occurrences and frequencies, it is also interesting to note that the customer journeys on which Online video is placed have the highest **conversion rate** than the others, with Instagram taking the second place, as in **Table** 5.

Table 5: Conversion rate of a customer journey when it contains channel

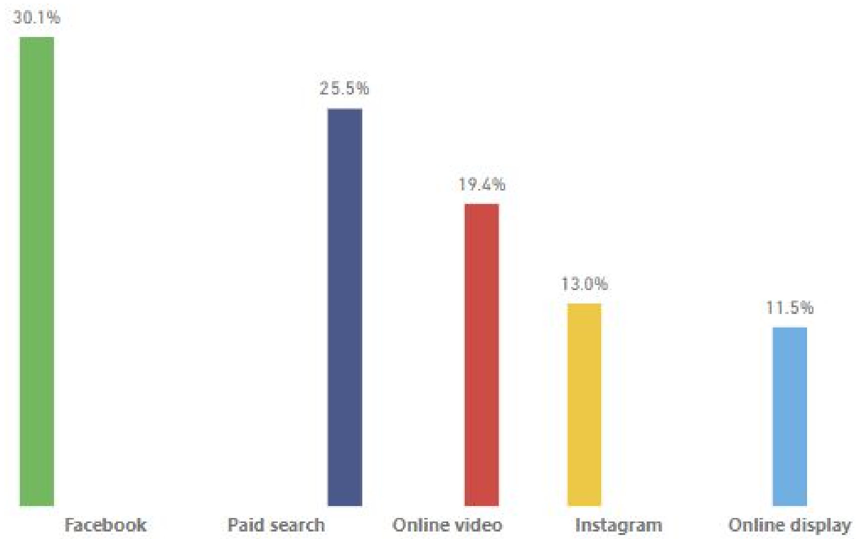
Instagram	Facebook	Paid search	Online display	Online video
10.1%	9%	6.9%	7%	10.4%

5 Result

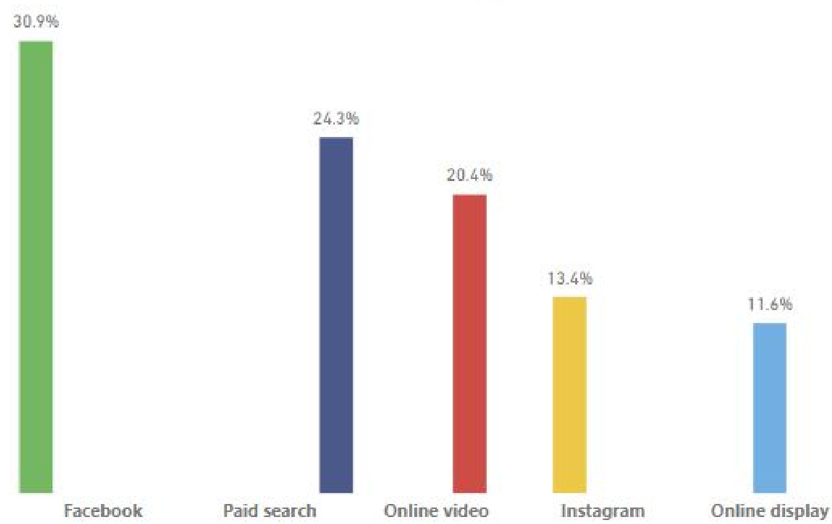
5.1 Method comparison

As shown in **Figure** 5, the First click, the Last click, and the Linear attribution give **similar** results. Despite of some differences in the percentages per channel, the order of channels with the credit from the most to the least amount is the same across the traditional methods, which is **Facebook, Paid search, Online video, Instagram, Online display.** In comparison, the Markov chain model gives a different ranking as output, which is **Facebook, Paid search, Instagram, Online video, Online display.**

