

Universidade de São Paulo – USP

Escola de Engenharia de São Carlos – EESC

Departamento de Engenharia Elétrica e de Computação

BRENO R. S. NASCIMENTO

Modelos de atribuição para mídias digitais com uso de
Cadeias de Markov: Um estudo de caso da Agência de
Marketing Digital

Trabalho de Conclusão de Curso apresentado à Escola de Engenharia de
São Carlos, da Universidade de São Paulo

Curso de Engenharia Elétrica com ênfase em eletrônica

ORIENTADOR: PROF. MARCO TERRA

São Carlos, 2017

Modelos de atribuição para mídias digitais com uso de Cadeias de Markov: Um estudo de caso da Agência de Marketing Digital

BRENO R. S. NASCIMENTO

Trabalho de Conclusão de Curso apresentado à Escola de Engenharia de São Carlos da Universidade de São Paulo como parte dos requisitos para obtenção do título de Engenheiro Eletricista com ênfase em Eletrônica.

ÁREA DE CONCENTRAÇÃO: Cadeias de Markov e sua aplicação em marketing digital.

ORIENTADOR: Prof. Marco Terra

São Carlos, 2017

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes e Seção Técnica de Informática, EESC/USP com os dados fornecidos pelo(a) autor(a).

N244m NASCIMENTO, BRENO RENATO SARAVALLI NASCIMENTO
Modelos de atribuição para mídias digitais com uso de Cadeias de Markov: Um estudo de caso da Agência de Marketing Digital / BRENO RENATO SARAVALLI NASCIMENTO NASCIMENTO; orientador MARCO HENRIQUE TERRA. São Carlos, 2017.

Monografia (Graduação em Engenharia Elétrica com ênfase em Eletrônica) -- Escola de Engenharia de São Carlos da Universidade de São Paulo, 2017.

1. CADEIAS DE MARKOV. 2. MARKETING DIGITAL. 3. MODELO DE ATRIBUIÇÃO. I. Título.

Bibliotecário responsável pela estrutura de catalogação da publicação:

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE APROVAÇÃO

Nome: Breno Renato Saravalli Nascimento

Título: “Modelos de atribuição para mídias digitais com uso de Cadeias de Markov: Um estudo de caso da Agência de Marketing Digital”

Trabalho de Conclusão de Curso defendido e aprovado
em 29/11/2017,

com NOTA 8,3 (Oito, três), pela Comissão Julgadora:

Prof. Titular Marco Henrique Terra - Orientador - SEL/EESC/USP

Prof. Assistente Carlos Goldenberg - SEL/EESC/USP

Dra. Daiane Cristina Bortolin - SEL/EESC/USP

Coordenador da CoC-Engenharia Elétrica - EESC/USP:
Prof. Associado Rogério Andrade Flauzino

DEDICATÓRIA

Dedico este trabalho a meus pais e amigos, especialmente ao meu pai que sempre me deu o incentivo necessário.

AGRADECIMENTOS

Primeiramente, agradeço à minha família que me deu toda força e suporte para fazer este trabalho.

Agradeço também aos meus amigos por toda ajuda e motivação, em especial Bruno C. Melo.

Agradecimento especial aos fundadores da Raccoon Marketing Digital Ltda, Marco Tulio Kehdi e André Palis. Também agradeço a Leonardo Araújo, diretor da Raccoon Marketing Digital Ltda, que muito contribuiu para o meu crescimento profissional ao longo dos últimos três anos.

EPÍGRAFE

“Nada acontece de repente e de repente tudo acontece”

(Juliana Stein)

RESUMO

Nascimento, B. R. Saravalli. Modelos de atribuição para mídias digitais com uso de Cadeias de Markov: Um estudo de caso da Agência de Marketing Digital – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2017.

O tema proposto tem uma abordagem sobre como as cadeias de Markov podem ser aplicadas em marketing digital. O estudo foi feito em uma empresa cliente de uma agência de Marketing Digital de médio porte administrada pelos seus fundadores desde o seu surgimento. Por se tratar de uma empresa real, não serão divulgados valores absolutos dos dados da empresa, o que nada influencia nas conclusões obtidas. Foi constatado que o comportamento de uma pessoa em um determinado site pode mudar dependendo de diversos fatores que precedem essa interação. Uma análise qualitativa mostra que a probabilidade do usuário tomar uma determinada ação no site pode ser diferente dependendo de diversos fatores como o dia da semana, o gênero e a idade do usuário, o caminho pelas mídias que o levou até o site e seus pontos de contato, entre outros. Esse estudo analisa a importância de cada ponto de contato existente no caminho que o usuário percorre até a chegada em um determinado site. Um modelo de atribuição define regras que ajudam a metrificar qual o peso em que cada ponto de contato nesse caminho influencia na ação final do usuário no site. Com auxílio do uso das Cadeias de Markov, é possível definir um modelo de atribuição mais sensível às necessidades da empresa anunciante. É possível maximizar a receita de uma empresa no ambiente online tornando o investimento nas mídias online mais eficiente. Esse conceito pode ser chamado de marketing de performance, que visa o foco em resultados. Se isolarmos algum dos fatores citados que influenciam na decisão do usuário, podemos definir cadeias de estados e probabilidades, afim de desenvolver o modelo citado, buscando maior eficiência nos resultados finais. Para isso, algumas ferramentas online do Google foram usadas, tais como o Google Analytics e Google Adwords. Além disso, vamos discorrer sobre o que são cadeias de Markov, com uma pequena introdução teórica, vista que o foco do estudo é em como aplicá-las no desenvolvimento de um modelo de atribuição. Também será mostrado as principais funções das ferramentas citadas e como contribuíram para a coleta e análise dos dados obtidos. Com isso, é desenvolvido um modelo que permite dar a peso a cada ponto de

contato na jornada do consumidor, afim de buscar os resultados esperados pela empresa e agência.

Palavras chave: Cadeias de Markov, marketing digital, modelo de atribuição para mídias digitais.

ABSTRACT

Nascimento, B. R. Saravalli. Assignment models for digital media using Markov Chains: A case study of the Digital Marketing Agency - School of Engineering of São Carlos, Universidade de São Paulo, 2017.

The proposed theme takes an approach on how Markov chains can be applied in digital marketing. The study was done in a client company of a medium-sized Digital Marketing agency managed by its founders since its inception. Because it is a real company, absolute values of the company data will not be disclosed, which has no influence on the conclusions obtained. It has been found that a person's behavior on a particular site may change depending on several factors that precede this interaction. A qualitative analysis shows that the user's likelihood of taking a particular action on the site may be different depending on various factors such as the day of the week, the gender and age of the user, the path through the media that led to the site and its contact, among others. This study analyzes the importance of each contact point in the path that the user travels until the contact with a particular site. An attribution model defines rules that help to quantify the weight at which each point of contact in that path influences the final action of the user on the site. With the help of Markov Chains, you can define an attribution model that is more responsive to the needs of the advertiser. It is possible to maximize a company's revenue in the online environment by making investing in online media more efficient. This concept can be called performance marketing, which focuses on results. If we isolate some of the factors that influence the decision of the user, we can define chains of states and probabilities, in order to develop the cited model, seeking greater efficiency in the final results. For this, some online tools from Google have been used, such as Google Analytics and Google Adwords. In addition, we will discuss what are Markov chains, with a small theoretical introduction, seeing that the focus of the study is on how to apply them in the development of an attribution model. It will also be shown the main functions of the mentioned tools and how they contributed to the collection and analysis of the data obtained. With this, a model is developed that allows giving weight to each point of contact in the consumer journey, in order to seek the results expected by the company and agency.

Keywords: Markov Chains, digital marketing, attribution model for digital media.

LISTA DE FIGURAS

Figura 1 - Busca no Google	19
Figura 2 - Anúncio da Rede de Display.....	20
Figura 3 - Tipos de anúncio em vídeo.....	21
Figura 4 - Anúncio para dispositivo móvel.....	22
Figura 5 - CPC real (fonte: https://marketingdeconteudo.com/google-adwords).....	24
Figura 6 - Funil de conversão 1	30
Figura 7 - Funil de conversão 2	31
Figura 8 - Diagrama de Estados 1	38
Figura 9 - Diagrama de Estados com efeito de remoção para o canal 1	39
Figura 10 - Diagrama de Estados 3	41
Figura 11 - Diagrama de Estados	45
Figura 12 - Diagrama de Estados	47
Figura 13 - Dados obtidos do Google Analytics	49
Figura 14- Diagrama de Estados Final.....	53

LISTA DE QUADROS

Tabela 1 – Cadeia de Markov com divisão em pares.....	36
Tabela 2 – Canais.....	45
Tabela 3 - Canais	46
Tabela 4 - Total de conversões.....	47
Tabela 5- Transições	50
Tabela 6 - Tabela de transição, contagem e probabilidade	52

LISTA DE SIGLAS

CPC – Custo por clique.

