InternexT

INTERNATIONAL BUSINESS AND MANAGEMENT REVIEW

v.20, n. 3, e866, 2025| e-ISSN: 1980-4865 | http://internext.espm.br

WHICH TOUCHPOINTS MATTER MOST? A DATA-DRIVEN MODEL FOR UNDERSTANDING STUDENT JOURNEY

Rogério Ferraz dos Santos[1] & Luciana Florêncio de Almeida[1]

[1]Escola Superior de Propaganda e Marketing – São Paulo (SP), Brazil.

ARTICLE DETAILS

Received:
May 12, 2025

Accepted:
September 1, 2025

Available online:
October 10, 2025

Double Blind Review System

Editor in Chief
Priscila Rezende da Costa

ABSTRACT

Objective: This research aimed to develop an attribution model using data from individual customer journeys to assess marketing channels and allocate value to various touchpoints throughout those journeys. **Method:** Through a comprehensive case study, the research developed and applied an attribution model using data from individual customer journeys to assess marketing channels and allocate value to various touchpoints throughout those journeys. The model analyzed 662,838 online and offline touchpoints across 185,631 customer journeys from a Brazilian higher education institution database. **Main Results:** The research on a student's journey shows that email, live chat, call center, sales, and inbound interactions were responsible for over 70% of enrollments. It also emphasizes the significance of customer-initiated contacts over firm-initiated contacts, with brand-owned touchpoints making up over 80% of registered contacts. **Relevance / Originality:** This article offers a dual contribution to the literature on student journey by evaluating the effectiveness of marketing touchpoints and exploring robust measurement methodologies. **Theoretical / Methodological Contributions:** The primary contribution of this research to the field lies in demonstrating that models based on data collected throughout the customer journey—and which account for all marketing touchpoints—achieve superior performance in identifying the specific contribution of each channel. **Social contributions for management:** This study offers three key managerial insights: it underscores the strategic value of integrating the customer journey into planning, the importance of mapping and monitoring touchpoints through data systems, and the need for continuous, iterative improvements to enhance conversion and customer experience.

Keywords: Attribution Models, Consumer Behavior, Decision Making, Higher Education, Marketing.

QUAIS PONTOS DE CONTATO SÃO MAIS IMPORTANTES? UM MODELO BASEADO EM DADOS PARA ENTENDER A JORNADA DO ESTUDANTE

DETALHES DO ARTIGO

Recebido:
12 maio, 2025

Aceito:
1º setembro, 2025

Disponível online:
10 outubro, 2025

Sistema de revisão "Double blind review"

Editora Chefe
Priscila Rezende da Costa

RESUMO

Objetivo: Esta pesquisa teve como objetivo desenvolver um modelo de atribuição usando dados de jornadas individuais de clientes para avaliar canais de *marketing* e atribuir valor a vários pontos de contato. **Método:** Por meio de estudo de caso em uma instituição de ensino superior, a pesquisa desenvolveu e aplicou um modelo de atribuição que analisou 662.838 pontos de contato *online* e *offline* em 185.631 jornadas de estudantes. **Principais Resultados:** A pesquisa mostra que *e-mail*, chat ao vivo, *call center*, vendas e interações *inbound* foram responsáveis por mais de 70% das matrículas. Também enfatiza a importância dos contatos iniciados pelo cliente em relação aos contatos iniciados pela empresa, com pontos de contato de propriedade da marca representando mais de 80% dos contatos registrados. **Relevância / Originalidade:** Este artigo oferece uma contribuição dupla para a literatura sobre a jornada do estudante, avaliando a eficácia dos pontos de contato no *marketing* e explorando metodologias de medição robustas. **Contribuições Teóricas / Metodológicas:** A principal contribuição desta pesquisa para a área está em demonstrar que modelos baseados em dados coletados ao longo da jornada do cliente — e que consideram todos os pontos de contato de *marketing* — apresentam desempenho superior na identificação da contribuição específica de cada canal. **Contribuições sociais / para a gestão:** Este estudo oferece três principais *insights* gerenciais: destaca o valor estratégico de integrar a jornada do cliente ao planejamento, a importância de mapear e monitorar os pontos de contato por meio de sistemas de dados e a necessidade de melhorias contínuas e iterativas para aumentar a conversão e aprimorar a experiência do cliente.

Palavras-chave: Comportamento do Consumidor, Tomada de Decisão, Ensino Superior, Marketing, Modelos de atribuição.

*Corresponding author: luflorencio@gmail.com

https://doi.org/10.18568/internext.v20i03.866

INTRODUCTION

Researchers have developed various models to quantify the impact of touchpoints on customer journeys, often referred to as attribution models or path-to-purchase models (Lemon & Verhoef, 2016). In their comprehensive review, Lemon and Verhoef (2016) emphasize the critical need to consider the cumulative effects of multiple touchpoints on customer journey outcomes.

Despite the availability of more sophisticated attribution methods, many organizations still rely on simpler heuristic or rule-based models, such as first-touch or last-touch attribution. These traditional models fail to capture the customer's multi-touch experience before conversion, leading to inaccurate insights when measuring results across different channels (Li & Kannan, 2014). Moreover, these methods often analyze individual channels in isolation, neglecting potential interactions and synergies (de Haan et al., 2016).

Understanding attribution in the context of higher education is particularly relevant given the complexity and competitiveness of this sector. In Brazil, the higher education market has undergone significant transformations in recent decades, driven by government regulations, public funding, and increased private investment (Abreu et al., 2019). This sector has become fiercely competitive (Rosenbaum et al., 2017; Senhoras et al., 2012), with prominent educational groups vying for leadership through strategic investments in brand development, technological advancements, and geographical expansion via acquisitions, mergers, and online learning platforms (CADE, 2016; Piurcosky et al., 2019).

Navigating the enrollment process at a higher education institution (HEI) presents significant complexity. Prospective students engage in a multi-stage journey, encountering various touchpoints, evaluating numerous factors, experiencing diverse emotions, and being influenced by distinct reference groups (Følstad and Kvale, 2018; Galan et al., 2015). These dynamics underscore the need for a more nuanced approach to attribution in the higher education student journey (HESJ).

In response to this need, the present study proposes a novel, data-driven attribution model tailored to the HESJ. The model leverages panel data from both online and offline channels, including detailed social media engagement data. Employing a case study approach at a Brazilian HEI, the research follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) for data collection and analysis.

These research contributions offer valuable insights for marketing professionals and data analysts, enhancing their understanding of the student journey and the role of each touchpoint in driving enrollment decisions.

The following sections will explore previous literature on customer journey data-driven approaches and the empirical background. The methodology section will detail the case study, data collection methods and data analysis protocols. The final sections will present the results, discussion, and implications, followed by limitations and suggestions for future research.

1. CUSTOMER JOURNEY DATA-DRIVEN APPROACHES

The adoption of data-driven approaches in customer journey studies has gained prominence. Lemon and Verhoef (2016) underscore the significance of data science in comprehending and personalizing the customer journey. Previous works on customer journey theory have embraced data-driven approaches, exemplified by McColl-Kennedy et al.'s (2019) longitudinal study that assessed customer experience and value creation throughout the customer journey using data-mining techniques. Similarly, Ordenes et al. (2014) employed text mining to analyze customer feedback from a UK company, presenting empirical evidence of the benefits of this innovative method in enhancing the customer journey.

Archak et al. (2010) pioneered the application of a Markovian graph method in a data-driven attribution model, utilizing the steady-state probabilities of different random walks to capture structural correlations. Shao and Li (2011) analyzed a massive dataset from an advertising campaign, proposing a more accurate data-driven method considering all available data. Li and Kannan (2014) focused on estimating carryover and spillover effects between touchpoints, introducing a new model for attributing credit for conversions, incorporating data from online channels like email and site visits. Anderl et al. (2016)

expanded upon Archak et al. (2010), proposing an attribution model based on the first and higher-order Markov chain. Kakalejč et al. (2018) applied the same model to analyze data from an e-commerce website. Zhao et al. (2019) evaluated different regression models in various advertising channels, introducing the use of partially linear additive models as a specific contribution.

