

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ  
ESCOLA DE NEGÓCIOS  
CURSO DE MARKETING**

**ALEX ALEXANDRE DIAS  
ANTONIO AUGUSTO LIMA CASTAGNOLI  
LUCAS ANDRE DIETER**

**ANÁLISE DA EFICIÊNCIA DA PUBLICIDADE ONLINE NO CONTEXTO DE  
E-COMMERCE A PARTIR DA ORIGEM DE UTM'S: UMA ANÁLISE DE BIG DATA**

**CURITIBA  
2017**

**ALEX ALEXANDRE DIAS  
ANTONIO AUGUSTO LIMA CASTAGNOLI  
LUCAS ANDRE DIETER**

**ANÁLISE DA EFICIÊNCIA DA PUBLICIDADE ONLINE NO CONTEXTO DE  
E-COMMERCE A PARTIR DA ORIGEM DE UTM'S: UMA ANÁLISE DE BIG DATA**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Marketing da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de bacharel em Marketing.

Orientador: Prof. Dra. Eliane Cristine Francisco Maffezzoli

**CURITIBA  
2017**

**PÁGINA RESERVADA PARA FICHA CATALOGRÁFICA QUE DEVE SER  
CONFECCIONADA APÓS APRESENTAÇÃO E ALTERAÇÕES SUGERIDAS  
PELA BANCA EXAMINADORA.**

**DEVE SER IMPRESSA NO VERSO DA FOLHA DE ROSTO**



**ALEX ALEXANDRE DIAS  
ANTONIO AUGUSTO LIMA CASTAGNOLI  
LUCAS ANDRE DIETER**

**ANÁLISE DA EFICIÊNCIA DA PUBLICIDADE ONLINE NO CONTEXTO DE  
E-COMMERCE A PARTIR DA ORIGEM DE UTM'S: UMA ANÁLISE DE BIG DATA**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Marketing da Pontifícia Universidade Católica do Paraná, como requisito parcial à obtenção do título de bacharel em Marketing.

**COMISSÃO EXAMINADORA**

---

Professor 1 (Titulação e nome completo)  
Instituição 1

---

Professor 2 (Titulação e nome completo)  
Instituição 2

---

Professor 3 (Titulação e nome completo)  
Instituição 3

Curitiba, \_\_\_\_ de \_\_\_\_\_ de 2017.



## **AGRADECIMENTOS**

Temos o profundo agradecimento pela Prof. Dra. Eliane Cristine Francisco Maffezzolli pela sua fenomenal orientação e à Bruna Wundervald, nossa Especialista em Estatística, que nos ajudou muito no manejo da base de dados e ao Prof. Dr. Juan Jose Viacava, que teve grande importância para a realização com sucesso deste trabalho.

Agradecemos também à Prof. Dra. Miriam Stolses Mazo, por todo o suporte durante estes 4 anos de curso na PUCPR.



## RESUMO

A utilização de *Big Data Analytics (BDA)* é uma prática ainda pouco explorada pelas organizações – principalmente as de menor porte – para melhorar suas estratégias de Marketing. Muitas destas ainda permanecem utilizando ferramentas simples que fornecem informações rasas perto das informações que suas bases de dados têm o potencial de oferecer. Este estudo, então, procurou ir além destas informações já conhecidas e encontrar, a partir de uma análise de *Big Data* utilizando a origem de UTM's, dados adicionais com potencial de se transformarem em informações efetivamente úteis para as tomadas de decisões gerenciais. Após um processo de limpeza, foi possível extrair informações sobre compras e clientes que tiveram origem em campanhas, para, assim, gerar uma comparação: neste caso, se os compradores que tiveram origem em campanhas (publicitárias e e-mail marketing) demonstraram maior propensão de compra do que os consumidores que chegaram de forma orgânica. No caso da empresa analisada, não houve diferença significativa. Contudo, o conhecimento de processos adquiridos permite que uma análise como esta seja feita caso a caso, a ponto dos gestores de comércio eletrônico (*e-commerces*) poderem ter um maior controle sobre seus orçamentos de Marketing e entenderem melhor o comportamento de seus clientes, tanto os oriundos de campanha, quanto os que chegam de forma orgânica.