PPC – Pay per click

KPI – Key performance indicator

ROI – Return of investment

SUMÁRIO

1. INTRODUÇÃO.....	16
1.1 Contextualização	16
1.2 Objetivos	16
1.3 Estrutura do trabalho	17
2. CADEIAS DE MARKOV.....	17
2.1 Processo Estocástico	17
2.2 Definição de Cadeia de Markov.....	18
3. CONTEXTUALIZAÇÃO EM MARKETING DIGITAL.....	19
3.1 Tipos de anúncio – Google Adwords	19
1) Rede de Display ou GDN (Google Display Network)	20
2) Anúncios no Youtube	21
3) Anúncio para aplicativo	22
3.2 Leilão do Google Adwords	23
3.3 Custo por clique	24
3.4 Custo por clique máximo e custo por clique real.....	24
3.5 Índice de qualidade	25
3.6 Indicadores de Performance.....	25
3.7 Remarketing	27
3.7.1 Cookies	27
3.7.2 Como funciona o Remarketing	27
3.7.3 Algumas das vantagens do remarketing do Google Adwords.....	28
3.7.4 Maneiras de usar o remarketing com o Google Adwords.....	29
3.8 Funil de conversão.....	29
3.9 Tipos de negócios.....	31
4. MODELOS DE ATRIBUIÇÃO	32
4.1 Visão geral do modelo de atribuição.....	32

4.2 Modelo de atribuição multi canal	34
4.3 Modelo de atribuição no conceito de Cadeias de Markov	35
4.4 Uso do Google Analytics para definição do Modelo de Markov	40
4.5 A atribuição de canais R <i>package</i>	42
5. ESTUDO DE CASO	43
5.1 Introdução	43
5.2 Coleta de dados	44
5.3 Modelo seguido	44
5.4 Modelos de acompanhamento	48
5.5 Extraindo dados do Google Analytics	48
6. CONSIDERAÇÕES FINAIS	54
7. REFERÊNCIAS	55

1. INTRODUÇÃO

1.1 Contextualização

Modelos de atribuição têm como objetivo monitorar o caminho do consumidor desde que ele teve um primeiro contato com o site do anunciante (REZ. 2016). Com um bom modelo de atribuição, é possível dar peso e importância a cada etapa ou canal desse caminho. A maioria dos modelos atuais não representa fielmente o verdadeiro peso de cada etapa.

Se um consumidor tem seu primeiro contato com um site através de um anúncio via email marketing, seguido de um contato por um anúncio do Google e, por consequência compra de algum produto no site, no modelo mais utilizado, a compra será atribuída para a última interação, ou seja, o anúncio do Google.

Dessa forma, o primeiro contato que o consumidor teve com o site não é levado em consideração, mas sabe-se que esse contato tem extrema importância, pois foi a porta de entrada ao site para esse consumidor.

1.2 Objetivos

Este trabalho propõe uma alternativa de modelo de atribuição, baseado no conceito das Cadeias de Markov. É desenvolvido um modelo que permite entender qual a importância de cada canal e etapa no caminho do consumidor. Com isso, é possível tomar melhores decisões sobre o investimento a ser feito nas mídias digitais. Pode-se dizer que com um modelo coerente, os resultados obtidos tornam-se mais eficientes. O objetivo final, assim com os resultados buscados já são pré-definidos pelo anunciante.

1.3 Estrutura do trabalho

O trabalho apresenta-se da seguinte forma: no Capítulo 2 são introduzidos os conceitos de Processo Estocástico e Cadeias de Markov. No desenvolvimento do modelo de atribuição, esses conceitos serão utilizados.

Em seguida, é necessária uma contextualização em marketing digital, para que o leitor compreenda os termos e processos utilizados no dia a dia. É importante o conhecimento de algumas estratégias de marketing, visto que após a definição do modelo, decisões devem ser tomadas em busca do objetivo do anunciante. Outro ponto abordado, são os tipos de anúncio e conceitos sobre a jornada do consumidor.

Uma situação real é estudada para o desenvolvimento do modelo de atribuição. Nesse momento, são utilizados o conhecimento em marketing digital e os conceitos abordados sobre as cadeias de Markov. Quando o modelo já estiver definido, um estudo é feito para saber o impacto que cada mídia possui no caminho do consumidor. Assim, é possível analisar se, para esse caso estudado, é possível aumentar a eficiência dos resultados finais obtidos.

2. CADEIAS DE MARKOV

2.1 Processo Estocástico

Processo estocástico é uma coleção de variáveis aleatórias que, em geral, são utilizadas para estudar a evolução de fenômenos (ou sistemas) que são observados ao longo do tempo. Assim, ao invés de descrever o sistema através de equações determinísticas (como, equações diferenciais ordinárias), que dado uma condição inicial conhecemos toda a evolução do sistema, utiliza-se processos estocásticos, para o qual, dado uma condição inicial, ainda temos diversas trajetórias possíveis para a evolução do sistema (Karlin; Taylor, 1998).

2.2 Definição de Cadeia de Markov

A perda de memória é a base da caracterização das cadeias de Markov e ela estabelece que em um conjunto de estados discretos o futuro só depende do estado presente, ou seja, os estados anteriores são irrelevantes para a predição dos estados seguintes, desde que o estado atual seja conhecido (Norris, 1998).

Em termos de probabilidades, uma cadeia de Markov a tempo discreto com espaço de estados S é um processo estocástico $\{X_n\}_{n \in T}$, onde $T = \{0, 1, 2, \dots\}$, tal que se verificam as seguintes propriedades:

- Para qualquer $i \in S$ tem-se:

$$P(X_0 = i) = P_i$$

- Para quaisquer $i, j \in S$, e $n \in T$: $P(X_{n+1} = j | X_n = i) = P_{ij}$.
- Para quaisquer $n \in T$ e $i_0, i_1, \dots, i_{n-1}, i, j \in S$, vale a condição: $P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i)$.

Além disso, quando $X_n = i$ diz-se que o X_n está no instante n . Em especial, a terceira propriedade nos diz que dado o presente (X_n), o futuro (X_{n+1}) e o passado (X_0, X_1, \dots, X_{n-1}) são independentes.

Exemplo: Foram observados alguns dados do time de futebol Flamengo. Ele nunca empata dois jogos seguidos. Se ele empata, as probabilidades de ele ganhar ou perder o próximo jogo são iguais. Se a vitória ocorreu no jogo atual, com a empolgação, a probabilidade de ganhar ou empatar no próximo jogo é de $1/2$ e $1/4$, respectivamente. Se a derrota vier no jogo atual, a probabilidade de ganhar na próxima partida diminui para $1/4$ e a de perder novamente aumenta para $1/2$. Perceba que as probabilidades de resultado da partida atual, não depende de resultados anteriores ao atual. Então, é possível modelar uma cadeia de Markov.

3. CONTEXTUALIZAÇÃO EM MARKETING DIGITAL

Para o completo entendimento do estudo proposto, é necessário que o leitor fique familiarizado com os conceitos e ferramentas de marketing digital utilizados.

3.1 Tipos de anúncio – Google Adwords

Existem diversos tipos de anúncios do google tais como a Figura 1 demonstra.

Figura 1 - Busca no Google

The image shows a Google search interface for the query "sapatos femininos". The search bar shows the query and the number of results (approximately 5,440,000). Below the search bar, there are navigation tabs for "Todas", "Shopping", "Imagens", "Notícias", "Vídeos", "Mais", "Configurações", and "Ferramentas".

The search results are categorized into three sections:

- Google Shopping:** A red-bordered box highlights a grid of six shoe products with images, titles, and prices. The products are: Sapatilha Laura Prado (R\$129,90), Tênis Converse (R\$99,90), Sapato Metalizado (R\$119,00), Tênis Polo US 7007 (R\$129,99), AMARO Feminino (R\$179,90), and Sandália Vizzano (R\$99,90).
- Google Search:** A green-bordered box highlights two search results. The first is from amaro.com, advertising "Sapatos Femininos | Comprar Calçados Online" with a price of R\$199. The second is from kanui.com.br, advertising "Calçados Femininos - Compre Já Online".
- Busca orgânica:** A blue-bordered box highlights an organic search result from Mercado Livre Brasil, titled "Sapatos Feminino - Sapatos Femininos no Mercado Livre Brasil".

Os anúncios do Google são criados por uma plataforma chamada Google Adwords. Esses anúncios são pagos pelo anunciante, no formato CPC (custo por

clique), ou seja, é cobrado uma certa quantia, em reais, a cada vez que acontece um clique no anúncio.

O resultado da busca orgânica não necessita de investimento para exibição dos anúncios. O posicionamento dos anúncios na busca orgânica envolve fatores diferentes dos anúncios pagos.

O posicionamento dos anúncios pagos é decido por um leilão, gerenciado pelo Google. Os critérios dos leilões mudam dependendo do tipo de anúncio. Assim, os critérios que definem a posição dos anúncios de Google Shopping diferem daqueles para os anúncios do resultado de Pesquisa (Google Search). Na Figura 2, temos mais alguns tipos de anúncios do Google.

1) Rede de Display ou GDN (Google Display Network)

Figura 2 - Anúncio da Rede de Display

CIRCULAÇÃO 307.530 exemplares
 PÁGINAS VISTAS 200.371.162 + (fev.17)
 VISITANTES ÚNICOS 26.437.489 + (fev.17)

Google Display Network

PUBLICIDADE
 Anúncios Google

LAREIRA NO INVERNO
 CONFORTO E SEGURANÇA

POLYTEC
 CHUVEIRO, BARRIL, BARRIL, BARRIL

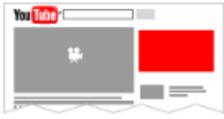
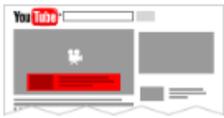
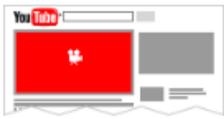
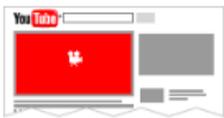
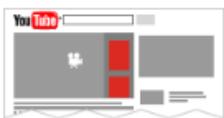
VALOR M
Refor
limite
 Cada pr
 ingresso

imp

2) Anúncios no Youtube

Na Figura 3, há exemplos de anúncios para Youtube.