Most previous studies primarily focused on online advertising, neglecting offline marketing, as also observed by de Haan et al. (2016). However, Wiesel et al. (2010) proposed a model evaluating the impact of both online and offline marketing channels on financial results, incorporating aggregated marketing actions metrics as inputs. Combining data from online and offline channels could enable a more accurate value allocation (Buhalis and Volchek, 2021). In contrast to similar approaches like Wiesel et al. (2010), the current research examines most marketing channels used by the analyzed HEI, including 17 different touchpoint categories encompassing online and offline channels. Notably, the current study employs user-level data to scrutinize offline touchpoints, avoiding the use of aggregated data.

Attribution models play a pivotal role in providing a more accurate evaluation of marketing channels, allocating value and significance to different touchpoints along customer journeys (Buhalis and Volchek, 2021; Kannan et al., 2016). The widespread use of digital marketing and Customer Relationship Management (CRM) systems has led to the availability of individual-level touchpoint data, spread across various company databases. Despite its inherent complexity, these detailed data on touchpoints present excellent research opportunities (Kannan & Li, 2017; Tueanrat et al., 2021). Individual-level data enable the detection of touchpoint sequences and the measurement of each one's relative contribution to conversions, offering a more realistic picture of the role of marketing channels along the customer journey and providing valuable insights for marketing planning (Buhalis & Volchek, 2021).

Traditional attribution models are typically categorized into two groups: single-touch and multi-touch models. Single-touch methods attribute the total value of a conversion to a single touchpoint in the journey, while multi-touch models derive value from the entire array of touchpoints that customers experience (Buhalis & Volchek, 2021). These earlier models, whether single or multi-touch, based their value allocation principles and computational techniques on simple heuristics (rule-based). While easy to comprehend and calculate, the main drawback of rule-based models is their failure to account for customer journey dynamics (Buhalis & Volchek, 2021) and their inadequacy for lengthy customer journeys (Nichols, 2013).

Li and Kannan (2014) also critique the fragilities of rule-based models, emphasizing that standard last-clicked models ignore spillover effects between channels. In their study, the last-click model underestimates the contribution of channels like email, display, and referral. Prior research contends that the last-click model provides biased insights to marketing practitioners, primarily because it completely disregards previous touchpoints in the journey (de Haan et al., 2016; Shao & Li, 2011). Metrics provided by linear and weighted models (such as time-touch and position-touch) consider all touchpoints leading to a conversion. However, these models do not account for paths that do not result in conversions, thereby missing valuable information (Anderl et al., 2016; Li & Kannan, 2014).

The availability of individual data on the customer journey, coupled with clear identification of touchpoints before conversions, has rekindled interest in the attribution problem (Kannan et al., 2016), leading to the emergence of a new class of models based on algorithmic and data-driven approaches. These data-driven models employ advanced statistical methods to analyze customer journey data, including logistic or linear regression models (Shao & Li, 2011; Zhao et al., 2019), hierarchical Bayesian models (Li & Kannan, 2014), game theoretical models (Berman, 2018), time series models (Kireyev et al., 2016), hidden Markov models (Abhishek et al., 2012), and discrete-time Markov models (Anderl et al., 2016; Archak et al., 2010). Regardless of the method chosen, data-driven attribution models require a solid theoretical foundation to guide parameter selection and configuration, as well as the interpretation of results (Kannan et al., 2016; Lemon & Verhoef, 2016).

Table 1 summarizes the timeline of the referenced studies, including dataset size, industry, channel types and count, and the attribution model used.

Table 1. Previous data-driven and multi-touch attribution studies.

Study	Industry	Type of chan-nels	Data	No. of Cha-nnels	Data sets	Time Frame	Model
Archak et al. (2010)	(Not informed)	Online Advertising	~ 27 thousand journeys ~ 2.6 million touchpoints Source: Google analytics tools.	6	8	Two months	First-order Markov random walks
Shao and Li (2011)	Software	Online Advertising	72.5 million journeys 2 billion touchpoints (impressions) Source: Google analytics tools.	39	1	One month	Logistic model and Probabilistic model
Li and Kannan (2014)	Hospitality	Online Advertising Digital channels	1,997 journeys 22,369 touchpoints Source: Double Click, Omniture Catalyst, and Online Ads platforms.	6	1	Two months	Hierarchical Bayesian model
Anderl et al. (2016)	Travel agency Fashion retail Luggage retail	Online Advertising	3 million journeys 4.8 million touchpoints Source: Google analytics tools.	10	4	One month	First and higher-order Markov random walks
Alblas (2018)	Travel Technical supplies	Online Advertising Digital channels	223 thousand journeys 1.2 million touchpoints Source: Google analytics tools	10	2	One year	First and higher-order Markov random walks
Kakalejč et al. (2018)	Electronics retail	Online Advertising Digital channels	~ 8,484 journeys ~ 49,123 touchpoints Source: Google analytics tools	7	1	Four months	First and higher-order Markov random walks
Zhao et al. (2019)	(Not informed)	Online Advertising	153,891 journeys 1,047,000 touchpoints (impressions) Source: Google analytics tools	18	1	Three months	Regression models
This Study (2022)	Higher Education	Online Advertising Digital channels Offline channels	185,631 journeys 662,838 touchpoints Source: RD Station, Rubeus CRM, internal ERP, and Instagram.	17	2	Two years	First and higher-order Markov random walks

2. EMPIRICAL BACKGROUND: HIGHER EDUCATION CUSTOMER JOURNEY

Choosing higher education is a complex and high-risk process (James-MacEachern, 2018) that requires candidates to engage in a journey that can span several months. Various models (Table 2) have been developed to conceptualize the main phases of a HEI customer journey and the factors associated with the decision-making process.

Early research by D. W. Chapman (1981), Hanson and Litten (1982) and Kotler (1976) laid the ground-

Table 2. Models proposed for Customer Journey stages in higher education.

Author	Customer Journey Stages
Kotler (1976)	(1) The decision to attend. (2) Information seeking and receiving. (3) Specific college inquiries. (4) Applications. (5) Admissions. (6) College choice. (7) Registration.
Hanson and Litten (1982)	(1) College aspirations. (2) Beginning the search process. (3) Gathering information. (4) Sending applications. (5) Enrolling.
Hossler and Gallagher (1987)	(1) Predisposition. (2) Search. (3) Choice.
Maringe (2006)	(1) Pre-search behavior. (2) Application stage. (3) Choice decision. (4) Registration.
Galan et al. (2015)	(1) Problem recognition. (2) Information search. (3) Evaluation of alternatives. (4) Purchase decision. (5) Post-purchase decision.
Gai et al. (2016)	(1) Predisposition. (2) Information seeking for targeting schools. (3) Application. (4) Evaluating Admission offers. (5) Final Decision.
James-MacEachern (2018)	(1) Awareness stage. (2) Information stage. (3) Decision stage.
Schuhbauer et al. (2020)	(1) Gathering information. (2) Application. (3) The initial phase of the degree program. (4) The second phase of the program. (5) Alumni phase.

work by investigating the customer journey within higher education. They proposed models where students and their parents navigate through various stages, influenced by a range of touchpoints during the decision-making process.

Building on this foundational work, later studies have expanded our understanding of these dynamics. Galan et al. (2015) emphasized that engaging in higher education requires considerable thought, characterizing it as a high-involvement decision process. This perspective aligns with the evolution of decision-making models over time, which, as noted by Gai et al. (2016), have transitioned from simpler to more complex sets of decision variables. This progression reflects a deeper engagement by candidates as they move through their educational journey, factoring in a broader spectrum of influences and information sources.

The role of modern technology and digital platforms in this decision-making process has also been highlighted in recent years. James-MacEachern and Yun (2017) specifically explored the impact of social media, suggesting that prospective students utilize these platforms to evaluate alternatives and gather insights from the experiences of alumni. This indicates a shift towards more interactive and peer-influenced phases of decision-making.

Recent research highlights the effectiveness of data-driven approaches in addressing educational challenges. For instance, Nascimento et al. (2018) explored issues of school dropout and failure by analyzing datasets provided by the Brazilian Institute for Study and Research on Education (INEP). Similarly, Vescovi (2020) applied supervised machine learning techniques to forecast and reduce student dropout rates within a HEI. In a broader context, Bolat and O'Sullivan (2017) advocate for the strategic use of data analytics to enhance marketing outcomes and inform business decisions in higher education settings.