**Palavras-chave:** big data, big data analytics (BDA), comércio eletrônico, *e-commerce*, origem de UTM's, marketing digital.

## **ABSTRACT**

The big data analytics (BDA) is an under explored practice in the organizations - mainly in the smaller ones - to improve their Marketing strategies. A lot of these organizations are still using simple tools that provide plain information compared to those databases potentially can provide. This research aimed to go forward those known informations and to find, from UTM's origin based big data analytics, adicional data effectively useful for management decision making. After a data cleaning process, it was possible to extract informations about client's purchases and if their origin were from campaigns (advertising or e-mail marketing), to then, generate a comparison: in this case, if buyers that had campaign origin tend to buy more than organic-originated buyers. In the case of the analysed company, there were no significative difference. However, the acquired process knowledge let that analysis like that to be made case by case, at the point that managers of e-commerces are able to have more control over their Marketing budget and also let them to better understand their consumer behaviors, both those coming from campaigns and those who come organically.

**Keywords:** big data, big data analytics (BDA), e-commerce, UTM's origin, digital marketing.

## LISTA DE FIGURAS

Figura 1 - Referencial Teórico do Estudo .....	19
Figura 2 - Modelo Integrado de Comércio Eletrônico .....	21
Figura 3 - Funcionamento de um sistema de UTM .....	31
Figura 4 - Modelo do 3 V's: volume, variedade e velocidade .....	33
Figura 5 - Modelo gráfico de hipótese .....	40
Figura 6 - Estrutura básica da base de dados de impressão de publicidade e custos. .....	43
Figura 7 - Estrutura básica da base de dados de sessões.....	44
Figura 8 - Estrutura da base de dados por usuário. ....	49
Figura 9 - Modelo gráfico de hipótese. ....	50
Figura 10 - Exemplo do tipo de anúncio.....	53
Figura 11 - Modelo gráfico da hipótese 1. ....	67

## LISTA DE GRÁFICOS

Gráfico 1 - Faturamento do comércio eletrônico ( <i>e-commerce</i> ) no Brasil .....	17
Gráfico 2 - Cruzamento dos dados da Tabela 5.....	55
Gráfico 3 - Cruzamento do orçamento diário efetivado com os dias. ....	57
Gráfico 4 - Cruzamento dos dias da semana com o número de visitas e números de visitas únicas.....	59
Gráfico 5 - Cruzamento das 24 faixas de 1h do dia com o número de visitas e número de visitas únicas.....	61
Gráfico 6 - Relação de compras pelo horário em que elas ocorreram. ....	64
Gráfico 7 - Relação entre a quantidade de compras efetuadas pelos clientes e os dias do mês analisados.....	66

## LISTA DE TABELAS

Tabela 1 - Definições dos formato de publicidade e modelos de precificação .....	27
Tabela 2 - Parâmetro de UTM.....	32
Tabela 3 - Sumário de Oportunidades com <i>Big Data</i> .....	34
Tabela 4 - Equações para métricas online .....	39
Tabela 5 - Cruzamento entre os dias do mês, dias da semana, nº de visitas, nº de visitas únicas e impressões de publicidade.....	54
Tabela 6 - Cruzamento entre os dias do mês e os dias da semana, com o orçamento gasto no dia, número de impressões e custo por mil (CPM).....	56
Tabela 7 - Cruzamento entre os dias de semana, número de visitas e número de visitas únicas.....	58
Tabela 8 - Cruzamento entre as 24h do dia, com o número de visitas e número de visitas únicas.....	59
Tabela 9 - Número de compras por faixa de horário.....	62
Tabela 10 - Cruzamento entre dias do mês, número de compras, orçamento investido no dia e custo por pedido.....	64
Tabela 11 - Tabulação Cruzada entre as variáveis "buy" e "cp". .....	67
Tabela 12 - Testes de Omnibus de Modelos de Coeficientes.....	68
Tabela 13 - Teste de Hosmer e Lameshow.....	68