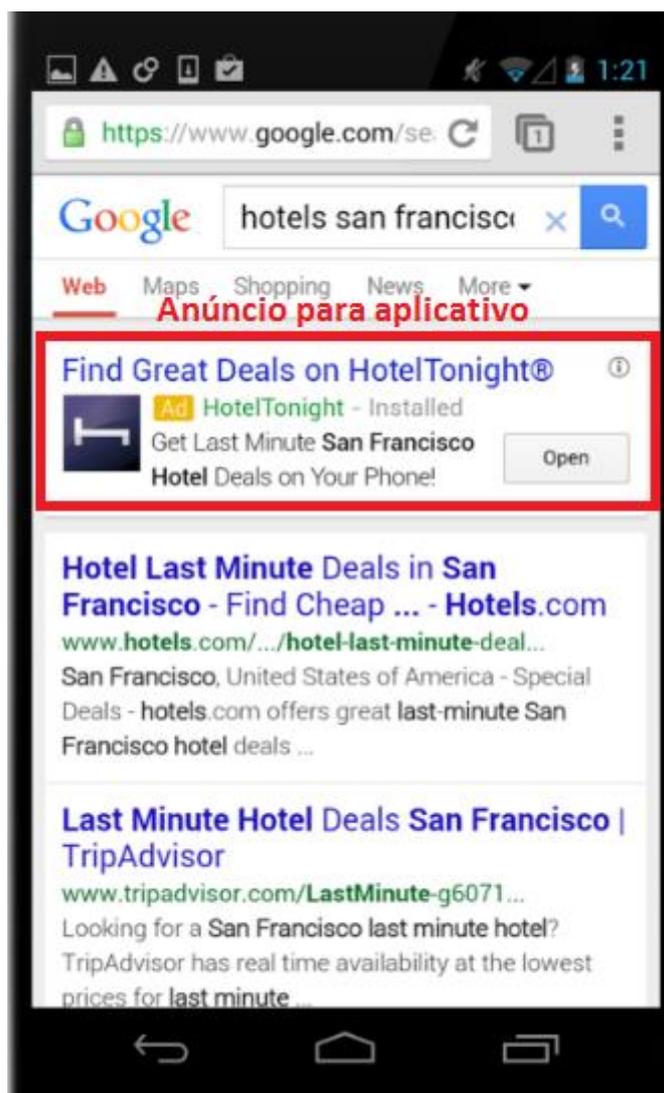
Figura 3 - Tipos de anúncio em vídeo

Formato do anúncio	Posicionamento	Plataforma	Especificações
Anúncios gráficos 	Aparece à direita do vídeo em destaque e acima da lista de sugestões de vídeo. Para players maiores, esse anúncio pode ser exibido abaixo do player.	Computador	300 x 250 ou 300 x 60
Anúncios de sobreposição 	Anúncios de sobreposição semitransparentes que são exibidos nos 20% da parte inferior do seu vídeo.	Computador	Anúncios gráficos ou de texto de 468 x 60 ou 728 x 90
Anúncios em vídeo puláveis 	Com os anúncios em vídeo puláveis, os espectadores podem pular anúncios depois de cinco segundos, se quiserem. Inseridos antes, durante ou depois do vídeo principal.	Computador, dispositivos móveis, TV e consoles de jogos	É reproduzido no player de vídeo.
Anúncios em vídeo impossíveis de pular e anúncios em vídeo longos e impossíveis de pular 	É necessário assistir a anúncios em vídeo impossíveis de pular para que o vídeo seja exibido. Os anúncios em vídeo longos e não ignoráveis podem ter até 30 segundos de duração. Esses anúncios podem ser exibidos antes, durante ou depois do vídeo principal.	Computador e dispositivos móveis	É reproduzido no player de vídeo. Possuem 15 ou 20 segundos de duração, dependendo dos padrões regionais. Os anúncios longos e não puláveis podem ter até 30 segundos de duração.
Anúncios curtos 	Anúncios em vídeo não puláveis de até seis segundos que precisam ser assistidos antes de visualizar o vídeo.	Computador e dispositivos móveis	Reproduzidos no player de vídeo, com até seis segundos de duração
Cartões patrocinados 	Os cartões patrocinados mostram conteúdo que pode ser relevante para seu vídeo, como produtos exibidos no vídeo. Os espectadores verão um teaser do cartão por alguns segundos. Eles também poderão clicar no ícone no canto superior direito do vídeo para navegar pelos cartões.	Computador e dispositivos móveis	Os tamanhos de cartões variam

3) Anúncio para dispositivo móvel

A Figura 4 mostra um exemplo de anúncio para dispositivo móvel.

Figura 4 - Anúncio para dispositivo móvel



Esses são os principais tipos de anúncio da rede Google. Dentro de cada tipo de anúncio, o Google oferece diferentes produtos. Por exemplo, existem tipos diferentes de campanhas de marketing dentro da rede de pesquisa, com diferentes formas de segmentações daquelas na rede de display.

Nesse texto, o foco será nos anúncios da rede de pesquisa e de display do Google. Serão feitas análises avançadas e o desenvolvimento de um modelo que possa representar um processo real.

3.2 Leilão do Google Adwords

Os conceitos abaixo são baseados no suporte do Google.

A cada vez que uma busca é realizada no Google, acontece um leilão para definir qual anúncio será exibido. Pela plataforma Google Adwords, o anunciante define algumas palavras chaves que vão disparar determinados anúncios. Para ir ao leilão, o anúncio precisa estar qualificado para exibição. A qualificação do anúncio depende de fatores que precisam estar de acordo com as normas e leis do país em questão. No Brasil, não é permitido anúncios de produtos com tabacos ou jogos de azar, por exemplo.

Sempre que um anúncio do Google Adwords está qualificado para exibição em uma pesquisa, ele passa pelo leilão de anúncios quando um termo de pesquisa dispara a palavra-chave escolhida pelo anunciante. Por exemplo, se o anunciante comprar a palavra-chave “chocolate”, o termo de pesquisa “chocolate” dispara essa palavra chave.

O leilão determina se o anúncio é de fato exibido e em qual posição será impresso no resultado da pesquisa.

O leilão funciona da seguinte forma: quando alguém pesquisa, o sistema do Google Adwords encontra todos os anúncios qualificados cujas palavras-chave correspondem à pesquisa.

A partir desses anúncios, o sistema ignora todos aqueles que não são qualificados, como anúncios reprovados ou segmentados para outra região.

Dentre os anúncios qualificados, apenas aqueles com uma classificação do anúncio suficientemente elevada podem ser exibidos. A classificação do anúncio é uma combinação do seu lance com a qualidade do seu anúncio e o impacto esperado de extensões e outros formatos de anúncio.

O ponto mais importante a ser lembrado é: mesmo que os lances da concorrência sejam mais alto, ainda é possível conquistar uma posição de maior destaque (por um preço menor) com palavras-chave e anúncios altamente relevantes.

Como o processo de leilão é repetido a cada pesquisa no Google, cada leilão pode ter resultados potencialmente diferentes, dependendo da concorrência no momento. Desse modo, é normal observar alguma oscilação na posição do anúncio na página e na possibilidade de exibição do anúncio.

3.3 Custo por clique

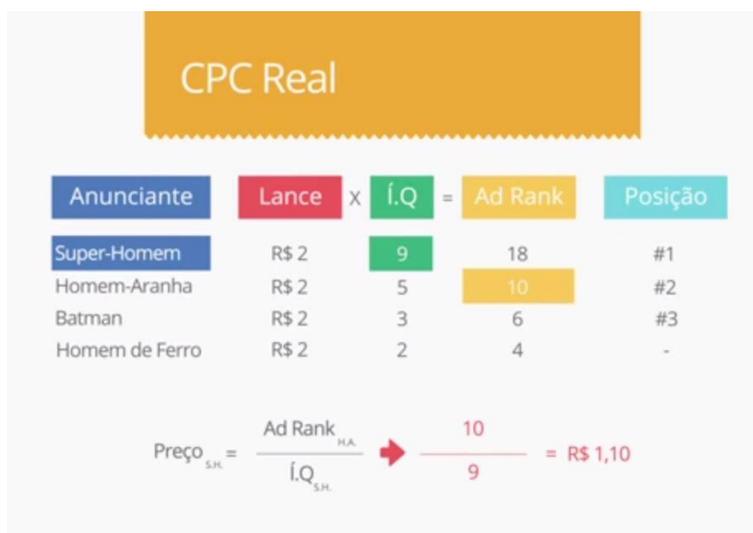
O CPC (custo por clique) é o valor que será pago para cada clique dado no anúncio. O usuário clica e é direcionado para a página de destino definida pelo anunciante.

3.4 Custo por clique máximo e custo por clique real

O anunciante dá um lance máximo que irá pagar por clique. Quando ocorre o leilão, o valor real do clique é definido sendo igual ou menor ao lance máximo definido.

A posição do anúncio é definida pela classificação que cada anúncio obtém no leilão. Um cálculo é feito para definir a classificação de cada anúncio. Esse cálculo envolve o índice de qualidade do anúncio (I.Q) e o CPC real como é mostrado na Figura 5.

Figura 5 - CPC real (fonte: <https://marketingdeconteudo.com/google-adwords>)



3.5 Índice de qualidade

“Cálculos de qualidade em tempo real e específicos do leilão para a taxa de cliques esperada, a relevância do anúncio e a experiência na página de destino são usados para definir a classificação do anúncio no leilão.”

Em resumo, para se obter melhores posições nos anúncios, é preciso ter um lance competitivo à concorrência e um bom índice de qualidade do anúncio.

O conteúdo da página de destino deve ser relevante para a busca feita, assim como o próprio anúncio. A taxa de clique por impressão (CTR, do inglês *click through rate*) é um dos parâmetros que compõe o índice de qualidade. O CTR é a razão de número de cliques por impressões. Então, quanto maior é a CTR maior é a taxa de clique em um anúncio.

Voltando ao exemplo da palavra-chave “chocolate” descrito no capítulo 3.2. Quando alguém abre a busca do Google e procura por chocolate, um anúncio sobre vendas de chocolate tem uma taxa de cliques (CTR) esperada maior do que um anúncio de sapatos, para um mesmo número de impressões. Dessa forma, a ferramenta entende que o anúncio de chocolate é mais relevante para a busca do que aquele sobre sapatos.