Moreover, a more detailed examination of touchpoints was conducted by Schuhbauer et al. (2020), who focused on the student journey at Nuremberg University. They defined touchpoints expansively as "all actions that students can take in connection with the university," such as visiting the university's website to obtain information about academic programs. This definition underscores the variety of interactions that can influence a student's decision, from digital engagements to direct personal experiences.

Collectively, these studies illustrate the dynamic nature of choosing higher education, where traditional models are being augmented by digital interactions and peer influences, reflecting broader societal shifts towards more connected and informed decision-making processes.

3. METHOD

3.1. The selected case study and protocol

The current research employs the case study method, following Yin's (2018) recommendation for a protocol guiding field research and the analytical process. The study adopts the CRISP-DM process, which is described as "a non-proprietary, documented, and freely available data mining model" (Shearer et al., 2000, p. 1). This conceptual model, outlined by Chapman et al. (2000) and utilized in Customer Journey longitudinal studies by McColl-Kennedy et al. (2019), consists of six interactive steps: (a) business understanding, (b) data understanding, (c) data preparation, (d) modeling, (e) evaluation, and (f) deployment.

The selected case for this study is a Brazilian private HEI referred to as XYZ University to maintain confidentiality. As of 2021, the university had approximately 20 thousand students and offered 36 higher education programs across three campuses in the state of Sao Paulo, along with a "virtual campus" that provided distance learning programs. Starting in 2017, XYZ University underwent extensive restructuring of its communication, marketing, and sales departments, implementing new processes and channels to recruit students.

This reorganization culminated in the development of a Student Journey Map (SJM), providing detailed information about the planned path and interactions with prospects and leads. XYZ University's SJM organizes the customer journey into eight stages: traffic, visitors, leads, opportunities, selection, scholarship, enrollment, and student retention, each comprising steps and detailed actions related to customer interactions.

The evaluated journey begins at the "leads" stage, where the lead is identified by email, and concludes at the "enrollment" stage. Anonymous data collected in previous stages are integrated into the customer journey once the prospect enters the "lead" stage. Individuals receive more firm-initiated contacts, such as phone calls and electronic messages, as they progress through the journey.

The current study incorporates data from five different sources, as detailed in Table 3, along with their relation to the channels evaluated and the respective number of records extracted.

For data preparation, the authors partitioned the data into two distinct datasets:

a) Period 1: Data from customers who applied or enrolled in the 2019 second semester admission (2019-S2) and the 2020 first semester admission (2020-S1). This dataset encompasses all journeys that did not result in conversion during the application process or enrollment between April 2019 and March 2020;
b) Period 2: Data from customers who applied or enrolled in the 2020 second semester admission

Table 3. Data sources for student journey data.

Data sources	Description	Number of records extracted	Period
Enrollment system	Info about enrolled students from a different campus.	46,651	11/1998 to 04/2021
Application system	Info about the application and affiliate channel.	1,000,922	03/2018 to 03/2021
MKT Automation system	Data related to Ad Display, Ad Facebook, Automation, Email, Events, Inbound, Lives, Messages, Live Chat, Promotion, Referral, SEO, Social FB, Social IG, and Social YT.	1,096,472	08/2017 to 04/2021
CRM	Commercial CRM database. Conversions related to Automation, Sales, and Call Center channels.	240,784	10/2018 to 04/2021
Instagram	Engagement with XYZ University official profiles (5 profiles). Likes and comments on posts.	4,453 posts 394,007 likes	04/2019 to 04/2021

(2020-S2) and the 2021 first semester admission (2021-S1). This set includes all journeys that did not culminate in conversion during the application process or enrollment between April 2020 and April 2021.

Some journeys (17.92%) have data in both sets, as certain candidates applied to the selection process at different times. The data spans touchpoints recorded between September 2017 and April 2021. In addition to digital channels, the study encompassed offline channels and interactions on Instagram.

Following Rosenbaum et al.'s (2017) recommendation to simplify the analysis, a critical pre-processing step was the consolidation of the raw data. Over 2,300 unique touchpoint classifications (representing granular events from the source data) were mapped and consolidated into 17 distinct online and offline marketing channels. This process was essential to create a cohesive dataset suitable for modeling the student journey.

- Channel owner: social, partner-owned, brand-owned, and customer owner. Lemon and Verhoef (2016) applied a similar classification in their work. The present study did not collect data on Customer Owner channels;
- Contact initiative: customer-initiated contacts (CICs) and firm-initiated contacts (FICs). This classification, utilized in previous studies by de Haan et al. (2016), Li and Kannan (2014), and Wiesel et al. (2010), distinguishes interactions initiated by customers from those initiated by the firm.

3.2. Attribution model based on a Markov graph approach

This study embraces the Markov graph model as proposed by Anderl et al. (2016) and Archak et al. (2010) to analyze the student journey as a multi-channel process. The model's capacity to incorporate a full spectrum of interactions, including offline and social media touchpoints, is critical for a holistic analysis. In accordance with the taxonomy from Buhalis and Volchek (2021), our methodology can be classified as a multi-touch, data-driven, and cross-channel attribution approach. This model creates "maps" to understand how students really move between marketing channels. It maps all the "roads" (channels) and analyzes the "turns" (touchpoints) to see how sequences lead to success. It then measures a channel's importance by asking: "If we close this road, how many students would get lost?"

At its core, the model is an application of discrete-time Markov Chains. The student journey is conceptualized as a directed Markov Graph characterized by a set of states, S. This set includes all marketing channels as transient states, along with three special states: a 'start' state to initiate all journeys, a 'conversion' state for successful enrollments, and a 'null' state for non-converting journeys. The transitions between these states are governed by a transition matrix, where the probability of moving to a future state depends only on the present state (the Markov property).

While a standard, first-order Markov model considers only the immediately preceding touchpoint, a higher-order model extends this logic by incorporating a memory of *k* previous states. This allows the model to capture more complex sequential patterns and dependencies. For instance, a first-order model evaluates a direct transition like "Inbound → Email", whereas a higher-order model can assess the probability of a longer sequence such as "Social Media → Inbound → Email → Event", thereby capturing the synergistic effects of specific channel combinations.

The central metric for attribution in this study is the Removal Effect. This metric quantifies a channel's importance by estimating the total loss in conversions if that channel were to be removed from the ecosystem. This effect is calculated through a dynamic simulation of thousands of 'random walks'—hypothetical journeys governed by the network's transition probabilities. By comparing the overall conversion probability of the complete system with a system where one channel is absent, the Removal Effect captures the cumulative and indirect influence of that channel within the journey's dynamic interplay. The transition probabilities that govern this system are derived empirically from the dataset, yielding a data-driven network as exemplified in Figure 1.

In summary, this model provides an integrated view of the student journey by considering the interdependencies among all touchpoints. The next section details the criteria adopted to select the most suitable model order.