First click attribution per channel



Last click attribution per channel



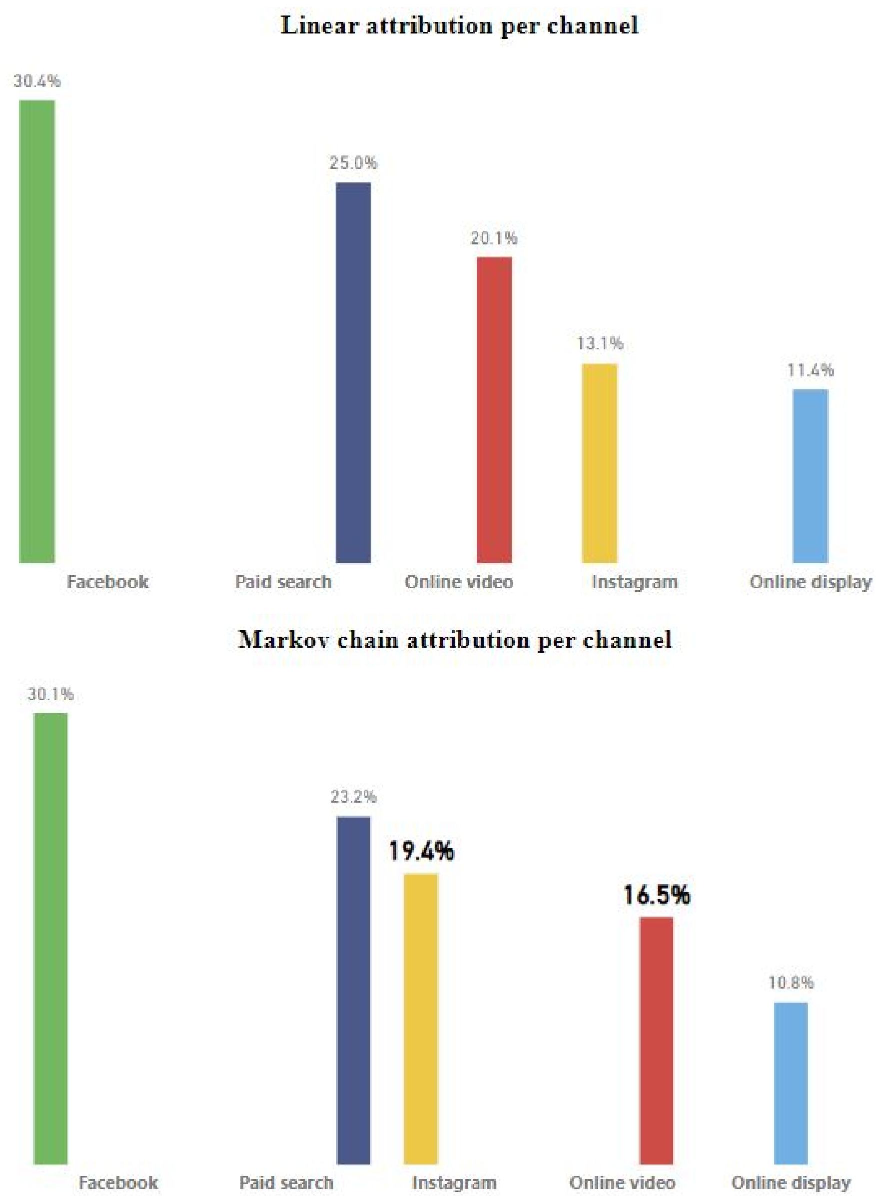


Figure 5: Channel attribution per four methods

The reasons behind above mentioned attribution can be better understood by looking into the **compositions** of the customer journeys, including the positions (**Figure** 6) and the number of visits (**Figure** 7) of the channels.

The **First** click method assigns the credit based on the occurrences of channels being the **first** touchpoint on the customer journeys. It is noteworthy that in this project I calculate the First and the Last click attribution, by **eliminating** the journeys which only consist of two touchpoints, as the channels that appear in the single-channel or the two-channel journeys are not technically the first or the last channel when there is no other channels in-between. Thus, counting their occurrences is not what the First click (and the Last click) attribution method aims for, but more what the Linear attribution would do. Based on this choice, I summarized the occurrences of channels being the first channel in **Figure** 6, and this observation fully explains the results shown in **Figure** 5: the channel appears more on the First clicks (in customer journeys more and equal to three channels) decides its high ranking in the attribution by the First click attribution method. The ranking is **Facebook, Paid search, Online video, Instagram, and Online display**. The **Last** click method follows the same logic as the First click. The attribution depends on the number of showing up on the Last touchpoint (**Figure** 6), which gives the result: **Facebook, Paid search, Online video, Instagram, Online display** (**Figure** 5).

As of **Linear** attribution, as the conversions are assigned evenly across every channel, the credit **solely** relies on the number of the occurrences of this channel. As **Figure** 6 outlines, the number of occurrences (visits on this channel) from high to low is **Facebook, Paid search, Online video, Online display**.

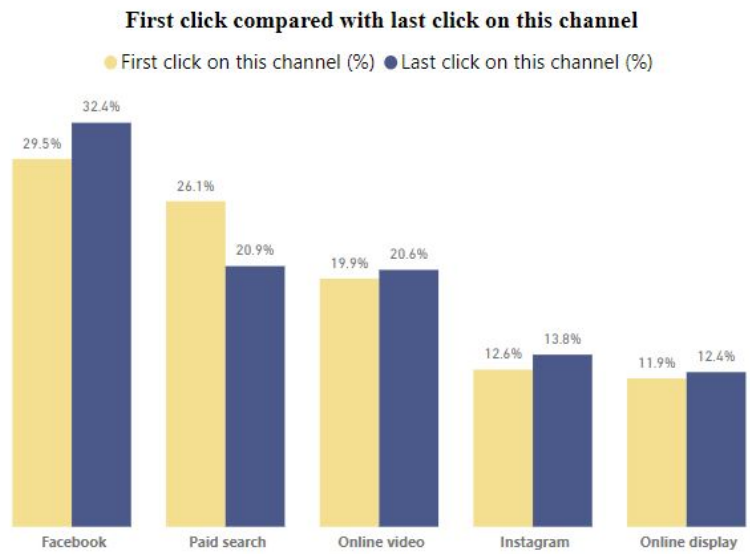


Figure 6: First click compared with last click on this channel (%)