Em suma, quanto maior a CTR, maior a relevância do anúncio. No caso oposto, se há uma busca por “sapatos femininos” e dois anúncios são exibidos, um de sapatos femininos e outro de uma loja de chocolate, o primeiro anúncio terá mais cliques do que o segundo para um mesmo número de impressões (CTR maior), pois sua relevância é maior do que a do segundo.

3.6 Indicadores de Performance

Os KPI's (indicadores de performance, do inglês *key performance indicator*) são os indicadores que auxiliam na mensuração das iniciativas de Marketing Digital: número de visitas, número de páginas por visita, quantidade de visitantes que retornam ao site, quantidade de assinantes do site, quantidade de leads gerados pelo site, entre dezenas de outros indicadores possíveis.

Atrair mais visitas para um site não gera, necessariamente, mais negócios para uma empresa. Os indicadores são os números que ajudarão a analisar o ROI – Return Over Investment, ou Retorno Sobre o Investimento em marketing. (REZ, 2016)

Para marketing digital, alguns KPIs são essenciais para o alcance dos objetivos das campanhas.

Alguns dos principais KPIs são:

Investimento: valor investido em dinheiro.

Impressões: quantidade de vezes que o anúncio é qualificado e exibido nos leilões.

Cliques: quantidade de cliques que determinado anúncio recebe ao ser exibido.

CTR: razão de cliques por impressões.

Conversões: definida pelo anunciante. Uma venda pode ser considerada conversão, assim como um preenchimento de formulário, download de aplicativo, etc.

CPA: custo por aquisição ou custo por conversão. É a razão do investimento pelo número de conversões. Ou seja, o valor investido para cada conversão recebida.

Taxa de conversão de cliques: razão de conversões por cliques.

Receita: valor gerado em dinheiro.

ROI (retorno sobre investimento do inglês *Return Over Investimento*): retorno sobre o investimento. Razão da receita pelo investimento. É importante saber que é comum o uso da sigla ROI no lugar de ROAS (do inglês *return over advertisement spent*). O verdadeiro ROI inclui outros custos, como o custo de produção do produto por exemplo. Para o cálculo do ROAS, é considerado apenas o investimento em anúncios como custo e a receita gerada. Mas, em geral, o termo ROI também é utilizado para calcular a receita sobre o investimento em anúncios.

Posição média: posição média dos anúncios no resultado da pesquisa.

CPC médio: razão do custo pelo número de cliques.

Parcela de impressão na rede de pesquisa: a porcentagem de impressões que os anúncios recebem em comparação com o número total de impressões que eles poderiam receber.

3.7 Remarketing

Para melhor entendimento do que é remarketing, vamos definir o que são cookies.

3.7.1 Cookies

Cada vez que alguém visita um site, algumas de suas informações básicas ficam armazenadas em seu computador no formato de texto. Esses arquivos simples de texto são chamados de cookies. Por exemplo, a maioria dos sites armazenam endereços IP (protocolo de internet, do inglês *internet protocol*) e preferências sobre idiomas, cores, etc. do usuário. Assim, quando essa pessoa retorna ao site, suas preferências são aplicadas de forma automática.

3.7.2 Como funciona o Remarketing

Sites que usam remarketing contém uma tag de remarketing em seu código fonte. A tag de remarketing do Google Adwords é um pedaço de código no site que é usado para salvar os IDs dos cookies de quem já visitou esse site. Quando as pessoas acessam essa página, o ID do cookie delas é adicionado à lista de remarketing do Google Adwords.

O remarketing possibilita ao anunciante se reconectar com pessoas que já acessaram seu site já ou usaram seu aplicativo para dispositivos móveis por exemplo. Quando uma pessoa abandona um site sem realizar uma conversão, como comprar algo, por exemplo, o remarketing auxilia na reexibição de anúncios relevantes para essas pessoas.

O objetivo do anunciante pode ser aumentar as inscrições, as vendas, os downloads de aplicativos ou promover o conhecimento da sua marca. Para qualquer um desses, o remarketing deve ser um componente estratégico da publicidade. Se bem estruturado, ele impulsiona o ROI para todos os tipos de anunciantes. Para anunciantes que definem inscrições como sendo as conversões em seu site, o remarketing pode ajudar na redução do CPA para esse anunciante. Ou seja, se antes era investido X reais

para cada inscrição, esse valor X é reduzido, deixando mais barato o CPA (custo por aquisição, veja página 23).

O remarketing busca alcançar as pessoas quando elas estão mais propensas a realizar uma conversão. É possível alcançá-las quando estiverem pesquisando seu produto ou até acessando outros sites. Quando uma pessoa entra em um site de pescaria e por vários dias fica vendo anúncios em banners sobre pesca em outros sites, isso é o remarketing impactando-a. Com o remarketing do Google Adwords, é usado pontos de contato oportunos para atrair os clientes ao seu site quando eles estão mais engajados.

3.7.3 Algumas das vantagens do remarketing do Google Adwords

Listas personalizadas: o anunciante pode criar listas personalizadas para as metas que definiu de publicidade. Por exemplo, criar uma lista “Abandono de carrinho – 1 dia”. Nessa lista, estarão pessoas que entraram no site até um dia atrás, adicionaram algum produto ao carrinho, mas não finalizaram a compra.

Alcance em larga escala: o anunciante pode alcançar pessoas nas suas listas de remarketing enquanto elas usam seus dispositivos para navegar em mais de 2 milhões de websites e aplicativos para dispositivos móveis.

Sistema de preços eficientes: é possível criar campanhas de remarketing de alto desempenho com estratégias de lances automáticos, como CPA e ROAS desejados. Os lances em tempo real calculam o lance ideal para a pessoa que visualiza seu anúncio, o que ajuda você a vencer o leilão de anúncios com o melhor preço possível. Não há custo adicional para usar o leilão do Google. O anunciante pode exibir anúncios de produtos com desconto para pessoas que já visitaram seu site, por exemplo.

Anúncios dinâmicos: com as campanhas de remarketing dinâmico, é possível criar anúncios dinâmicos que combinam seu “feed” de produtos com os “layouts” da Galeria de anúncios, o que resulta em anúncios criados automaticamente de acordo com os produtos listados, com base no interesse de cada usuário.

3.7.4 Maneiras de usar o remarketing com o Google Adwords

Remarketing padrão: exibe anúncios a visitantes anteriores enquanto eles navegam em websites e usam aplicativos da Rede de Display.

Remarketing dinâmico: impulsiona os resultados com o remarketing dinâmico, que eleva o remarketing a outro nível com anúncios de produtos ou serviços que as pessoas já viram no website do anunciante.

Lista de remarketing para anúncios de texto: exibe anúncios a visitantes anteriores enquanto eles continuam pesquisando o que precisam no Google, depois de sair do website do anunciante.

(GOOGLE, 2016).

3.8 Funil de conversão

Um funil de conversão é um caminho por onde os clientes potenciais passam e tem como resultado final uma conversão.

No site, por causa da quantidade de tráfego (visitantes) que recebe, esse número é maior do que a quantidade que converte no final, assim obtém-se o formato de funil como apresentado na Figura 6.

Figura 6 - Funil de conversão 1



Como na Figura 6, um site pode ter muitos visitantes. Quando um usuário começa a ter mais engajamento com o site, como deixando seu email por exemplo, esse visitante se torna um lead. Esse lead deve ser trabalhado para que se torne um “prospect” e depois um “lead” qualificado. Enfim, aquele visitante torna-se um cliente e realiza uma conversão.

Remarketing é uma forma de nutrir os visitantes e “leads” já gerados no site. No topo de um funil de interesses e ações, existem pessoas que estão no estágio de conhecimento da marca (“awareness”) e consideração. Depois da consideração, a pessoa passa para avaliação. Por fim, a compra.

Figura 7 - Funil de conversão 2



Cada tipo de anúncio do Google tem um impacto diferente sobre cada visitante em determinada etapa do funil. Podemos chamar que cada tipo de anúncio é um canal diferente. Por exemplo, os canais de display impactam pessoas que não necessariamente estão buscando por aquele produto naquele momento, diferentemente dos canais da rede de pesquisa, onde o usuário já está buscando sobre algum serviço ou produto.

O remarketing funciona bem o fundo do funil, pois é direcionado para pessoas que já demonstraram algum tipo de interesse naquele serviço ou produto do anúncio que sendo impactado.

3.9 Tipos de negócios

As estratégias de marketing devem estar completamente alinhadas a cada tipo de negócio. Estratégias para uma empresa de varejo não são as mesmas para uma empresa B2B (do inglês, "business to business"). Enquanto o varejo tem como objetivo final a venda de um produto, a empresa B2B pode ter como métrica de conversão a captação de um email.

4. Modelos de atribuição

4.1 VISÃO GERAL DO MODELO DE ATRIBUIÇÃO

O modelo de atribuição tem como objetivo atribuir crédito de vendas e conversões a cada ponto de contato / canal no caminho de conversão.

Um modelo de atribuição é a regra ou conjunto de regras que determina como o crédito de vendas e as conversões são atribuídas aos canais nos caminhos de conversão. Por exemplo, o modelo “última interação” no Google Analytics atribui 100% de crédito aos pontos de contato finais (ou seja, cliques) que precedem imediatamente vendas ou conversões. Por outro lado, o modelo primeira interação atribui 100% de crédito aos pontos de contato que iniciam caminhos de conversão. Isso pode ser um problema, porque não analisa a importância exata de cada canal no caminho de conversão.

É possível usar a ferramenta de comparação de modelos para comparar como diferentes modelos de atribuição afetam a valorização dos canais de marketing. Na ferramenta, o valor de conversão calculado (e o número de conversões) para cada um dos canais de marketing varia de acordo com o modelo de atribuição utilizado. Um canal que principalmente inicia caminhos de conversão terá um valor de conversão maior de acordo com o modelo de atribuição primeira interação do que teria de acordo com o modelo de atribuição última interação.