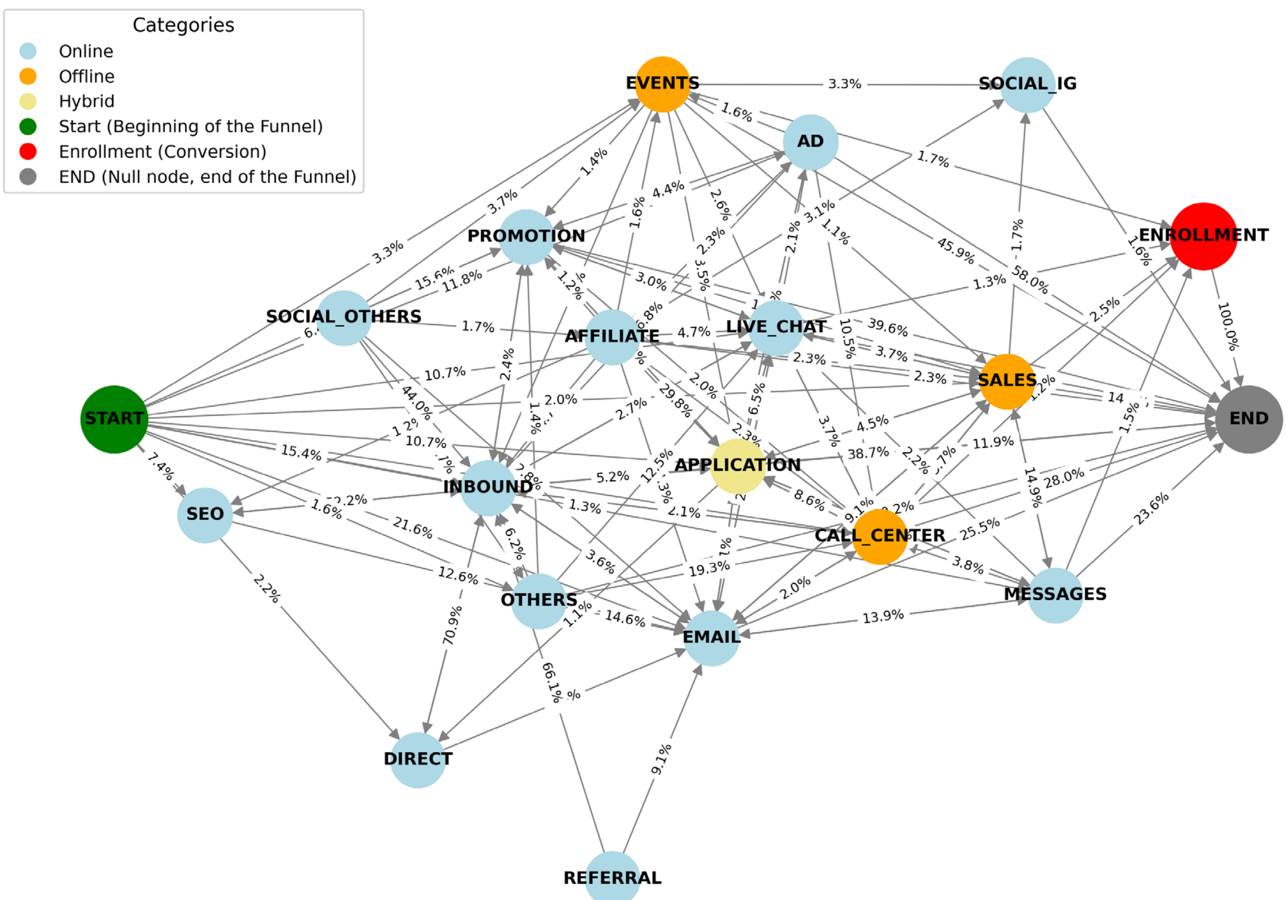

Figure 1. Markov graph for dataset 1, showing only transitions > 1% with self-loops removed.

3.3. Model order selection and predictive evaluation

The primary methodological challenge in applying higher-order Markov models is the selection of an optimal order (*k*) that balances predictive power with model complexity. To address this, we based our model assessment on predictive efficacy, as proposed by Shao and Li (2011).

The highly imbalanced nature of our dataset, where conversions represent only 3% of cases, renders standard information criteria like Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) unsuitable for model selection. While prior research has employed other specialized metrics for this task, such as top-decile lift (Anderl et al., 2013) and the Global Dependency Level (GDL) criterion (Kakalejč et al., 2018), we selected Receiver Operating Characteristics (ROC) Analysis as the primary evaluation method due to its proven robustness in such conditions. The Area Under the Curve (AUC) derived from this analysis quantifies the model's ability to effectively distinguish between converting and non-converting journeys, irrespective of the class imbalance.

In contexts where class imbalance is significant—such as when only a small proportion of students enroll—traditional accuracy metrics may provide misleading results, as they tend to favor the majority class. The AUC offers a more robust alternative by evaluating the model's ability to correctly rank instances. Specifically, AUC assesses whether the model can distinguish between students likely to enroll and those unlikely to do so, thereby providing a more meaningful measure of predictive performance under imbalance conditions.

As observed by Anderl et al. (2016), predictive accuracy measured by AUC tends to increase with the Markov order. However, this gain is accompanied by an exponential rise in model complexity, increasing the risk of overfitting. To navigate this trade-off, we adopted a Penalized AUC measure, as suggested by Altomare and Loris (2016). This metric, conceptually analogous to an adjusted R^2 in regression models, discounts the standard AUC based on model complexity, allowing for the selection of a more parsimonious model without sacrificing significant predictive capability. The formula is defined as Equation 1:

$$\text{Penalized AUC} = 1- [(1-\text{AUC}) (np-1) / (np-nnodes-1)] \quad (1)$$

Where:
np: the number of individual paths observed;
nnodes: the number of nodes in the graph, a measure of the model's complexity, which is dependent on the number of channels and the model order (k).

Penalized AUC introduces a complexity adjustment to the model's performance evaluation, aiming to balance predictive accuracy with model simplicity. By incorporating a penalty for overfitting, this metric discourages the selection of overly complex models that may perform well on training data but lack generalizability. Consequently, Penalized AUC supports the identification of models that are not only accurate but also robust and reliable for future predictions.

By calculating both AUC and Penalized AUC for models of varying orders, as illustrated in Figure 2, we identified the fourth-order model as the optimal choice. This model demonstrated a strong predictive performance, with AUC values of 0.876 for dataset 1 and 0.860 for dataset 2. Although a fifth-order model yielded a marginal improvement in AUC, the Penalized AUC indicated that this gain did not compensate for the significant increase in complexity.

Therefore, to ensure consistency across the analysis and favor a simpler model, the fourth-order Markov model was adopted for the attribution analysis presented in this study.

To facilitate the application and replication of our methodology, a toolkit notebook is available at https://bit.ly/AttributionToolkit. This resource provides the code and instructions for processing data exported from CRM and Digital Marketing platforms to construct an attribution model using the method presented in this study.

4. MAIN FINDINGS

4.1. Descriptive analysis of customer journey

Table 4 consolidates descriptive information regarding customer journeys for each evaluated period.

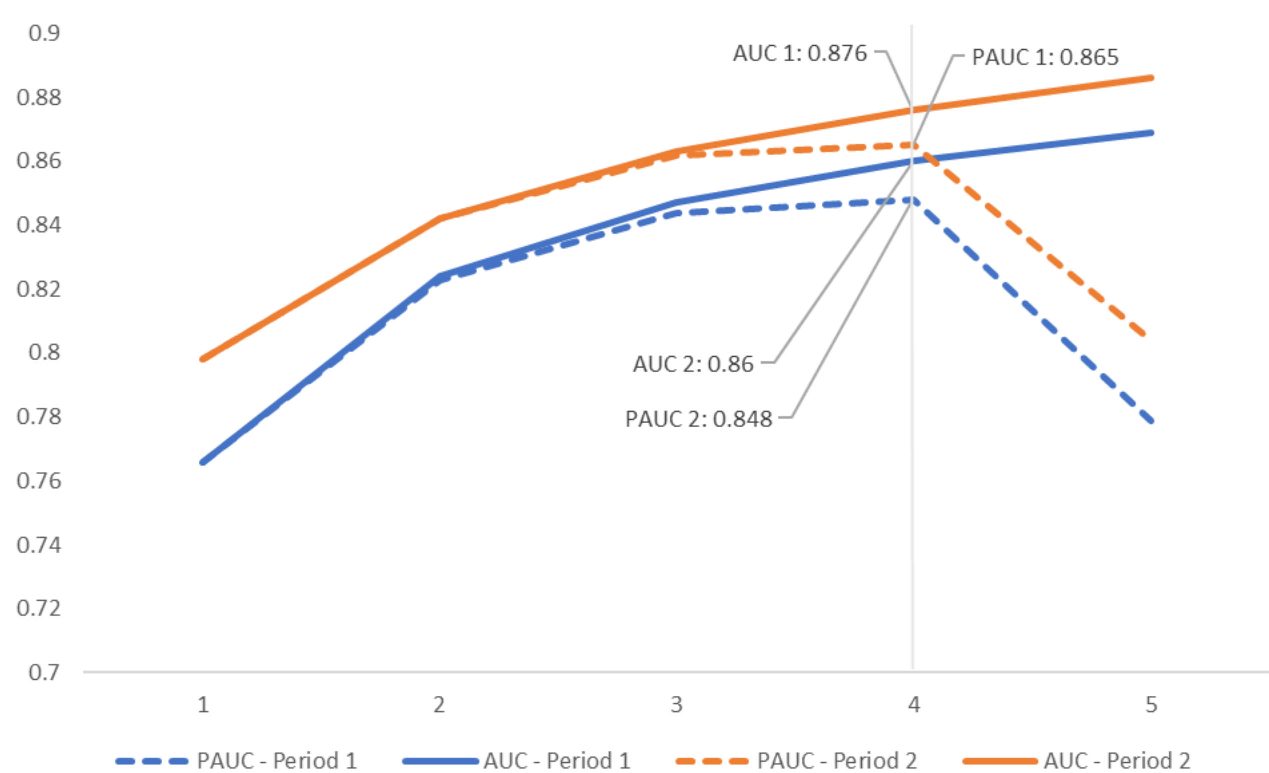

Figure 2. Area under the curve (AUC) and Penalized AUC for each dataset using different Markov Orders.