## LISTA DE QUADROS

Quadro 1 - Diferenças entre tipos de links .....	31
--	----

## LISTA DE ABREVIATURAS E SIGLAS

B2B	<i>Business to Business</i>
B2C	<i>Business to Consumer</i>
B2G	<i>Business to Government</i>
BDA	<i>Big Data Analytics</i>
C2B	<i>Consumer to Business</i>
C2C	<i>Consumer to Consumer</i>
C2G	<i>Consumer to Government</i>
CSV	<i>Comma Separated Value</i>
G2G	<i>Governemnt to Government</i>
GA	<i>Google Analytics</i>
IAB	<i>Interactive Advertising Bereau</i>
JSON	<i>Java Script Object Notation</i>
PB	<i>PetaBytes</i>
SQL	<i>Structured Query Language</i>
TB	<i>TeraBytes</i>
TI	Tecnologia da Informação
UTM	<i>Urchin Tracking Module</i>
RTB	<i>Real Time Bidding</i>
ZB	<i>ZetaBytes</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>16</b>
1.1 PROBLEMATIZAÇÃO.....	18
1.2 OBJETIVOS.....	18
<b>1.2.1 Objetivo Geral</b> .....	<b>18</b>
<b>1.2.2 Objetivos Específicos</b> .....	<b>18</b>
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>19</b>
2.1 COMÉRCIO ELETRÔNICO ( <i>E-COMMERCE</i> ).....	20
<b>2.1.1 Tipos de Comércio Eletrônico (<i>e-commerce</i>)</b> .....	<b>21</b>
<b>2.1.2 Cenário do Comércio Eletrônico (<i>e-commerce</i>)</b> .....	<b>22</b>
2.2 MARKETING DIGITAL.....	23
<b>2.2.1 Ferramentas de Marketing Digital</b> .....	<b>25</b>
2.3 PUBLICIDADE ONLINE.....	25
<b>2.3.1 Tipos de Conteúdo</b> .....	<b>26</b>
<b>2.3.2 Formatos de Publicidade Online</b> .....	<b>27</b>
<b>2.3.3 Mídia Programática</b> .....	<b>29</b>
2.4 UTM'S ( <i>URCHIN TRACKING MODULES</i> ).....	30
2.5 BIG DATA NO CONTEXTO DE MARKETING.....	32
<b>2.5.1 <i>Big Data Analytics (BDA) – Análise de Big Data</i></b> .....	<b>36</b>
2.6 EFICIÊNCIA DE PUBLICIDADE ONLINE.....	36
<b>2.6.1 <i>Indicadores de Desempenho de Marketing</i></b> .....	<b>37</b>
<b>2.6.2 <i>Métricas para Eficiência Online em E-Commerce</i></b> .....	<b>37</b>
<b>3 METODOLOGIA</b> .....	<b>40</b>
3.1 BASE DE DADOS SOBRE NÚMEROS, DIAS E HORÁRIOS DE VISITAS.....	41
3.2 BASE DE DADOS DE CUSTOS DE PUBLICIDADE E QUANTIDADE DE IMPRESSÕES.....	42
3.3 BASE DE DADOS DAS SESSÕES DE NAVEGAÇÃO.....	43
<b>3.3.1 <i>Identificação de Sessões Provenientes de Campanhas (Variável Independente)</i></b> .....	<b>45</b>
<b>3.3.2 <i>Identificação de Sessões que Finalizaram em Compra (Variável Dependente)</i></b> .....	<b>47</b>
3.4 BASE DE DADOS INDIVIDUAIS .....	48
<b>3.4.1 <i>Contagem de Casos</i></b> .....	<b>50</b>