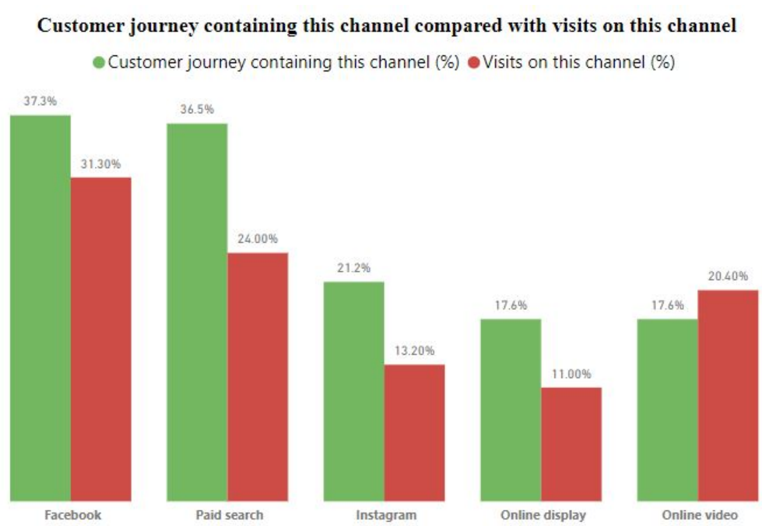


Figure 7: Customer journeys containing this channel compared with visits on this channel (%)

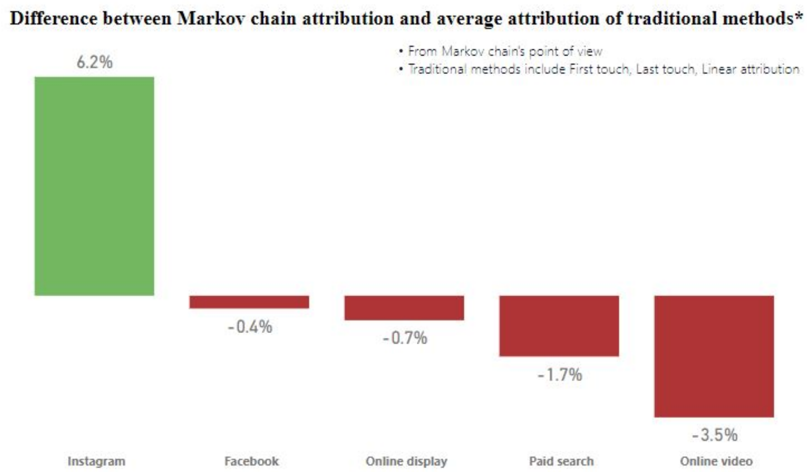


Figure 8: Difference between Markov chain and average attribution of traditional methods

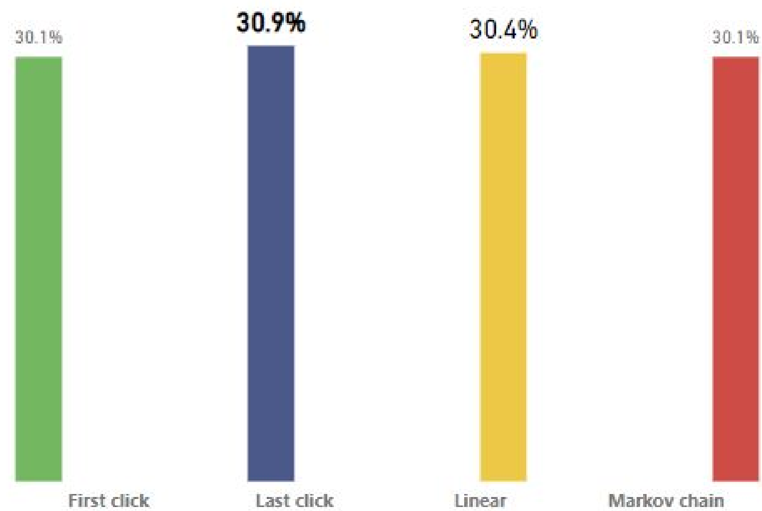
When it comes to **Markov chain model**, an evident **difference** noticed is that **Instagram** has moved up one place, and Online video has moved down one place in the ranking. Delving into the details, **Figure** 8 better shows the difference between the Markov chain attribution and the average attribution of the traditional methods. It is noted that Instagram has been significantly increasingly credited (+6.2%), so that it moves up a place in the said ranking, while all the other channels have been decreasingly credited (-3.5%), wherein the decrease of credit of Online video is much greater than the ones of Paid search (-1.7%), Online display (-0.7%) and Facebook (-0.4%). As a result, the **Online video** has moved down one place in the ranking of the Markov chain, and the other ones did not move the ranking, but just received less credit. The reduced credit of all the other channels has been **added** to the credit of **Instagram**.