Table 4. Description of the student journeys data sets.

Description	Period 1	Period 2	% var
Corpus	Semesters: 2019-S2 and 2020-S1	Semesters: 2020-S2 and 2021-S1	-
Number of channels	17	17	-
Number of touchpoints	293,012	369,826	+26.22
in pre-purchase stage	228,252 (77,90%)	253,288 (68.49%)	+10.97
in purchase stage	52,618 (17,96%)	93,834 (25.37%)	+78.33
in post-purchase stage	12,142 (4,14%)	22,704 (6.14%)	+86.99
Share by type of channel			
Branded owner channels	86.47%	80.35%	+26.76
Partner owner channels	12.85%	19.58%	+107.97
Social	0.68%	0.06%	-87.07
Share by contact initiative			
Customer Initiated Contacts (CIC)	66.61%	58.55%	+19.92
Firm Initiated Contacts (FIC)	33.39%	41.45%	+69.34
Number of journeys	95,715	89,916	-6.06
Journey length (touchpoints)	$\bar{x}$ 3.09 (s29.60)	$\bar{x}$ 4.14 (s2 13.81)	+33.98
Duration of journey (months)	$\bar{x}$ 2.06 (s24.79)	$\bar{x}$ 3.17 (s2 6.22)	+53.88
Time until conversion (months)	$\bar{x}$ 3.88 (s25.25)	$\bar{x}$ 7.84 (s29.02)	+102.06
Number of conversions (enrollments)	3,113	2,529	-18.76
Journey conversion rate	3.25%	2.81%	-13.54

Period 2 recorded 6% fewer journeys compared to Period 1; however, it exhibited 26% more touchpoints. Journeys in Period 2 also consumed more time. Notably, journeys resulting in conversions had an average duration of 7.84 months, which is twice the duration observed in Period 1 (3.88 months).

It can be inferred that the pandemic period had an impact on the decision-making process. The HESJ, typically spanning extended periods (James-MacEachern, 2018), even further extended during the evaluated journeys in Period 2. While most touchpoints traditionally occur in the pre-purchase phase, there is a noticeable increase in contacts during the purchase and post-purchase phases in Period 2.

For instance, the average journey duration until conversion doubled from 3.88 months in Period 1 to 7.84 months in Period 2. This extended decision-making timeline was accompanied by a clear shift toward digital channels. Touchpoint volume from instant messaging grew by 217%, email by 106%, and social media (specifically Instagram) by 117%. While these figures represent a correlation rather than direct causation, they strongly suggest an adaptation to the social isolation and increased digitalization that characterized the pandemic period.

The data also reveal that brand-owned channels account for most touchpoints (86.47% in Period 1 and 80.35% in Period 2). However, partner-owned touchpoints saw a substantial increase in Period 2 (107.97%), primarily attributed to the growth in touchpoints on the Instagram channel. It is important to note that the boundaries between these classes may become less distinct when involving technologically enabled channels (Lemon & Verhoef, 2016).

Touchpoints related to FIC and CIC exhibited an increase in both periods, particularly for FIC, which experienced a growth of 63.34% in Period 2, mainly due to an increase in touchpoints through the Email, Sales, and Instant Messages channels. As observed by de Haan et al. (2016), CIC recorded a higher volume of touchpoints than FIC in both periods (66.61% in Period 1 and 58.55% in Period 2). Buhalis and Volchek

(2021) emphasize the significance of customer-initiated communications, which generally exert a substantial influence on the final consumer decision.

4.2. Touchpoints

The findings revealed significant variations in the distribution of touchpoints per channel across the two periods. Notable changes include a decrease in contacts per journey and share of contacts for inbound channels (-23.74%), application channels (-21.55%), and call center (-13.99%). Conversely, there were notable increases in contacts per journey and share of contacts for email (106.66%), social IG (115.17%), promotion (92.24%), and sales channels (173.79%). Some channels experienced relatively minor changes, such as live chat (15.39%) and SEO (63.36%), while others saw declines, such as events (-34.97%) and direct channels (-15.98%). Messages experienced a substantial increase of 217.80%, while affiliate channels showed a significant decrease of-87.07%.

Additionally, between the two periods, the number of touchpoints increased by 26.22%. However, this increase impacted each channel differently, altering the share of contacts for each one. Notably, touchpoints in the inbound channel, present in at least 50% of the journeys, experienced an overall decrease of 23.74%. This reduction is directly linked to a significant decrease in the implementation of the marketing action known as “Vocational Tests,” which is part of the inbound channel. The Sales department typically utilizes these tests not only in online promotions but also in face-to-face events, such as visits to secondary schools. Unfortunately, these events did not take place in Period 2 due to the COVID-19 pandemic.

In Period 2, 89% of the journeys featured at least one email-related contact, with emails witnessing a substantial increase of 106% in volume. XYZ University primarily employs the Email channel in conjunction with Inbound Marketing, sending emails to users interested in specific programs or students actively involved in some application or subscription process.

The Call-Center channel, crucial in both volume and results, experienced a 14% decrease in phone contacts between periods 1 and 2. This decline is mainly attributed to the reduction in applications (as most calls occur in the purchase phase) and a slight decrease in the success rate of phone calls, in line with observations made by de Haan et al. (2016), who noted a growing aversion to this type of contact among consumers.

The analysis reveals contrasting trends in touchpoint distribution across various channels during the observed periods. While traditional phone contacts witnessed a decline, alternative channels like live chat and instant messaging saw substantial growth, with a notable 15 and 200% increase, respectively, managed by the call center team. Notably, Instagram’s Social IG channel experienced a remarkable growth of 115% between the two periods, emerging as the second-largest channel in terms of contact volume in Period 2. This surge may reflect the impact of social isolation measures during the pandemic, as suggested by Tankovska (2020).

The pandemic-induced shift to virtual platforms is evident in the significant decrease of 34% in touchpoints for the events channel due to the cancellation of in-person gatherings. However, online events saw a remarkable 174% increase in contacts, highlighting the adaptation to digital alternatives. Additionally, channels like Promotion and Sales witnessed growth in touchpoints, likely driven by increased investment in activities associated with these channels by XYZ University. These findings underscore the dynamic nature of customer engagement strategies, which have evolved in response to changing circumstances and preferences, particularly considering the pandemic’s disruptions to traditional modes of interaction.

4.3. Carryover and spillover effects

The order in which customers experience touchpoints could have “synergic” or “antagonistic” effects on conversions. Previous interactions may have overlapping effects on subsequent touchpoints (Buhalis & Volchek, 2021), referred to as “carryover” and “spillover” effects.

Carryover effects occur when one channel receives sequential contacts, often influenced by consumer preferences for that channel or other circumstantial aspects of the customer journey. Minor carryover effects are noticeable on the following channels: Affiliate, Direct, Email, Events, Inbound, Referral, and Social Others (Youtube and Facebook). Moderate carryover effects are observed on Application, Call-Center, Live-Chat, and Sales, indicating sequential contacts with these channels by students. A strong carryover effect is evident on Instagram (Social_IG), reflecting a behavior where students engage with various brand posts during the same visit.