<b>3.4.2 Contagem de Compras por horário do dia .....</b>	<b>51</b>
<b>3.4.3 Contagem de Compras por dia do mês .....</b>	<b>52</b>
<b>3.4.4 Análise do tipo de anúncio utilizado.....</b>	<b>52</b>
<b>4 ANÁLISE DOS RESULTADOS .....</b>	<b>54</b>
4.1 ANÁLISE DESCRITIVA .....	54
4.2 ANÁLISE DA RELAÇÃO ENTRE EXPOSIÇÃO POR CAMPANHA E EFETIVAÇÃO DA COMPRA.....	66
<b>5 CONSIDERAÇÕES FINAIS.....</b>	<b>69</b>
5.1 APLICAÇÕES TEÓRICAS.....	69
5.2 APLICAÇÕES GERENCIAIS.....	69
5.3 LIMITAÇÕES E SUGESTÕES DE ESTUDOS FUTUROS .....	71
<b>REFERÊNCIAS.....</b>	<b>73</b>
<b>APÊNDICE A – ANÚNCIOS MODIFICADOS UTILIZADOS PELA EMPRESA X NO PERÍODO DE ANÁLISE .....</b>	<b>80</b>

## 1 INTRODUÇÃO

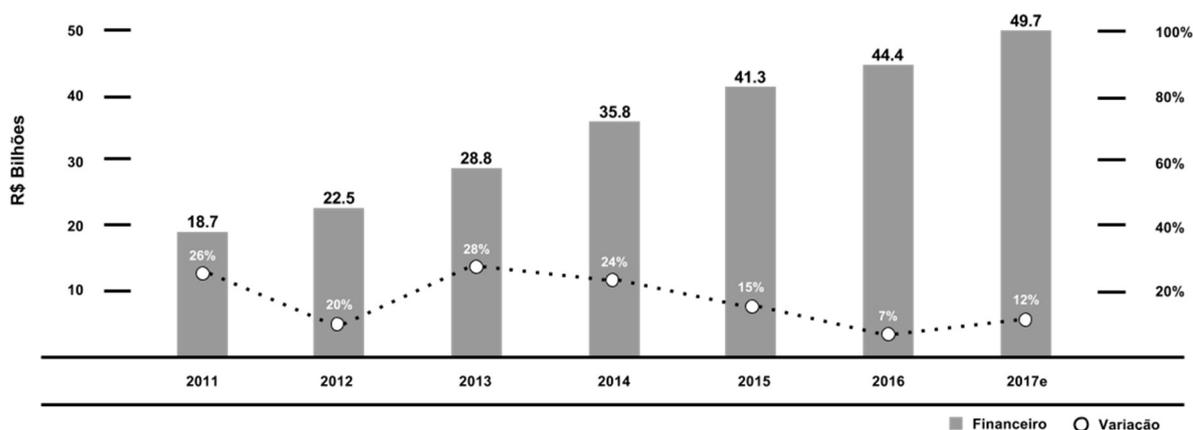
O constante avanço de tecnologias digitais dentro do contexto de marketing nos últimos anos possibilitou inúmeras oportunidades para a área, desde interpretar decisões de marketing até executar campanhas mais eficientes. Segundo Prange (2017), o mundo do marketing e da publicidade está vivendo seu melhor momento em relação à inteligência de negócios e isso se deve à disponibilidade de informações estratégicas – antes indisponíveis – processadas por ferramentas de *Big Data Analytics* (BDA), o que permitiu uma grande otimização nas rotinas de diversos setores. As ferramentas utilizam dados não estruturados coletados de diversas fontes, tais como: sites, mídias sociais, processos de atendimento ao cliente, relacionamentos e vendas. As análises oriundas destes dados permitem gerar informações que possibilitam vantagens estratégicas para muitos setores, inclusive, o de comércio eletrônico (*e-commerce*), uma vez que permite um melhor mapeamento dos comportamentos de consumidores na plataforma, e assim pode possibilitar a criação de estratégias de publicidade online mais assertivas.