5.2 In-channel comparison

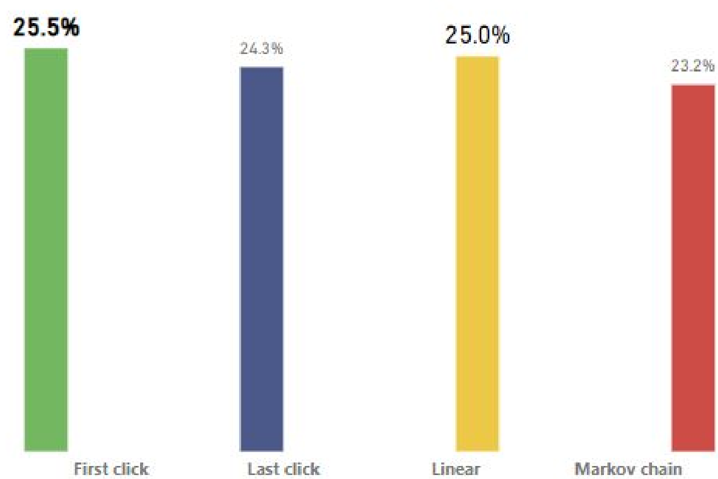
Figure 9 delves more into the **in-channel comparisons**. This ankle can provide us more insights into why certain channels always seem of high or low contribution to traditional methods and what is the **pitfall** of not seeing the **real reasons** behind such differences.

As elaborately explained in the last paragraph, the **Linear** attribution solely depends on the number of visits to the channels. The **First** touch and **Last** touch attribution is decided by the count of occurrences of this channel on the First or the Last click on a (more than or equal to 3 channel-) journey. The Facebook attribution, for example, perfectly showcases the rationale. The Last-click count is Facebook's most preceding measure, where it has the largest **advancement** (in **percentage**) than the other channels, and this advancement is larger than in the comparison in the other measures. In other words, the positions and number of visits decide which traditional attribution methods are more advantageous for this channel.