Spillover effects occur when one channel guides the consumer to another channel. For instance, minor spillover is noted from Online Advertising (AD) to Inbound and Promotion channels. However, substantial effects from AD to channels like Inbound, Promotion, and Application would be expected due to landing pages set in these campaigns.

Application has a minor spillover effect on email, where the initiation of the application results in the student receiving more emails related to application stages and the program. Direct and Referral channels exhibit a strong spillover effect on Inbound, possibly due to offline links, links in partner websites, and links in social media (typically identified as direct in mobile traffic). The Email channel has a minor spillover effect on Inbound, as most emails suggest links to content.

Messages have a moderate effect on Call-Center, possibly because they are used to reach students when they do not answer phone calls. Messages also have a minor effect on Email, which may occur when users receive more calls and messages (during the purchase phase), leading to more automated emails.

Sales have minor effects on Call-Center, confirming some entanglement among channels like Call-Center, Emails, and Sales. SEO demonstrates substantial effects on Inbound, primarily driven by the Blog and E-books content that receives significant traffic from SEO. Minor effects on Email are observed due to newsletter subscriptions or marketing automation after downloading some content. Links on Facebook and Youtube (Social Others channel) also have a moderate effect on Inbound and minor effects on Promotion.

4.4. The attribution model findings

Our analysis unveils the critical role specific channels played in driving student enrollment at the Brazilian HEI under study. Over 70% of conversions stem from just five channels: Emails, Live Chat, Call Center, Sales, and Inbound. These channels exhibit consistent performance across two analyzed periods, suggesting their strategic placement within the student journey. In Table 5 the results are showcased by contrasting the percentage of conversions ascribed to each channel utilizing the Markov model with two conventional rule-based methods: First Touch and Last Touch. The presented percentages illustrate the conversions allocated to each channel, accompanied by supplementary details on removal effects (% RE).

The effectiveness of Emails, Call Centers, and Sales likely arises from their proximity to the enrollment decision. These touchpoints cater to students nearing the final stages, offering crucial support and addressing last-minute inquiries that solidify enrollment choices. Conversely, Live Chat functions as a student-initiated touchpoint, ideal for those who have already identified their educational needs and actively seek information (de Haan et al., 2016).

The Inbound channel plays a vital role by guiding students throughout their entire journey, from initial career exploration to starting classes. This holistic approach aligns with Kumar et al. (2018)'s findings on the significant impact of firm-generated content (offered through Inbound channels) on conversions. The Inbound channel likely synergizes with email, social media, and SEO channels, fostering brand awareness and nurturing leads throughout the decision-making process.

To further explore these dynamics, as suggested by the distinction between CIC and FIC, we can aggregate the attribution results. Channels primarily driven by customer initiative include Live Chat, Inbound, Social IG, Direct, and SEO. Conversely, channels like Email, Call Center, Sales, Messages, and Ads are predominantly firm-initiated. An analysis of the Period 2 data reveals a compelling insight: while CICs constitute the majority of touchpoints by volume (58.55% as shown in Table 4), the FIC channels collectively account for a larger share of the attributed conversions (approximately 61 vs. 39% for CICs). This highlights a critical tension: the student journey is largely driven by the prospect's own exploration, but strategic, firm-initiated interventions at key moments are disproportionately effective in securing the final conversion.

Beyond this core group of consistently high-performing channels, the findings also reveal a notable dynamic among a secondary set of touchpoints. A group of digitally-centric channels (including Social IG, Promotion, Messages, and Ads) saw their attributed contribution to conversions grow between the two periods. Conversely, channels more reliant on partnerships or in-person engagement, such as Affiliate

Table 5. Results for the attribution model.

Channels	Period 1			Period 2		
	Markov graph (4th) (%)	First Touch (%)	Last Touch (%)	Markov graph (4th) (%)	First Touch (%)	Last Touch (%)
EMAIL	19.61	26.32	19.72	17.22	20.53	13.81
LIVE CHAT	12.93	15.14	14.37	14.71	16.61	16.55
CALL CENTER	17.41	13.09	31.19	14.14	13.37	15.72
SALES	9.23	5.39	7.96	12.91	7.46	18.61
INBOUND	13.57	12.58	9.06	12.13	15.44	6.22
SOCIAL IG	2.71	1.17	5.89	5.71	2.14	16.07
PROMOTION	4.14	3.35	4.05	5.41	3.85	4.44
AD	3.74	5.11	1.35	4.71	7.03	1.89
MESSAGES	3.64	1.14	2.36	4.17	1.83	4.49
DIRECT	3.85	4.36	0.00	2.86	4.29	0.00
SEO	2.34	3.36	0.00	2.29	3.00	0.00
EVENTS	2.81	3.60	2.37	2.10	3.00	1.98
OTHERS	0.45	0.20	0.16	0.70	0.21	0.16
AFFILIATE	3.04	4.72	1.52	0.46	0.77	0.06
SOCIAL OTHERS	0.23	0.28	0.00	0.30	0.26	0.00
REFERRAL	0.31	0.19	0.00	0.18	0.22	0.00

programs, Events, and Referrals, received a diminishing portion of the attribution. This divergence among lower-attribution channels suggests an adaptive shift in the HEI's marketing tactics, possibly influenced by evolving student preferences or the operational constraints of Period 2, which coincided with the COVID-19 pandemic.

The Markov model offers a significant advantage over traditional First Touch and Last Touch models by providing a more balanced view of channel impact. First Touch models tend to overestimate the importance of initial interactions (Email, Inbound) while undervaluing Sales and Messages—crucial touchpoints closer to conversion. Conversely, Last Touch models overemphasize the impact of final interactions (sales, social media). The Markov model addresses these limitations by considering the sequential nature of touchpoints and their cumulative influence on enrollment decisions.

An important consideration when interpreting attribution models is the potential endogeneity of firm-initiated touchpoints (Kannan et al., 2016). CRM and marketing automation systems strategically target students with a higher propensity to enroll. This targeted approach results in increased FIC for students further along the journey, potentially influencing attribution results.

Brand-owned channels consistently hold a dominant position in driving conversions. This highlights the importance of a strong brand presence across various touchpoints. Interestingly, social touchpoints associated with marketing actions involving digital influencers (Affiliate channel) exhibited a decrease. Despite this decline, the Affiliate channel still achieved results comparable to online advertising or events in the initial period, suggesting its potential effectiveness with a smaller investment.

A practical tool for quantifying channel impact within the Markov chain attribution model is the "Removal Effects" metric (Anderl et al., 2016; Archak et al., 2010). This metric estimates the impact of removing specific channels from the customer journey on the conversion rate. For instance, removing the email channel could potentially lead to a dramatic drop in conversions. This hypothetical scenario offers a quantifiable measure of each channel's significance, providing marketers with a clear understanding of how individual channels contribute to the overall enrollment success.

In conclusion, this analysis demonstrates the effectiveness of a data-driven approach to quantifying channel impact within the student enrollment journey. By identifying key channels and their influence throughout the journey, HEIs can optimize marketing strategies and resource allocation to maximize enrollment outcomes. The insights gleaned from the Markov chain model, particularly the removal effects metric, empower marketers to prioritize channels with the most significant impact on student enrollment.

FINAL REMARKS

This research introduces a novel data-driven attribution model that leverages individual customer journey data to assess the effectiveness of marketing channels and assign value to various touchpoints. Conducted as a case study of a Brazilian HEI, the research analyzes a substantial dataset—662,838 online and offline touchpoints from 185,631 customer journeys.

The analysis yields valuable insights into two key areas of consumer journey research, which in turn provide a clear foundation for the managerial contributions discussed subsequently:

1. Tangible Marketing Performance: The study reveals that Email, Live Chat, Call Center, Sales, and Inbound channels are the primary drivers of enrollment, collectively accounting for over 70% of conversions across two analyzed periods. This aligns with de Haan et al. (2016)'s findings on the prominence of CICs compared to FICs. The dominance of brand-owned touchpoints (over 80% in Period 2) underscores the importance of a strong brand presence throughout the student journey;
2. Evolving Customer Journeys: By comparing results from two periods, the study revealed that the average customer journey duration leading to enrollment increased significantly (from 3.8 to 7.8 months), and online channels like Instant Messaging (+217%), Social Media (+117%), and Email (+106%) experienced a substantial rise in touchpoint volume.