EXEMPLO DE MODELO DE ATRIBUIÇÃO

Um cliente encontra um site clicando em um dos anúncios do Google Adwords. Ele volta uma semana depois clicando em uma rede social. Nesse mesmo dia, ela volta pela terceira vez por meio de uma campanha de e-mail e, algumas horas mais tarde, retorna de novo diretamente e realiza uma compra.

O Google Analytics fornece sete modelos de atribuição:



No modelo de atribuição Última interação, o último ponto de contato (nesse caso, o canal Direto) receberia 100% do crédito da venda.



No modelo de atribuição Último clique não direto, todo o tráfego direto é ignorado, e 100% do crédito da venda vai para o último canal por meio do qual o cliente chegou ao site antes de concluir a conversão (neste caso, o canal E-mail).



No modelo de atribuição Último clique do Google Adwords, o último clique do Google Adwords (nesse caso, o primeiro e único clique no canal Pesquisa paga) receberia 100% do crédito da venda.



No modelo de atribuição Primeira interação, o primeiro ponto de contato (nesse caso, o canal Pesquisa paga), receberia 100% do crédito da venda.



No modelo de atribuição Linear, cada ponto de contato no caminho de conversão (nesse caso, os canais Pesquisa paga, Rede social, E-mail e Direto) dividiriam a mesma porcentagem de crédito (25% cada) para a venda.



No modelo de atribuição Redução de tempo, os pontos de contato mais próximos em termos de tempo da venda ou conversão recebem a maior parte do crédito. Nessa venda específica, os canais Direto e E-mail receberiam a maior parte do crédito porque o cliente interagiu com eles dentro de algumas horas da conversão. O canal Rede social receberia menos crédito do que os canais Direto ou E-mail. Como a interação da pesquisa paga ocorreu uma semana antes, esse canal receberia um crédito significativamente menor.



No modelo de atribuição Com base na posição, 40% do crédito é atribuído a cada primeira e última interação, e os 20% de crédito restantes são distribuídos

uniformemente para as interações intermediárias. Nesse exemplo, os canais Pesquisa paga e Direto receberiam 40% do crédito, enquanto os canais Rede social e E-mail receberiam 10% do crédito.

4.2 MODELO DE ATRIBUIÇÃO MULTI CANAL

(KAELIN, 2016)

Como dito anteriormente, um cliente geralmente passa por um caminho ou sequência de diferentes canais ou pontos de contato antes de uma compra em comércio eletrônico ou conversão em outras áreas. No Google Analytics, é possível encontrar alguns pontos de contato mais propensos a auxiliar a conversão do que outros que mais propensos a ser o “touchpoint” do último clique.

Como a maioria dos canais são pagos (em termos de dinheiro ou tempo gasto), é vital ter um algoritmo para distribuir conversões e o valor entre esses canais e comparar com seus custos em vez de creditar, por exemplo o último canal não-direto apenas. Este é um problema do Modelo de Atribuição Multi-Channel. A definição já citada do Google Analytics ajuda: um modelo de atribuição é uma regra ou conjunto de regras que determina como o crédito para vendas e conversões é atribuído a pontos de contato em caminhos de conversão.

Atualmente, o Google Analytics fornece sete modelos de atribuição predefinidos e até mesmo um modelo personalizado que pode ser adaptado. Esse é um campo muito interessante para analistas e comerciantes. Porém, no Google Analytics, é necessário tomar uma decisão de escolha ou gerencial sobre o modelo a ser usado e o porquê.

Pode-se ver diferentes resultados com diferentes modelos, mas dificilmente o anunciante sabe qual é o mais adequado para seu negócio. Em outras palavras, o Google Analytics fornece modelos heurísticos com seus prós e contras. O maior contra é o que foi já foi citado, nem sempre o anunciante sabe qual o melhor modelo de atribuição a ser usado.

As vantagens do Google Analytics: não é necessário organizar um armazenamento e infra-estrutura para coletar dados, pois possui uma vasta variedade de modelos heurísticos, além de ser fácil e gratuito para uso. Portanto, para uma empresa relativamente pequena, seria lógico usar a abordagem do Google Analytics, mas se os resultados da atribuição teriam um impacto significativo nos orçamentos de

marketing, preços dos produtos, compreensão das viagens dos clientes, etc., é possível explorar outras ideias.

O uso de cadeias de Markov é uma delas. Em breve, estudaremos aspectos práticos de sua implementação.

4.3 MODELO DE ATRIBUIÇÃO NO CONCEITO DE CADEIAS DE MARKOV

Um modelo a partir da Cadeia de Markov determina a probabilidade de um usuário passar da Sequência A para a Sequência B com base nas etapas que cada usuário leva através de um site. Cada sequência exibe um caminho diferente do usuário.

Então, o uso de cadeias Markov nos permite mudar de modelos heurísticos para probabilistas. Pode-se representar todas as passagens de clientes (sequência de canais ou pontos de contato) como uma cadeia em um gráfico de Markov direcionado, em que cada vértice é um possível estado (canal ou ponto de contato) e as bordas representam a probabilidade de transição entre os estados (incluindo a conversão).

Calculando o modelo e estimando probabilidades de transição, pode-se atribuir cada canal / ponto de contato. É possível começar com um exemplo simples do gráfico de Markov, de primeira ordem ou sem memória, para uma melhor compreensão do conceito.

Por exemplo, os caminhos dos clientes contêm três canais únicos C1, C2 e C3. Além disso, deve-se adicionar manualmente três estados especiais a cada gráfico: (início), (conversão) e (nulo). Esses estados adicionais representam o ponto de partida, compra ou conversão e conversão malsucedida. Transições de canais idênticos são possíveis (por exemplo, C1 -> C1), mas podem ser omitidos por diferentes motivos.

Nesse caso, o uso do termo “canais” pode ser entendido como as diferentes mídias, como as citadas na Seção 3.8.

Assumindo três canais na jornada do cliente:

C1 -> C2 -> C3 -> compra

C1 -> conversão malsucedida

C2 -> C3 -> conversão malsucedida

Após isso, é necessário fazer uma transformação e, em seguida, dividir as transições em pares, como na Tabela 1 e 2. Na transformação, são adicionados os estados “conversion” para caracterizar uma compra, “start” para o início da jornada e “null” para demonstrar uma compra malsucedida.

Tabela 1 – Cadeia de Markov com divisão em pares

Jornada do consumidor	Transformação	Divido em pares
C1 -> C2 -> C3 -> purchase	(start) -> C1 -> C2 -> C3 -> (conversion)	(start) -> C1, C1 -> C2, C2 -> C3, C3 -> (conversion)
C1	(start) -> C1 -> (null)	(start) -> C1, C1 -> (null)
C2 -> C3	(start) -> C2 -> C3 -> (null)	(start) -> C2, C2 -> C3, C3 -> (null)

Em seguida, é possível calcular as probabilidades da transição de estado para estado, contabilizando o número de ocorrência de cada transição.

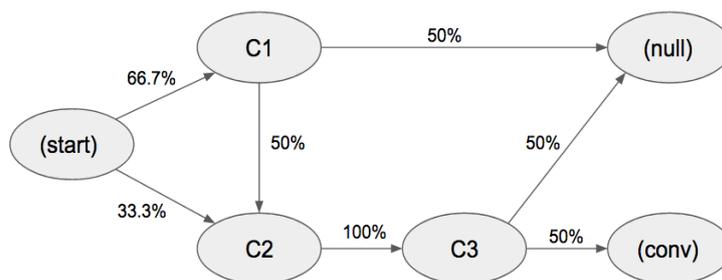
Tabela 2 - Cadeia de Markov com adição de estados

De	Para	Probabilidade	Probabilidade Total
(start)	C1	2/3	66.7%
(start)	C2	1/3	33.3%

De	Para	Probabilidade	Probabilidade Total
total from (start)		3/3	100%
C1	C2	1/2	50%
C1	(null)	1/2	50%
total from C1		2/2	100%
C2	C3	2/2	100%
total from C2		2/2	100%
C3	(conversion)	1/2	50%
C3	(null)	1/2	50%
total from C3		2/2	100%

Finalmente, tem-se o modelo na Figura 8 onde são mostradas as probabilidades de transição de cada estado.

Figura 8 - Diagrama de Estados 1



(fonte: <https://analyzecore.com/2016/08/03/attribution-model-r-part-1/>)

O último passo é estimar e entender a importância de cada canal / ponto de acesso.

É muito fácil fazer isso utilizando o princípio do **Efeito de Remoção**.

O **Efeito de Remoção** é a probabilidade de conversão quando um passo é completamente removido. Todas as sequências que tiveram que passar por esse passo agora são enviadas diretamente para o nó de saída. Este cálculo é feito executando um grande número de simulações no modelo de Markov com a etapa removida.

Ou seja, o núcleo do **Efeito de Remoção** é remover cada canal do gráfico consecutivamente e medir quantas conversões (ou quanto de valor) podem ser feitas (obtidas) sem a mesma.