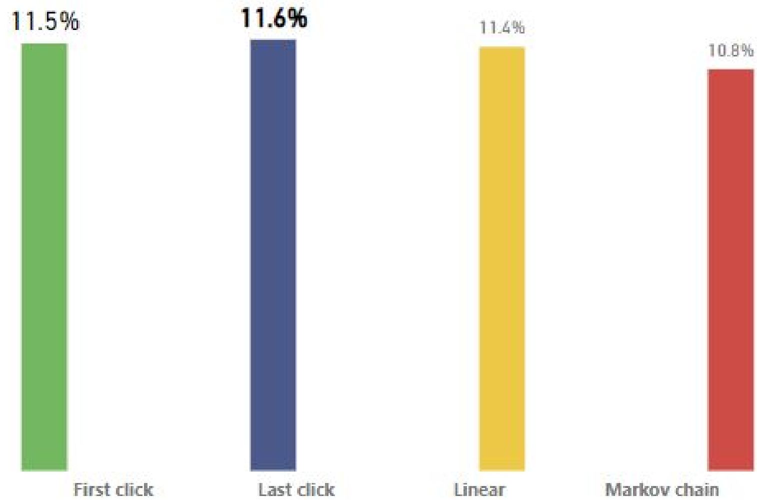
Facebook attribution per method



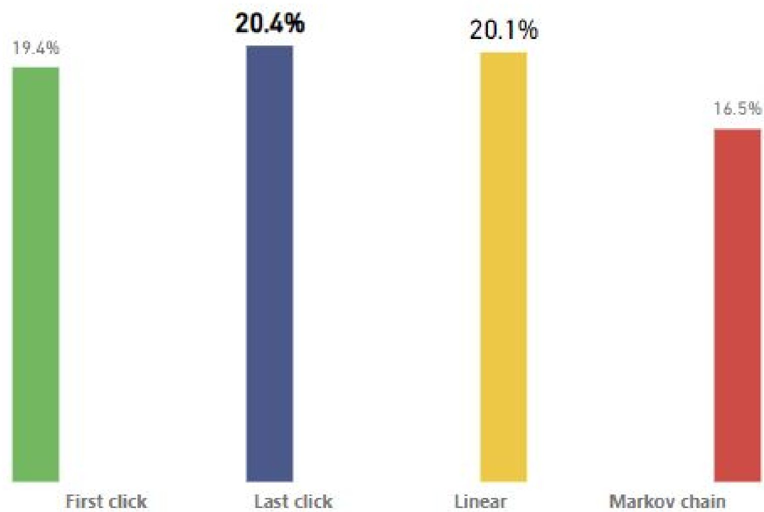
Paid search attribution per method



Online display attribution per method



Online video attribution per method



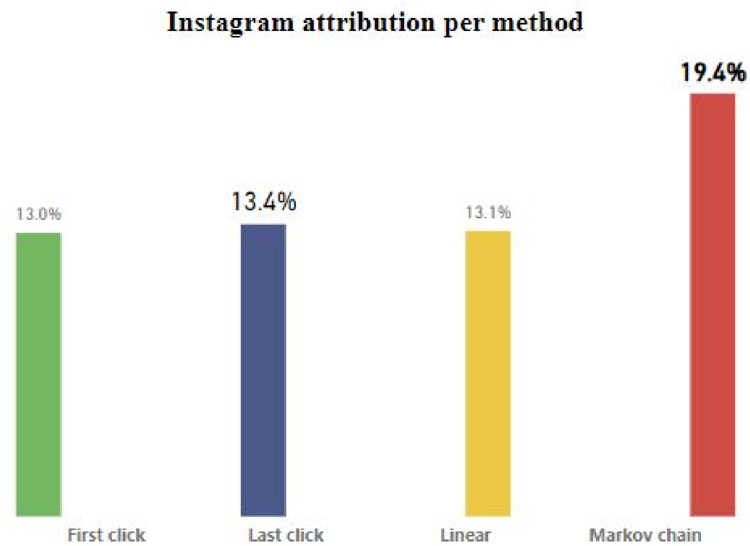


Figure 9: In-channel comparison of attributions per five channels

The remarkable number of being the First click has assigned **Paid search** the highest attribution using the First click method. It is important to note that as though a channel does not take any advanced position in any measure such as occurrences as First clicks or the number of visits, the relative higher percentage in a certain measure still makes the advantage of using a certain method felt. An example is **Online display**. It has the least number of visits among all the channels and the least occurrences as Last clicks and First clicks. However, in the in-channel comparison, Last click method still yields the highest attribution across methods, due its relatively higher percentage of being the Last clicks. The attribution of **Online video** can be explained in the same fashion, where the relatively advanced percentage makes the Last click method the most favorable. On the other hand, **Instagram** has the highest attribution on Markov chain method, which is an ineluctable signal of reconsidering if channel attribution of traditional methods is sound and valid, as such simple-heuristic models are also criticized by H. Li and Kannan (2014) compared with their Bayesian models.

5.3 Comparing results of varying lengths

Apart from showing the result derived from the entire data, it is also interesting to investigate how the channel attribution will be with **different lengths** of customer journeys. This aim is one of the research inquiries of this thesis, in the attempt to discover **if** it is proper to generalize market insights while not paying attention to the possible impact of the journey's length, which might bring **surprise** to the 'common sense' of channel attribution.

The data was sampled into two parts, with each consisting of 67 customer journeys. As in Section 4 enclosed, the lengths of journeys vary from 1 to 135 touchpoints (in short: tp), this thesis took the average number of touchpoints 68 (tp), and sampled the data into **two** parts. One part consists of every customer journey of an **above-average** length, while the other set of data contains customer journeys with a **below-average** length. The Markov chain attributions on both parts of data is shown in the **Table** 6.

Table 6: Markov chain attributions on shorter and longer paths

Shorter path (1 – 68 tp)	Paid search, Facebook, Online display, Online video, Instagram
Longer path (69 – 136 tp)	Facebook, Instagram, Online video, Paid search, Online display

It is noticeable that, according to the Markov chain attribution, **Instagram** is the least contributive in the shorter paths while being the second most important in the **longer** ones. This is similar to the result of Abhishek, P. Fader, and Hosanagar (2012), where they find **Display** differently higher credited by the HMM model than the simplistic models, due to its more observed incremental value in moving a customer from a disengagement stage to having the **attention** of the product. Accordingly, the Markov chain model provides a **different and valid** allocation for online channels, as extant research concludes (Abhishek, P. Fader, and Hosanagar 2012; Anderl et al. 2016; Xu, Duan, and Whinston 2014).