Segundo a *Interactive Advertising Bureau* (IAB Brasil, 2017) – a maior rede de associações do mundo em mídia interativa – no ano de 2016 foram investidos R\$ 11,8 bilhões em publicidade online, sendo 16,5% destes investimentos em mídia programática (*displays, banners, etc*). A estimativa de investimento para o período de 2017 tende a R\$ 14,8 bilhões, vislumbrando um crescimento de 26% em relação ao ano anterior, o que representaria aproximadamente um terço do total investido em publicidade no Brasil. De acordo com Turchi (2012), a *web* tem se tornado cada vez mais protagonista como um canal de mídia publicitária e por isso passou a ser notada pelas organizações como uma forma de dar mais velocidade aos negócios e como um campo atrativo para ser estudado.

Um ponto a ser considerado com a propagação da internet e conseqüentemente das mídias sociais, conforme Verhoef, Kooge e Walk (2016), foi a grande quantidade de dados proporcionados, que se tornaram disponíveis para muitas empresas, diariamente ou em tempo real, dando início ao conceito de *Big Data*. Para Bello-Orgaz, Jung e Camacho (2015), os dados vêm de toda a parte: desde sensores usados para condições climáticas, informações de voos e trânsito, *posts* em mídias sociais, e até vídeos ou fotos digitais, essenciais para o comércio eletrônico (*e-commerce*).

Segundo o relatório divulgado pela E-Bit (*Webshoppers*, 2017), estudo demonstrativo da situação do comércio eletrônico (*e-commerce*) brasileiro em 2016, mostra que ao menos 48 milhões de brasileiros realizaram uma compra virtual no período. O relatório também detalha o crescimento do segmento e a estimativa de faturamento para 2017, totalizando um crescimento nominal de 12% ao atingir um total de R\$ 49,7 bilhões, mostrado no gráfico 1.

Gráfico 1 - Faturamento do comércio eletrônico (*e-commerce*) no Brasil



Fonte: E-Bit Informação (*Webshoppers*, 2017 p. 23).

De acordo com Kotler (2010), a disseminação do digital gerou impacto no comportamento de consumidores e marcas, e portanto, no marketing como um todo. As facilidades no acesso à informação possibilitaram a comparação entre produtos e serviços, removendo os hiatos da relação marca-consumidor, tanto geográficos quanto temporais. Desta forma, segundo Torres (2009, p. 25), “não há mais separação entre produtor e consumidor. Não há mais exclusividade de produção nem na mídia nem no software [...]”.

Dado isto, segundo Limeira (2001), as organizações necessitam compreender quais as características próprias da internet, para assim, poder aperfeiçoar seus efeitos. No contexto do marketing, de acordo com Hofacker, Malthouse e Sultan (2016), os consumidores consomem cada vez mais digitalmente, e o consumo de mídia também já é quase totalmente digital.

Pesquisadores de marketing acreditam que o *Big Data Analytics* (BDA) e a computação em nuvem oferecem uma oportunidade única para que empresas obtenham opiniões de um grande número de clientes, melhorando as estratégias mais tradicionais (Bello-Orgaz, Jung e Camacho, 2015).

## 1.1 PROBLEMATIZAÇÃO

Diante do cenário apresentado, formulou-se o seguinte problema de pesquisa: “como avaliar a eficiência da publicidade online através de uma análise de *Big Data* rastreando a origem de acessos em sites de comércio eletrônico (*e-commerce*) por meio de UTM's?”

Segundo Turchi (2015), anteriormente a publicidade era realizada através de negociações diretas com os veículos de mídia. Apesar de ter existido por bastante tempo, este antigo modelo de publicidade foi evidenciando suas deficiências, como desperdício de mão de obra, muitos processos manuais onerando o rendimento e a eficiência dos resultados que nem sempre atingia o público-alvo. Assim, para corrigir esses problemas, a principal tendência da análise de publicidade é o *Big Data*, que disponibiliza uma quantidade de dados e métricas onde é possível apurar informações precisas sobre o público-alvo de um produto ou marca, que para a publicidade se torna uma possibilidade de obter melhores resultados.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Este estudo tem como objetivo geral como avaliar a eficiência da publicidade online através de uma análise de *Big Data* rastreando a origem de acessos em sites de comércio eletrônico (*e-commerce*) por meio de UTM's.