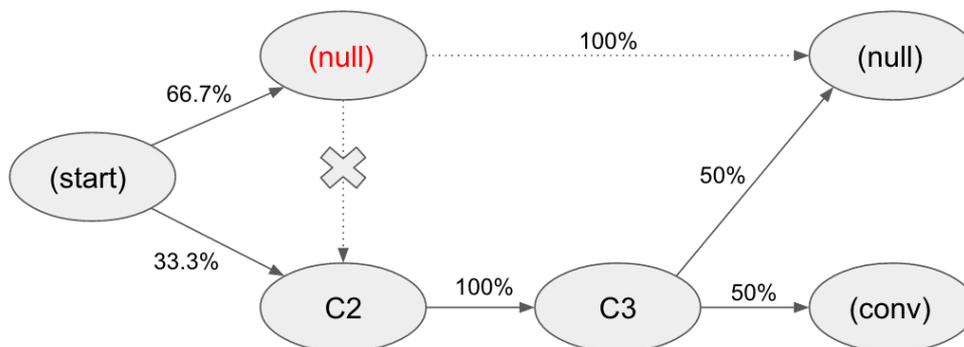
A lógica é a seguinte: se houver N conversões sem um determinado canal ou ponto de contato em relação ao total de conversões T do modelo completo, isso significa que o canal reflete a mudança no total de conversões (ou valor). Afinal, os canais ou pontos de contato são estimados: é preciso pesá-los porque a soma total de (T-Ni) seria maior do que T e normalmente é.

Outra maneira eficaz de medir o efeito de remoção está em porcentagens, por exemplo: o canal afetou probabilidades de conversão em X%.

É possível ilustrar como é feito isso em um exemplo simplificado. Remover um canal ou ponto de contato do gráfico significa que o mesmo deve ser substituído em pares de canais. Caso o canal exista em um estado "de", substitui-se o "de" por NA (e depois omite-se esse par) e é substituído o canal por (nulo) se estiver em um estado de "para". Em outras palavras, não haverá caminhos do canal e a transição dos outros canais para o estado (nulo) quando o canal estiver em uma parte "para".

O efeito de remoção para o **canal C1** é como demonstra a Figura 9 abaixo:

Figura 9 - Diagrama de Estados com efeito de remoção para o canal 1



Portanto, a probabilidade de conversão do modelo completo (com o canal C1) é 33,3% ($0,667 * 0,5 * 1 * 0,5 + 0,333 * 1 * 0,5$). Vide Figura 8.

A probabilidade de conversão após a remoção do canal C1 é 16,7% ($0,333 * 1 * 0,5$). Vide Figura 9.

Portanto, o **efeito de remoção do canal C1** é de 0,5 ($1 - (0,167 / 0,333)$), ou seja, 50%. Em outras palavras, se não houvesse o canal C1 nos caminhos dos clientes, 50% das conversões seriam perdidas, o que causaria um impacto enorme no negócio do anunciante.

O **Efeito de Remoção** de C2 e C3 é 1, porque todas as conversões seriam perdidas, uma vez que ambos os canais (C2 e C3) estão no único caminho que leva à conversão. ($1 - (0 / 0,333)$), ou seja, 100%. Além disso, é preciso pesar os índices e multiplicá-los pelo número total de conversões (1 nesse caso):

$$C1: 0,5 / (0,5 + 1 + 1) = 0,2 * 1 \text{ conversão} = 0,2$$

$$C2: 1 / (0,5 + 1 + 1) = 0,4 * 1 \text{ conversão} = 0,4$$

$$C3: 1 / (0,5 + 1 + 1) = 0,4 * 1 \text{ conversão} = 0,4.$$

4.4 USO DO GOOGLE ANALYTICS PARA DEFINIÇÃO DO MODELO DE MARKOV

Os dados permitem o melhor entendimento dos usuários e seus comportamentos, enquanto os métodos fornecem os meios para análise.

Esses métodos, variando em qualquer lugar de simplista (ou seja, frequência) para complexo (ou seja, clustering), permitem a escolha do que se queira entender a partir dos dados. Uma maneira popular de entender os usuários e seus comportamentos no Google Analytics é através da atribuição multicanal nos relatórios de funis multi-canal usando heurísticas simples: primeiro clique, último clique e atribuição linear.

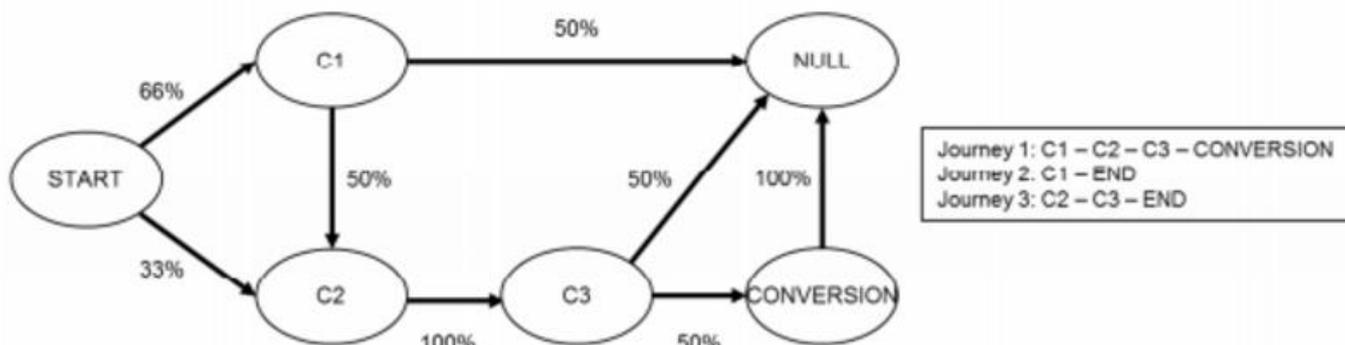
Embora esses métodos, respectivamente, forneçam informações sobre a frequência do primeiro ponto de contato de marketing, a frequência do último ponto de contato de marketing ou uma Sequência de pontos de contato de marketing igualmente importantes, um consumidor de dados pode querer um instantâneo diferente. Por exemplo, alguém que queira entender o nível de importância e / ou o valor de cada canal em relação às conversões deve usar um método diferente. A modelagem de Markov é um exemplo que pode ser usado.

Algumas razões para usar esse método:

- 1) Se o anunciante não tem o Google Analytics Premium (GA 360), então ele não tem acesso ao modelo de atribuição de dados.
- 2) Mesmo que o anunciante tenha o Google Analytics Premium, o algoritmo de atribuição de dados é um pouco de uma caixa preta.
- 3) É possível usar qualquer dado sequenciado – não há limitação a usar apenas canais de marketing.

Para determinar as sequências, o conteúdo dessas sequências é determinado pela ordem de Markov, que varia de 0 a 4.

Figura 10 - Diagrama de Estados 3



Ordem 0: Não se sabe de onde o usuário veio ou em que etapa o usuário está ativado, apenas a probabilidade de ir para qualquer página.

Ordem 1: de volta a zero passos. Estando no Passo A (Sequencia A). A probabilidade de ir a qualquer lugar é baseada em estar naquele passo.

Ordem 2: Refere-se para trás um passo. Vindo do Passo A (Sequência A) e está atualmente no Passo B (Sequência B). A probabilidade de ir a qualquer lugar é baseada em onde você estava e o usuário está.

Ordem 3: Refere-se dois passos para trás. Vindo do Passo A > B (Sequência A) e o usuário está atualmente no Passo C (Sequência B). A probabilidade de ir a qualquer lugar é baseada em onde ele estava e onde ele está.

Ordem 4: Refere-se para trás três etapas. Vindo do Passo A > B > C (Sequência A) e atualmente está no Passo D (Sequência B). A probabilidade de ir a qualquer lugar é baseada em onde ele estava e onde ele está.

Os modelos de Markov contam caminhos de usuários que se estendem após o número da ordem, atuando como uma janela deslizante.

Digamos que as etapas do Usuário X foram as seguintes: A > B > C > D > E > F > G. Este modelo mostraria o Usuário X passando da sequência A (A > B > C > D) para a sequência B (B > C > D > E) para a sequência C (C > D > E > F), e assim por diante até que o usuário X seja removido ou convertido.

Escolher a melhor ordem de Markov pode ser difícil. Sem entrar em muitos detalhes, uma maneira é traçar a precisão do treinamento do modelo versus o desvio padrão do treinamento. O objetivo é encontrar onde estas duas linhas se cruzam, ou onde o modelo gera variabilidade e perde a precisão igualmente.

Embora possa parecer complicado, ter uma compreensão básica da matemática por trás do modelo também pode ser útil.

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

Por sorte, isso pode ser simplificado em 3 partes principais:

- 1) A Probabilidade de Transição (w_{ij}) = A Probabilidade do Estado Anterior (sequência A, X_{t-1}) dado o Estado Atual (sequência B, X_t);
- 2) A Probabilidade de Transição (w_{ij}) não é inferior a 0 e não é superior a 1;
- 3) A soma das Probabilidades de Transição é igual a 1 (todos devem ir a algum lugar).

4.5 A ATRIBUIÇÃO DE CANAIS R PACKAGE

A **Atribuição de Canais R-package**, constrói os modelos de Markov que permitem calcular o número de conversões e/ou o valor de conversão que pode ser atribuído a cada canal de marketing. Em outras palavras, **Atribuição de Canais** usa modelos Markov para determinar a contribuição de cada canal para conversão e/ou valor (HARMON, KAELEN).

Este modelo concentra-se na resolução dos seguintes problemas:

- 1) **Objetividade** - Nenhum sentimento de o achismo. Apenas fatos.
- 2) **Precisão preditiva** - Previsões de eventos de conversão.
- 3) **Robustez** - Resultados válidos e confiáveis.
- 4) **Interpretabilidade** - Transparente e relativamente fácil de interpretar.
- 5) **Versatilidade** - Não depende do conjunto de dados. Capaz de se adaptar a novos dados.
- 6) **Eficiência Algorítmica** - Fornece resultados oportunos.

Também é importante ter em mente as seguintes limitações de atribuição:

- 1) **Endógeno** - A atribuição é relativa às condições subjacentes.
- 2) **Interpretação causal não estrita** - Os modelos de Markov não explicam 100% da variação entre as contribuições dos canais de marketing. Por exemplo, certos canais de comercialização podem ser inerentemente mais eficazes em uma determinada configuração.