Online video is observed to have **idiosyncratic solid carryover** effects as H. Li and Kannan (2014) and Anderl et al. (2016) defined and encountered in their research. The reason is many customers experience multiple Online video touchpoints in one journey, which mostly leading up to a conversion. This sequence of Online video is mostly observed in **longer** paths, which endorses the higher contribution toward a longer path's conversion. **Anticipating** such carryovers can help marketers improve targeting measures and reach customer groups more efficiently. Another reason for Online video receiving more credit in longer paths is that the longer paths usually cover a more complete set of customer decision process (Wolny and Charoensuksai 2014), the channels that foster a **primary stage** such as brand awareness or attention are much likely to show their value there, and so are **Instagram, Facebook and Online video**. This is similar to the results about Display in Abhishek, P. Fader, and Hosanagar (2012).

Paid search shows the opposite. It is the most impressive channel in the **shorter** paths but almost the last in the longer paths. Paid search ads are triggered by customers' searching of certain keywords, which action already infers a potential close to purchase state of the customer (i.e., Information Searches - Alternative Evaluation) as Wolny and Charoensuksai (2014) mentioned. Thus, after Paid search, a conversion may follow. Similarly, **Online display** influences customers much better in the **shorter** than the longer paths, from which a relatively immediate not long-lasting impact is discovered. This can be interpreted in the way that the display ads are normally visuals, which mostly create a stimulating effect in the short term for the customers, which impact might decrease over time. However, the result of this thesis about Online display, i.e., being not as contributive in none of the models and not in a longer path, is **different** than the ones in Abhishek, P. Fader, and Hosanagar (2012). The reason could be the **heterogeneity** across data (in customers, product, time), the order of the Markov chain and much more differences in the research environment.

5.4 Empirical insights

Should this data be representative, some **general empirical insights** are distilled into **occurrences** and **positions** of channels:

- Facebook and Paid search are frequently visited and frequently appearing at **First** click or **Last** clicks of the online customer journeys, with Facebook's visit numbers being higher than the Paid search's.

- Online display **seldom** appears in the First or Last clicks.

- Online display and Instagram are **not** as frequently **revisited** as other channels. This is deductible by their low total visits and **high** ratio of customer journeys **containing** such channels. It means that there is a decent number of journeys containing (at least one) Online display or Instagram, however, not as many revisits of them.

- Online video shows the **opposite**. Not as many customer journeys contain Online video, but the number of visits is high, which implies that **many** customers have **never** watched Online video, however those who have **watched** tend to revisit it **a lot**.

- Online video is also highly populated on the **Last** clicks (only after Facebook and Paid search, while these two are of considerably higher visits number than Online video). Combined with the last point, we could interpret that a **group** of preference for Online video exists in the data, who like to revisit and close the journey with this channel.

- Instagram is highly observed in customer journeys (only after Facebook and Paid search, while these two are of considerably higher visits number than Instagram). It is not highly visited, but every five customers at least has one Instagram ad in their journey. This might explain the high credit attached to Instagram by the Markov chain model, and infers a potential **spillover effect**.

- This ranking of the most contributing channels **roughly** corresponds to the ones that Oskam (2021) (2021) has shared, although noteworthy is that the channels mentioned there and in this data are not identical.

When it comes to the **varying length** of customer journeys, the followings should be paid attention to :

- Instagram has a significantly higher impact on customers in **longer** customer journeys.
- Online video shows a slightly higher contribution in **longer** customer journeys.
- Facebook has a slightly higher contribution in **longer** paths but still quite high in shorter paths.
- Paid search, on the other hand, contributes better in **shorter** paths.
- Online display, similarly, helps convert better in **shorter** journeys.

Delving into the **higher attribution of Instagram** calculated by the **Markov chain** than the traditional methods should Markov chain better represent the channel contributions, managers who mostly use the traditional methods may ask themselves:

- If Instagram has been **insufficiently** attributed in the history, as Markov chain remarks significant more credit to Instagram despite of its less occurrences and visits than the other channels?
- If Online video has been **overly** sufficiently attributed in the history, as Markov chain shows a remarkably less contribution to this channel despite its high visits, which are concentrated to some customer journeys?
- If Online display and Facebook have also been **overly** sufficiently attributed, although maybe not as much as Online video, as these channels also receive less attribution by Markov chain, which shows their slightly overestimated contribution?

- Should the **cost** of investment in each channel have been included, the slightly overestimated channels such as Facebook can be as costly as the significantly over-attributed ones.

6 Conclusion

With the rapid development of the Internet and mobile technologies, individual online customer journeys have become more challenging than ever to understand and impact. The effectiveness of single channels and their interplay are **hyper important** to optimize multi-channel attribution and alleviate customer experience. Concepts such as **conversion funnel** (similar to AIDA) help researchers conceptualize the underlying consumer behavior i.e., the **deliberation process** from a customer's perspective. The **carryover** and **spillover** effects are also observed among the channels in the previous research. Moreover, extant literature finds the **difference** in channel attribution in different **industries**, **length** settings, at which **stages** are the customers on their customer journeys. This thesis provides empirical results which mostly **confirm** the previous research.