The research strongly advocates for data-driven multi-touch attribution models as a superior method for evaluating touchpoints across the customer journey and their impact on conversion rates. This approach aligns with Tueanrat et al. (2021), who emphasize the value of data-driven methods in understanding consumer behavior dynamics along the path to conversion. These models, like the one employed in this study, share common features:

- Inclusion of multiple touchpoints;
- Individual-level data utilization;
- Consideration of overlapping touchpoint effects;
- Dynamic touchpoint value allocation through custom computational techniques;
- Potential for cross-channel and cross-platform analysis.

The study further demonstrates the limitations of single-touch models by comparing their results to the multi-touch model. As documented in prior research, single-touch models yield inaccurate results, especially for complex and lengthy customer journeys. The model in this study, drawing on data from thousands of individual journeys, offers a reliable picture of touchpoint interactions based on actual consumer behavior. This approach provides empirical support for the argument that data-driven models are powerful tools for enhancing our understanding of the Customer Journey.

Additionally, the findings are aligned with Kannan and Kulkarni (2022), demonstrating the increasingly shifting to online channels as replacements for traditional offline purchases due the extended period of pandemic restrictions. This adaptation has fostered the development of new decision marking routines, ultimately reducing the barriers to accessing online and mobile channels.

The primary contribution of this research to the field lies in demonstrating that models based on data collected throughout the customer journey—and which account for all marketing touchpoints—achieve superior performance in identifying the specific contribution of each channel.

Managerial contributions

Integrate the customer journey into strategic planning

Marketing and sales strategies should be structured around the customer journey, which encompasses pre-purchase, purchase, and post-purchase interactions. This long-term perspective enables institutions to foster meaningful relationships with prospective students. The customer journey should be understood as part of a broader relationship marketing framework, incorporating principles such as customer experience, satisfaction, loyalty, and customer centricity. These concepts collectively enhance the effectiveness of marketing initiatives and support higher enrollment conversion rates.

Systematically map and monitor touchpoints

Effective customer journey management requires early identification and mapping of key touchpoints and phases. Institutions must determine where and how customers interact with the brand, who initiates contact, and which channels are under institutional control. Implementing mechanisms to register and analyze these interactions—such as CRM systems—enables the collection of user-level data. These data is essential for understanding channel usage, journey duration, and touchpoint sequences, thereby facilitating more accurate and actionable insights. Furthermore, an attribution model makes these insights more powerful by prioritizing data collection. It pinpoints the most valuable data sources (such as *Live Chat* and *Inbound* in our study), ensuring that monitoring resources are concentrated on the channels that truly drive the journey.

Optimize and evolve through iterative improvements

Customer journey management is an ongoing, iterative process that demands continuous refinement. Each touchpoint influences customer perception and can have synergistic or antagonistic effects on conversion outcomes. Therefore, touchpoints should be planned to deliver a coherent, cross-channel experience. Attribution models, when grounded in a well-defined customer journey framework, can provide valuable insights into touchpoint effectiveness. However, without a solid theoretical foundation and structured data, such models risk producing misleading or non-actionable results. Institutions should aim for incremental improvements, integrating technical and organizational capabilities over time. A robust model makes this improvement process manageable and data-driven. It identifies the handful of channels driving the majority of conversions (as the five in our study did), allowing managers to focus high-impact optimization efforts—such as A/B testing, script refinement, or UX improvements—precisely where they will yield the greatest return.

Research limitations and future studies

The research acknowledges limitations related to the contextual interpretation of Markov chain models and the generalizability of findings to other HEIs. Future investigations could explore simplified Shapley Value algorithms (Mahboobi et al., 2018) or delve into machine learning approaches (Li et al., 2018; Pattanayak et al., 2022). Additionally, the model could be extended to examine the relationship between students and the HEI brand, analyzing the influence of touchpoints on loyalty, brand engagement, and word-of-mouth recommendations. Bergamo et al. (2010) highlight that the HEI customer journey is shaped not only by utilitarian factors but also by brand perception and relational bonds. Future studies could integrate attitudinal data based on self-reported surveys (Kumar et al., 2018) to capture these relational aspects. By incorporating these refinements and expanding the scope of analysis, this research paves the way for a more comprehensive understanding of student enrollment journeys and empowers HEIs to optimize marketing strategies for maximizing enrollment success.

REFERENCES

Abhishek, V., Fader, P., & Hosanagar, K. (2012). Media exposure through the funnel: a model of multistage attribution. *SSRN*. https://doi.org/10.2139/ssrn.2158421

Abreu, W. F., Araújo, R. S., & Lima, J. P. C. (2019). Políticas de financiamento e expansão no Ensino Superior do Brasil. *Quaestio - Revista de Estudos em Educação*, *21*(2), 583-609. https://doi.org/10.22483/2177-5796.2019v21n2p583-609

Alblas, R. (2018). *Attribution modeling customer journey online*. Retrieved from https://research.tue.nl/en/studentTheses/attribution-modeling

Altomare, D., & Loris, D. (2016). Python Package "ChannelAttribution. Retrieved from https://pypi.org/project/ChannelAttribution/

Anderl, E., Becker, I., von Wangenheim, F., & Schumann, J. H. (2013). Putting attribution to work: a graph-based framework for attribution modeling in managerial practice. *SSRN Electronic Journal*, 1-36. https://doi.org/10.2139/ssrn.2343077

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. https://doi.org/10.1016/j.ijresmar.2016.03.001

Archak, N., Mirrokni, V. S., & Muthukrishnan, S. (2010). Mining Advertiser-specific User Behavior Using Adfactors Categories and Subject Descriptors. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 31-40). Association for Computing Machinery, Raleigh.

Bergamo, F., Ponchio, M., Zambaldi, F., Giuliani, A. C., & Spers, E. (2010). De prospect a aluno: fatores influenciadores da escolha de uma instituição de ensino superior. *Base – Revista de Administração e Contabilidade da Unisinos*, *7*(3), 182-193. https://doi.org/10.4013/base.2010.73.01

Berman, R. (2018). Beyond the last touch: Attribution in online advertising. *Marketing Science*, *37*(5), 771-792. https://doi.org/10.1287/mksc.2018.1104

Bolat, E., & O'Sullivan, H. (2017). Radicalising the marketing of higher education: learning from student-generated social media data. *Journal of Marketing Management*, *33*(9-10), 742-763. https://doi.org/10.1080/0267257X.2017.1328458

Buhalis, D., & Volchek, K. (2021). Bridging marketing theory and big data analytics: The taxonomy of marketing attribution. *International Journal of Information Management*, *56*, 102253. https://doi.org/10.1016/j.ijinfomgt.2020.102253

Chapman, D. W. (1981). A model of student college choice. *The Journal of Higher Education*, *52*(5), 490. https://doi.org/10.2307/1981837

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C, & Wirth, R. (2000). Crisp-Dm 1.0. *CRISP-DM Consortium* (Vol. 76).

Conselho Administrativo de Defesa Econômica (CADE) (2016). *Atos de concentração no mercado de prestação de serviços de ensino superior*. CADE. Retrieved from https://cdn.cade.gov.br/Portal/centrais-de-conteudo/publicacoes/estudos-economicos/cadernos-do-cade/atos-de-concentracao-no-mercado-de-prestacao-de-servicos-de-ensino-superior-2016.pdf

de Haan, E., Wiesel, T., & Pauwels, K. (2016). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. International *Journal of Research in Marketing*, *33*(3), 491-507. https://doi.org/10.1016/j.ijresmar.2015.12.001

Følstad, A., & Kvale, K. (2018). Customer journeys: a systematic literature review. *Journal of Service Theory and Practice*, *28*(2), 196-227. https://doi.org/10.1108/JSTP-11-2014-0261

Gai, L., Xu, C., & Pelton, L. E. (2016). A netnographic analysis of prospective international students' decision-making process: implications for institutional branding of American universities in the emerging markets. *Journal of Marketing for Higher Education*, *26*(2), 181-198. https://doi.org/10.1080/08841241.2016.1245233

Galan, M., Lawley, M., & Clements, M. (2015). Social media's use in postgraduate students' decision-making journey: an exploratory study. *Journal of Marketing for Higher Education*, *25*(2), 287-312. https://doi.org/10.1080/08841241.2015.1083512

Hanson, K. H., & Litten, L. H. (1982). Mapping the road to academe: a review of research on women, men, and the college selection process. In P. J. Perun (Ed.), *The undergraduate woman: Issues in education equity* (pp. 73-98). Lexington Books.