5. ESTUDO DE CASO

5.1 INTRODUÇÃO

Para o estudo de caso, foram coletados dados de uma empresa real. O estudo analisa o melhor caminho do usuário até a conversão desejada. Nesse caso, a empresa estudada é uma construtora, no qual tem como conversão a conversa do usuário com um corretor online. Então cada vez que alguém entra no site e abre um chat online com um corretor, é contabilizada uma conversão.

Como objetivo de marketing, o anunciante busca aumentar os números mensais de chats online, afim de atingir mais vendas de apartamentos. Por se tratar de um grande anunciante, são utilizados diversos canais, tais como anúncios em Rede de Pesquisa, anúncios na Rede de Display, exibição na busca orgânica e entrada direta do usuário no site.

Com o modelo de atribuição adequado, é possível entender qual o melhor caminho do usuário, que trará maior probabilidade da conversão final, ou seja, o chat online. Vamos estudar, analisar e compreender se para esse caso, é melhor um usuário ser impactado com um anúncio da Rede de Display como primeira interação ou com um anúncio da Rede de Pesquisa, por exemplo.

Após o modelo ser criado, fica mais fácil decidir a melhor forma de alcançar possíveis clientes para a empresa, reduzindo os custos com marketing e aumentando o número de conversões obtidas.

5.2 COLETA DE DADOS

Os dados foram coletados pelo Google Analytics, com uma janela de período de um mês. Pode-se concluir quais as melhores estratégias a curto, médio e longo prazo para esse anunciante. O foco é em um estudo na análise a longo prazo, visto que essa sofre menos influências externas, como por exemplo, algum evento imobiliário que esteja acontecendo ou algum feriado, no qual os dados se comportam de forma anômala à realidade desse mercado. Foi coletado uma janela de tempo apenas para exemplificar a construção do modelo. Na prática, deve-se coletar várias janelas de tempo afim de diminuir a variância.

5.3 MODELO SEGUIDO

Antes de iniciar a análise dos dados obtidos, é ilustrado o modelo a ser seguido em um exemplo.

Pode-se considerar o caminho de conversão como uma série baseada em que cada estado é um canal ou ponto de contato do usuário. Como caminho do usuário, é possível considerar diversas sequências de canais, como:

Start > Orgânico > CPC > (conversão)

Start > CPC > Orgânico > Orgânico > nulo

Start > Orgânico > Orgânico > (conversão)

Os canais e pontos de contato utilizados nesse exemplo são:

- CPC – “cost per click”, ou seja, mídia paga
- Nulo ou saída - o usuário não volta mais
- Orgânico – resultado da busca orgânica
- Conversão – objetivo final

Para que o diagrama de estados e probabilidades seja feito, o primeiro passo é quebrar os caminhos em transições individuais, para simplificar a contagem de ocorrências em cada estado. Depois, com base no número de instâncias de uma transição versus possíveis transições, calcula-se as probabilidades individuais, como apresentado na Seção 4.4.

Tabela 2 – Canais

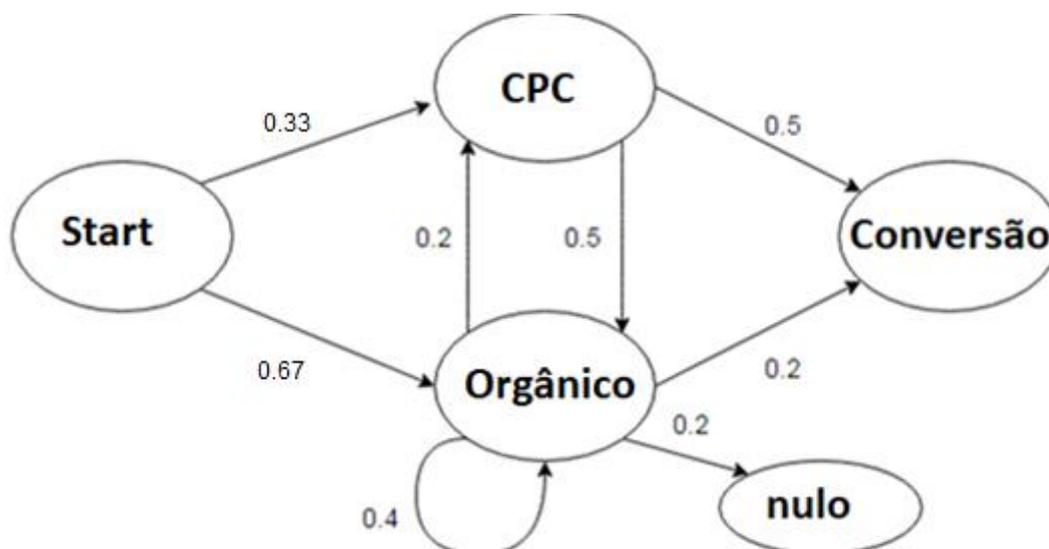
Transições:	
CPC > (conversão)	
CPC > Orgânico	
Orgânico > (conversão)	
Orgânico > CPC	
Orgânico > nulo	
Orgânico > Orgânico	
Orgânico > Orgânico	
Start > CPC	
Start > Orgânico	
Start > Orgânico	



Estado atual	Transição	Contagem	Probabilidade
CPC	Orgânico	1	1/2
	(conversão)	1	1/2
Orgânico	(conversão)	1	1/5
	CPC	1	1/5
	nulo	1	1/5
	Orgânico	2	2/5
Start	CPC	1	1/3
	Orgânico	2	2/3

A Figura 11 demonstra o diagrama.

Figura 11 - Diagrama de Estados



Para esse caso, é possível observar que, se a última visita foi no canal de busca orgânica, tem-se a probabilidade de quarenta por cento de que seja seguida por outra visita derivada da busca orgânica, vinte por cento que o usuário nunca mais será visto (nó nulo), vinte por cento que o próximo contato é via PPC (pay por click) e vinte por cento de probabilidade de conversão como próximo estado.

Agora, para que esse modelo tenha utilidade, é necessário fazer uma avaliação do impacto da remoção de um nó específico (ou um canal). Usando as probabilidades remanescentes, pode-se avaliar a queda nas conversões concluídas na ausência desse canal ou ponto de contato. Como citado anteriormente, essa prática tem base no **Efeito de Remoção**. Se o canal PPC for removido, as sequências ficam da seguinte forma:

Start > Orgânico > CPC > (conversão)

Start > CPC > Orgânico > Orgânico > nulo

Start > Orgânico > Orgânico > (conversão)

Quebrando as sequências de transições em individuais:

Tabela 3 - Canais

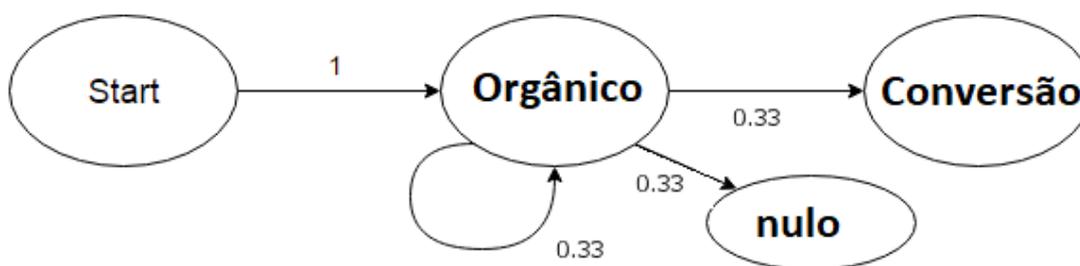
Transições:			
Start > Orgânico			
Start > Orgânico			
Orgânico > Orgânico			
Orgânico > (conversão)			
Orgânico > nulo			



Estado atual	Transição	Contagem	Probabilidade
Orgânico	Orgânico	1	1/3
	(conversão)	1	1/3
	nulo	1	1/3
Start	Orgânico	2	1

Com isso, o diagrama:

Figura 12 - Diagrama de Estados



Com o **Efeito de Remoção** de excluir o canal CPC, perde-se uma das duas conversões medidas anteriormente.

Se o mesmo for feito para o canal de busca orgânica, o resultado fica ligeiramente diferente. Pois, nenhuma das conversões ocorrem devido ao impacto do canal orgânico na conversão concluída pelo canal CPC. Deste modo, o **Efeito de Remoção** é a redução para zero.

Start > Orgânico > CPC > (conversão)

Start > CPC > Orgânico > Orgânico > nulo

Start > Orgânico > Orgânico > (conversão)

Total de conversões: 2

Tabela 4 - Total de conversões

Canal	Efeito de Remoção	%	Peso	Conversões atribuídas
CPC	1	50%	33%	0,67
Orgânico	2	100%	67%	1,33

Com essa análise, conclui-se que os resultados de busca orgânica têm um peso significativamente maior do que a busca paga (CPC). Análises como esta ajudam nas tomadas de decisões dos anunciantes. Em seguida, é abordado um caso real.

5.4 MODELOS DE ACOMPANHAMENTO

Na definição da página de destino dos anúncios, é possível inserir as informações de acompanhamento, formando assim, a URL do anúncio.

Os modelos de acompanhamento são usados para identificar a origem dos cliques dados nos anúncios. Por exemplo, pode-se adicionar o acompanhamento “display” ou “pesquisa” na url do anúncio, afim de saber se determinado clique foi originado de uma campanha paga de display ou pesquisa.

5.5 EXTRAINDO DADOS DO GOOGLE ANALYTICS

No Google Analytics, há uma ferramenta chamada Funis Multicanal. Os Relatórios de funis multicanais são gerados a partir de caminhos de conversão, as sequências de interações (ou seja, cliques/referências de canais) que levaram a cada conversão e transação. (GOOGLE, 2016).