This thesis provides empirical results of a large dataset with customer journeys on the individual level. The results mostly **endorse** the insights given by the extant research. At first the **entire dataset** (with an imbalanced but largely ranged lengths setting) was analyzed, which gives such conclusions. First, the Markov chain model provides different attribution than the simple-heuristics, especially on **Instagram**. Instagram has much more credit using the two-order Markov chain model, and also later in the analysis with only longer paths. This refers to the finding of Abhishek, P. Fader, and Hosanagar (2012). They find **Display** differently credited from the HMM model than the simplistic models, due to its more observed incremental value in moving a customer from a disengagement stage to having the attention of the product. Accordingly, the Markov chain model provides a different and mostly more valid attribution for marketers, as extant research concludes (Abhishek, P. Fader, and Hosanagar 2012; Anderl et al. 2016; Xu, Duan, and Whinston 2014). Instagram is also showing a potential **spillover** effect, as it is observed mostly once in many of the customer journeys, with a not ending position and limited revisits. This implies that it is good at moving the customers forward and passing by them to visits of other channels. Second, **Online video** is observed to have **idiosyncratic solid carryover** effects as H. Li and Kannan (2014) and Anderl et al. (2016) encounter in their research. Online video can take up multiple touchpoints on its own for a sequence in many customer journeys, especially in longer paths with conversion. **Anticipating** such carryovers can help marketers improve targeting measures and reach customer groups more efficiently.

Third, when it comes to the **varying length** of customer journeys, the thesis' results of the Markov chain model show that **Instagram, Facebook and Online video** have **more** contribution toward a **longer** path's conversion. This result is similar to the results about Display in Abhishek, P. Fader, and Hosanagar (2012) again. Because the longer paths usually cover a more complete set of customer decision process (Wolny and Charoensuksai 2014), the channels that foster a **primary stage** such as brand awareness or attention are much likely to show their incremental value. So are Instagram, Facebook and Online video. In contrast, **Paid search** has much higher attribution in fostering a shorter path's conversion, which also follows the same logic of Abhishek, P. Fader, and Hosanagar (2012) and Wolny and Charoensuksai (2014), since Paid search mostly contributes on a **later stage** of a customer, such as 'Information searches - Alternatives evaluation'. The result of this thesis about **Online display**, i.e., being not as contributive in none of the models and not in a longer path, is different than the ones in Abhishek, P. Fader, and Hosanagar (2012). The reason could be the **heterogeneity** across data (in customers, product, time), the order of the Markov chain and much more differences in the research environment. Fourth, seeing the **remarkable difference** among the attribution with various lengths, and also the rule-of-thumbs from Oskam (2021), the thesis suggests marketers to be **careful** to adopt any generalization of channel attribution as **heterogeneity** (in industry, customers and time etc.) may bring vastly different channel value distributions. Fifth, the Markov chain model provides the transition matrix, which contains the probability a customer moves through different stages, which can be depicted in a graph or a matrix as described in Chapter Methodology. This can be a good tool for managers to **automate predicting** a customer's next move and putting the best channel(s) upfront to due customers for higher conversions.

7 Limitation

The thesis has many limitations, mainly regarding the data and modeling part. First, the **data** collects exposures on a cookie level, more specifically, an individual device level. Thus, multiple customer journeys might belong to one customer and his or her journey but the data did not have such information allowing combine them and map to a complete customer journey. Second, the data does not contain offline channels, which does not allow looking at the total attribution of online channels in the whole picture, and analyzing the interactions between online and offline channels, similar to the limitation of Abhishek, P. Fader, and Hosanagar (2012). Third, the background of the data is not clearly given, which does not allow gaining insights into a particular industry or product. As Wolny and Charoensuksai (2014) shared, high-involvement products normally have an entire decision-making process. Thus, (future) research looking into the above **aspects** (i.e., cross-devices customer journeys, specific data source, offline and online channel interactions) may lead to more meaningful implications.

When it comes to the **modeling**, first, the thesis did not control for customer heterogeneity, unlike Abhishek, P. Fader, and Hosanagar (2012). This does not allow the managers to separate the effects from the offline channels that may have also influenced the customers. Second, the thesis did not extensively evaluate the models' predictive power and robustness (stability). The reason is that this thesis aims more for descriptive analysis rather than predictive. However, this might reduce the plausibility of the model fits and the insights. Third, the thesis only separates the customer journeys into two parts based on length. However, the dataset can be divided into more different lengths to observe the difference on a more detailed level. Moreover, the researchers can also make segments according to a certain pattern (e.g., customers with many Online video touchpoints in a long sequence) to focus on a specific group of customers. Thus, (future) research that addresses these above **aspects** (i.e., control offline channels effects/customer bias, testing predictive power and robustness of models, focusing on a particular customer group) may yield more significant and robust implications.