Hossler, D., & Gallagher, K. S. (1987). Studying student college choice: a three-phase model and the implications for policymakers. *College and University*, *62*(3), 207-221.

James-MacEachern, M. (2018). A comparative study of international recruitment–tensions and opportunities in institutional recruitment practice. *Journal of Marketing for Higher Education*, *28*(2), 247-265. https://doi.org/10.1080/08841241.2018.1471014

James-MacEachern, M., & Yun, D. (2017). Exploring factors influencing international students' decision to choose a higher education institution: A comparison between Chinese and other students. *International Journal of Educational Management*, *31*(3), 343-363. https://doi.org/10.1108/IJEM-11-2015-0158

Kakalejč, L., Bucko, J., Resende, P., & Ferencova, M. (2018). Multichannel marketing attribution using Markov chains. *Journal of Applied Management and Investments*, *7*(1), 49-60.

Kannan, P. K., & Kulkarni, G. (2022). The impact of Covid-19 on customer journeys: implications for interactive marketing. *Journal of Research in Interactive Marketing*, *16*(1), 22-36. https://doi.org/10.1108/JRIM-03-2021-0078

Kannan, P. K., & Li, H. A. (2017). Digital marketing: A framework, review and research agenda. *International Journal of Research in Marketing*, *34*(1), 22-45. https://doi.org/10.1016/j.ijresmar.2016.11.006

Kannan, P. K., Reinartz, W., & Verhoef, P. C. (2016). The path to purchase and attribution modeling: Introduction to special section. *International Journal of Research in Marketing*, *33*(3), 449-456. https://doi.org/10.1016/j.ijresmar.2016.07.001

Kireyev, P., Pauwels, K., & Gupta, S. (2016). Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing*, *33*(3), 475-490. https://doi.org/10.1016/j.ijresmar.2015.09.007

Kotler, P. (1976). A role for marketing in college admissions. In *Applying marketing theory to college admissions* (pp. 54-72). College Entrance Examination Board.

Kumar, A., Bezawada, R., Rishika, R., Janakiraman, R., & Kannan, P. K. (2018). The effects of firm generated content (FGC) in social media on customer behavior. *Journal of Marketing*, *53*(9), 1689-1699.

Lemon, K. N., & Verhoef, P. C. (2016). Understanding customer experience throughout the customer journey. *Journal of Marketing*, *80*(6), 69-96. https://doi.org/10.1509/jm.15.0420

Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: an empirical model and a field experiment. *Journal of Marketing Research*, 51(1), 40-56. https://doi.org/10.1509/jmr.13.0050

Li, N., Arava, S. K., Dong, C., Yan, Z., & Pani, A. (2018). Deep neural net with attention for multi-channel multi-touch attribution. *AdKDD*. https://doi.org/10.48550/arXiv.1809.02230

Mahboobi, S. H., Usta, M., & Bagheri, S. R. (2018). Coalition game theory in attribution modeling: Measuring what matters at scale. *Journal of Advertising Research*, *58*(4), 414-422. https://doi.org/10.2501/JAR-2018-014

Maringe, F. (2006). University and course choice: Implications for positioning, recruitment and marketing. *International Journal of Educational Management*, *20*(6), 466-479. https://doi.org/10.1108/09513540610683711

McColl-Kennedy, J. R., Zaki, M., Lemon, K. N., Urmetzer, F., & Neely, A. (2019). Gaining customer experience insights that matter. *Journal of Service Research*, *22*(1), 8-26. https://doi.org/10.1177/1094670518812182

Nascimento, R. L. S., Cruz Junior, G. G., & Fagundes, R. A. A. (2018). Mineração de dados educacionais: um estudo sobre indicadores da educação em bases de dados do INEP. *RENOTE*, *16*(1), 1-11. https://doi.org/10.22456/1679-1916.85989

Nichols, W. (2013). Advertising Analytics 2.0. *Harvard Business Review*, *91*(3), 60-68.

Ordenes, F. V., Theodoulidis, B., Burton, J., Gruber, T., & Zaki, M. (2014). Analyzing customer experience feedback using text mining: a linguistics-based approach. *Journal of Service Research*, *17*(3), 278-295. https://doi.org/10.1177/1094670514524625

Pattanayak, S., Pati, P. B., & Singh, T. (2022). Performance analysis of machine learning algorithms on multi-touch attribution model. In *3rd International Conference for Emerging Technology* (pp. 1-7). IEEE, Belgaum. https://doi.org/10.1109/INCET54531.2022.9824865

Piurcosky, F. P., Calegário, C. L. L., Silva, G., Frogeri, R. F., & Portugal, N. S. (2019). Fusões e aquisições no mercado de ensino superior: uma revisão sistemática. *Brazilian Journal of Development*, *5*(9), 15738-15751. https://doi.org/10.34117/bjdv5n9-146

Rosenbaum, M. S., Otalora, M. L., & Ramírez, G. C. (2017). How to create a realistic customer journey map. *Business Horizons*, *60*(1), 143-150. https://doi.org/10.1016/j.bushor.2016.09.010

Schuhbauer, H., Brockmann, P., & Mustafayev, T. (2020). Mapping the students' journey to develop student-centered tools. In *IEEE Global Engineering Education Conference* (pp. 56-60). https://doi.org/10.1109/EDUCON45650.2020.9125139

Senhoras, E. M., Takeuchi, K. P., & Takeuchi, K. P. (2012). A análise estrutural do ensino superior privado sob perspectiva. In *Simpósio de Excelência em Gestão e Tecnologia* (pp. 1-12). Universidade Federal Fluminense.

Shao, X., & Li, L. (2011). Data-driven multi-touch attribution models. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 258-264). Association for Computing Machinery. https://doi.org/10.1145/2020408.2020453

Shearer, C., Watson, H. J., Grecich, D. G., Moss, L., Adelman, S., Hammer, K., & Herdlein, S. A. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13-22. Retrieved from https://www.scirp.org/(S(vtj3fa45qm1ean45vvffcz55))/reference/ReferencesPapers.aspx?ReferenceID=1592780

Tankovska, H. (2020). Share of social media users in the United States who believe they will use select social media more if confined at home due to the coronavirus as of March 2020. *Statista*. Retrieved from https://www.statista.com/statistics/1106343/social-usage-increase-due-to-coronavirus-home-usa/

Tueanrat, Y., Papagiannidis, S., & Alamanos, E. (2021). Going on a journey: A review of the customer journey literature. *Journal of Business Research*, *125*, 336-353. https://doi.org/10.1016/j.jbusres.2020.12.028

Vescovi, P. V. S. (2020). *Análise preditiva na detecção de evasão de alunos no ensino superior privado brasileiro: abordagem de algoritmos de aprendizado de máquina, com base nas perspectivas acadêmicas, financeiras, geográficas e socioeconômicas* [dissertation, Fundação Getúlio Vargas]. Retrieved from https://bibliotecadigital.fgv.br/dspace/handle/10438/28907

Wiesel, T., Pauwels, K., & Arts, J. (2010). Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science*, *30*(4), 604-611. https://doi.org/10.1287/mksc.1100.0612

Yin, R. K. (2018). *Case study research and applications: Design and methods*. SAGE Publications.

Zhao, K., Mahboobi, S. H., & Bagheri, S. R. (2019). Revenue-based attribution modeling for online advertising. *International Journal of Market Research*, *61*(2), 195-209. https://doi.org/10.1177/1470785318774447

How to cite this article:

Santos, R F., & Almeida, L. F. (2025). Which touchpoints matter most? A data-driven model for understanding student journey. *Internext*, 20(3), e866. https://doi.org/10.18568/internext.v20i03.866