Esses canais incluem, entre outros:

- Pesquisa paga e orgânica (em todos os mecanismos de pesquisa junto com as palavras-chave específicas pesquisadas);
- Sites de referência;
- Afiliados;
- Redes sociais;
- Boletins informativos por e-mail;
- Campanhas personalizadas que você criou, incluindo campanhas off-line que enviam tráfego para URLs curtos.

Para o desenvolvimento do modelo, foi usado o relatório de *caminhos de conversão mais comuns*. Esse relatório mostra os caminhos de conversão que os clientes percorreram até a conversão.

Foi definido um período de um mês. Pelo relatório do Google Analytics, foram obtidos os seguintes dados:

Figura 13 - Dados obtidos do Google Analytics

Caminho de origem/mídia ?	Conversões ?	↓
1. google-GDN / display × 2	4.958 (8,50%)	
2. google / cpc × 2	3.829 (6,57%)	
3. (direct) / (none) × 2	2.119 (3,63%)	
4. google / organic → (direct) / (none)	1.703 (2,92%)	
5. google-GDN / display × 3	1.676 (2,87%)	
6. (direct) / (none) × 3	1.283 (2,20%)	
7. google / cpc × 3	1.167 (2,00%)	
8. google / cpc → google / organic	1.052 (1,80%)	
9. (direct) / (none) × 4	924 (1,58%)	
10. google / cpc → (direct) / (none)	902 (1,55%)	

Antes da montagem das sequências e, posteriormente, as transições individuais dos estados, foi definido o seguinte modelo de acompanhamento utilizado.

- Google-GDN / display – Rede de display;
- Google / cpc – Rede de pesquisa;
- (Direct) / (none) – Direto;
- Google / Organic – Busca Orgânica.

Nesse relatório, é possível observar apenas usuários que converteram no final do caminho. Por isso, pode-se dizer que o primeiro estado de cada sequência é “Start” e o último estado sempre será “Conversão”.

Então, pelos dados coletados no período pré-definido, tem-se as sequências:

1. Start > Display > Display > Conversão
2. Start > Pesquisa > Pesquisa > Conversão
3. Start > Direto > Direto > Conversão
4. Start > Orgânico > Direto > Conversão
5. Start > Display > Display > Display > Conversão
6. Start > Direto > Direto > Direto > Conversão
7. Start > Pesquisa > Pesquisa > Pesquisa > Conversão
8. Start > Pesquisa > Orgânico > Conversão
9. Start > Direto > Direto > Direto > Direto > Conversão
10. Start > Pesquisa > Direto > Conversão

Quebrando em sequências individuais, a pares:

Tabela 5- Transições

Transições:
Direto > Conversão
Direto > Direto
Display > Conversão
Display > Conversão

Display > Display
Display > Display
Display > Display
Orgânico > Conversão
Orgânico > Direto
Pesquisa > Conversão
Pesquisa > Conversão
Pesquisa > Direto
Pesquisa > Orgânico
Pesquisa > Pesquisa
Pesquisa > Pesquisa
Pesquisa > Pesquisa
Start > Direto
Start > Direto
Start > Direto
Start > Display
Start > Display
Start > Orgânico
Start > Pesquisa

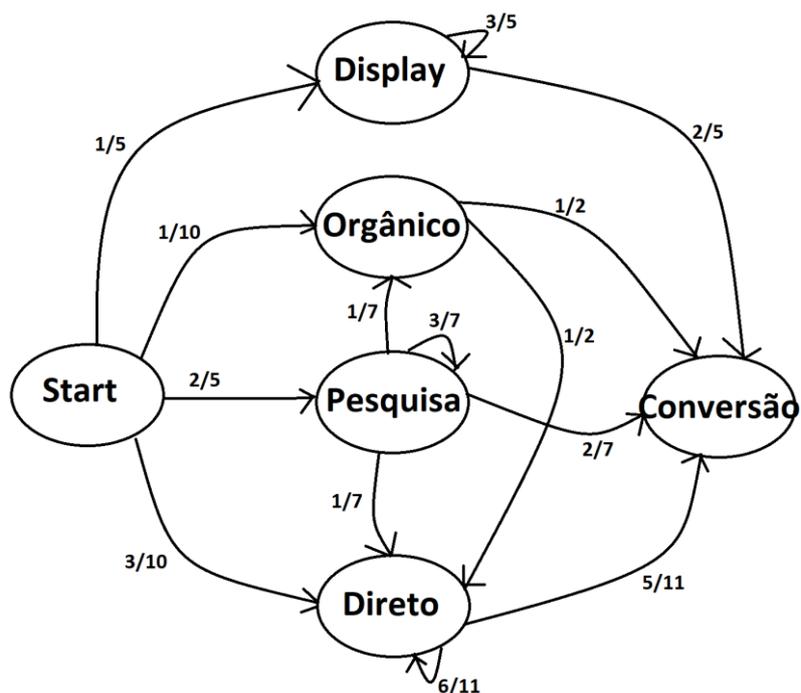
A tabela de transição, contagem e probabilidade fica da seguinte forma:

Tabela 6 - Tabela de transição, contagem e probabilidade

Estado atual	Transição	Contagem	Probabilidade
Direto	Conversão	5	5/11
	Direto	6	6/11
Display	Conversão	2	2/5
	Display	3	3/5
Orgânico	Conversão	1	1/2
	Direto	1	1/2
Pesquisa	Conversão	2	2/7
	Direto	1	1/7
	Orgânico	1	1/7
	Pesquisa	3	3/7
Start	Direto	3	3/10
	Display	2	1/5
	Orgânico	1	1/10
	Pesquisa	4	2/5

A partir dessa tabela, o diagrama e probabilidade de transição foi montado:

Figura 14- Diagrama de Estados Final



Agora, aplica-se o **Efeito de Remoção**. Mas antes de excluir qualquer canal, vale uma rápida análise do diagrama, afim de reduzir as possibilidades de remoção.

Percebe-se que o canal Pesquisa é bem importante, pois faz parte de diversas sequências no caminho de conversão. No entanto, após a interação com o canal de pesquisa, a maior probabilidade da próxima interação é também o canal pesquisa. Nesse caso, podemos utilizar do remarketing para exibir anúncios de display para quem já está no estado canal de pesquisa, uma vez que o canal de display tem a probabilidade maior de conversão. Então, para os usuários que estão no estado da rede de pesquisa, convém diminuir o remarketing em pesquisa e aumentar o remarketing em display, afim de levar esses usuários para o canal display, buscando mais conversões. Esse é um teste que deve ser feito e analisado posteriormente para ver se as conversões realmente vão aumentar.

O canal Orgânico também é muito importante, visto que tem 50% ($1/2$) de probabilidade de conversão ser o próximo estado a partir dele. Com essas análises, fica claro que o canal de pesquisa é essencial para os resultados obtidos.

Para melhor entendimento dos resultados eficientes para esse cliente, deve-se fazer as tabelas de transição e probabilidades, como foi feito para o exemplo da Seção 4.4. Assim, é possível entender o peso desse canal no caminho de conversão.

6. CONSIDERAÇÕES FINAIS

Este trabalho teve o objetivo de propor um modelo de atribuição. A contextualização em marketing digital, a análise de modelos existentes e o desenvolvimento de um modelo próprio de atribuição foram como previsto. Um modelo no qual apenas analisa-se o último clique, ou o primeiro é impróprio e ineficiente para a maioria dos casos. Provou-se que existe uma forma melhor de definir a importância de cada canal na jornada do consumidor. Com o efeito de remoção, é possível excluir caminhos dentro da jornada do consumidor, caso seja um caminho onde haja muito investimento e pouco retorno de conversões.

O desenvolvimento do modelo foi feito em um caso hipotético na Seção 4.4. Para o caso real, optou-se fazer apenas a demonstração de como os dados são obtidos pelo Google Analytics e como são definidos os modelos de acompanhamento. Desenvolver o modelo para o caso real seria redundante com o que já havia sido apresentado.

7. REFERÊNCIAS

REZ, R. Marketing de Conteúdo: A Moeda do Século XXI. São Paulo: DVS Editora, 2016.

BRYL', SERGEY. Marketing Multi-Channel Attribution model with R (part 1: Markov chains concept). Disponível em: <<http://analyzecore.com/2016/08/03/attribution-model-r-part-1>>. Acesso em: 5 de agosto de 2017.

FOTHERGILL, DAVID. Beyond heuristics: Algorithmic multi-channel attribution. Disponível em: <<https://martechtoday.com/beyond-heuristics-algorithmic-multi-channel-attribution-191679>>. Acesso em 4 de julho de 2017.

HARMON, KAELIN. Marketing Channel Attribution with Markov Models in R. Disponível em: <<https://www.lunametrics.com/blog/2016/06/30/marketing-channel-attribution-markov-models-r/>>. Acesso em agosto de 2017.

GOOGLE, Formatos de publicidade no YouTube. Disponível em: <<https://support.google.com/youtube/answer/2467968?hl=pt-BR>>. Acessado em 2017.

REZ, KPI's: O que são Key Performance Indicators (Indicadores Chaves de Performance. Disponível em: <<http://www.marketingdeconteudo.com.br/infograficos/kpis-o-que-sao-key-performance-indicators-indicadores-chaves-de-performance>>. Acessado em 2017.

GOOGLE, Visão geral do modelo de atribuição. Disponível em: <<https://support.google.com/analytics/answer/1662518?hl=pt-BR>>. Acessado em 2017.

BRYL', Marketing Multi-Channel Attribution model with R (part 1: Markov chains concept). Disponível em: <<http://analyzecore.com/2016/08/03/attribution-model-r-part-1>>. Acessado em 2017.

FOTHERGILL, 2016. Disponível em: <<https://martechtoday.com/beyond-heuristics-algorithmic-multi-channel-attribution-191679>>. Acessado em 2017.

GOOGLE, Sobre funis multicanal. Disponível em: <<https://support.google.com/analytics/answer/1191180?hl=pt-BR>> Acessado em 2017.