References

Abhishek, V., P. Fader, and K. Hosanagar (2012). "Media exposure through the funnel: A model of multi-stage attribution". In: *Available at SSRN 2158421*.

Anderl, E. et al. (2016). "Mapping the customer journey: Lessons learned from graph-based online attribution modeling". In: *International Journal of Research in Marketing* 33.3, pp. 457–474.

Berman, R. (2018). "Beyond the last touch: Attribution in online advertising". In: *Marketing Science* 37.5, pp. 771–792.

Brown, L. (June 2020). *7 Types of Attribution Models*. URL: `https://www.empiricalpath.com/insights/7-types-of-attribution-models/`.

Bruce, N. I., K. Peters, and P. A. Naik (2012). "Discovering how advertising grows sales and builds brands". In: *Journal of marketing Research* 49.6, pp. 793–806.

Bryl', S. (Aug. 2016). *Marketing Multi-Channel Attribution Model with R (Part 1: Markov Chains Concept)*. URL: `https://www.analyzecore.com/2016/08/03/attribution-model-r-part-1/`.

Chierichetti, F. et al. (2012). "Are web users really markovian?" In: *Proceedings of the 21st international conference on World Wide Web*, pp. 609–618.

Dalessandro, B. et al. (2012). "Causally motivated attribution for online advertising". In: *Proceedings of the sixth international workshop on data mining for online advertising and internet economy*, pp. 1–9.

Hennig-Thurau, T. et al. (2010). “The impact of new media on customer relationships”. In: *Journal of service research* 13.3, pp. 311–330.

Jansen, B. J. and S. Schuster (2011). “Bidding on the buying funnel for sponsored search and keyword advertising”. In: *Journal of Electronic Commerce Research* 12.1, p. 1.

Lee, G. (2010). “Death of ‘last click wins’: Media attribution and the expanding use of media data”. In: *Journal of Direct, Data and Digital Marketing Practice* 12.1, pp. 16–26.

Li, H. and P. Kannan (2014). “Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment”. In: *Journal of Marketing Research* 51.1, pp. 40–56.

Lukmani, M. (Sept. 2021). *Ditch the Last-Click Model. Here’s How to Overhaul your Attribution.* URL: `https://www.i-com.org/news-articles/ditch-the-last-click-model-heres-how-to-overhaul-your-attribution`.

Moffett, T. (Nov. 2012). *Google Rolls Out Its Attribution Modeling Tool To A Wider Audience.* URL: `https://www.forrester.com/blogs/12-11-01-google_rolls_out_its_attribution_modeling_tool_to_a_wider_audience/`.

Montgomery, A. L. et al. (2004). “Modeling online browsing and path analysis using clickstream data”. In: *Marketing science* 23.4, pp. 579–595.

Montoya, R., O. Netzer, and K. Jedidi (2010). “Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability”. In: *Marketing Science* 29.5, pp. 909–924.

Mulpuru, S. (2011). “The purchase path of online buyers”. In: *Forrester report* 5, p. 12.

OrbitalAds. (Nov. 2021). *Five digital marketing channels that will help you increase conversions.* URL: `https://www.orbitalads.com/blog/5-digital-marketing-channels`.

Oskam, J. (Mar. 2021). *New Media Analytics - Ecommerce [Lecture Slides].* Erasmus School of Economics.

Parment, A., P. Kotler, and G. Armstrong (2011). *Principles of marketing: Swedish edition*. Prentice-Hall.

Salazar, K. (Dec. 2016). *How channels, devices, and touchpoints impact the customer journey*. URL: `https://www.nngroup.com/articles/channels-devices-touchpoints/`.

Schwartz, E. M. et al. (2011). "'Children of the HMM': Modeling Longitudinal Customer Behavior at Hulu. com". In: *Com (August 1, 2011)*.

Schweidel, D. A., E. T. Bradlow, and P. S. Fader (2011). "Portfolio dynamics for customers of a multiservice provider". In: *Management Science* 57.3, pp. 471–486.

Shao, X. and L. Li (2011). "Data-driven multi-touch attribution models". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 258–264.

Singal, R. et al. (2022). "Shapley meets uniform: An axiomatic framework for attribution in online advertising". In: *Management Science*.

Turow, J. (2010). *Media Today, 2010 Update: An Introduction to Mass Communication*. Routledge.

Wolny, J. and N. Charoensuksai (2014). "Mapping customer journeys in multichannel decision-making". In: *Journal of Direct, Data and Digital Marketing Practice* 15.4, pp. 317–326.

Xu, L., J. A. Duan, and A. Whinston (2014). "Path to purchase: A mutually exciting point process model for online advertising and conversion". In: *Management Science* 60.6, pp. 1392–1412.