### 1.2.2 Objetivos Específicos

Desta forma, os objetivos específicos do trabalho são:

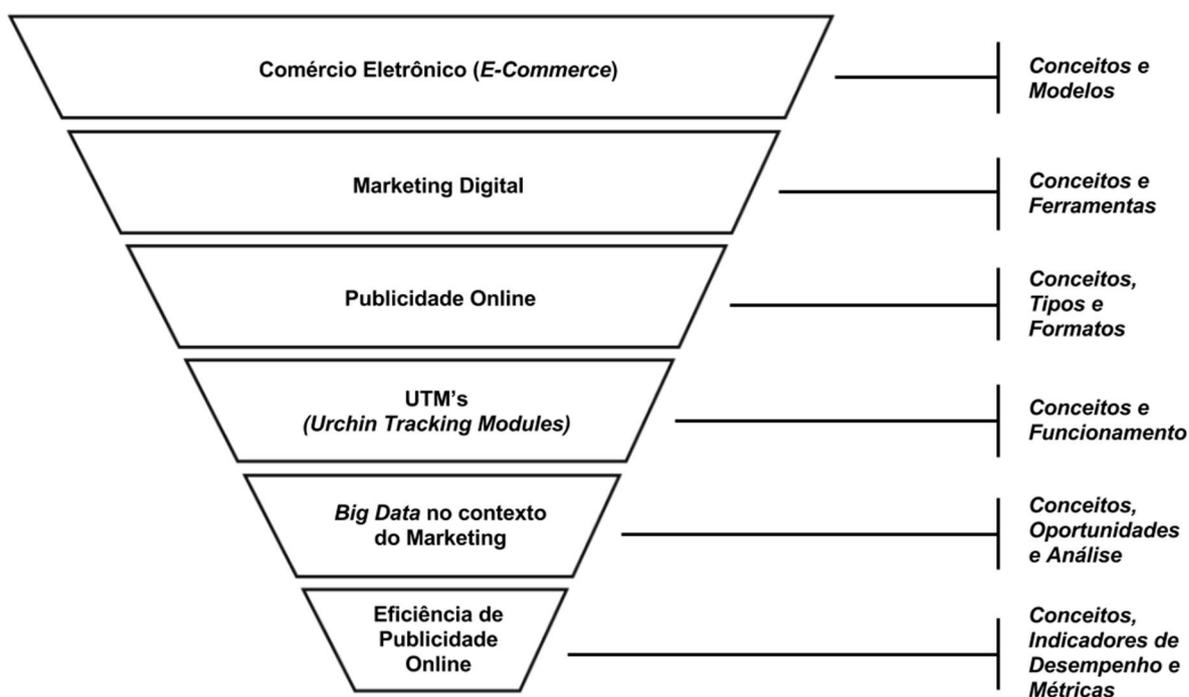
- a) Analisar a quantidade de consumidores provenientes de campanhas de publicidade online e de forma orgânica;
- b) Identificar os principais indicadores de desempenho de eficiência de publicidade online;

- c) Verificar o custo por pedido de clientes provindos de publicidade *online* que efetivaram uma compra;
- d) Compreender se consumidores oriundos de campanhas de publicidade *online* são mais propensos a comprar.

## 2 REFERENCIAL TEÓRICO

A metamorfose estimulada pelo digital, observada na sociedade, no mercado e no consumidor demonstra mudanças em seus comportamentos (Gabriel, 2010). Assim sendo, o universo do marketing necessita também evoluir, considerando novas visões, ferramentas e plataformas de ação disponíveis. No presente estudo serão abordados os conceitos de comércio eletrônico (*e-commerce*), marketing digital, publicidade online, *UTM's*, *big data*, eficiência e indicadores de desempenho. Com a finalidade de contextualizar as temáticas do trabalho em questão, os conteúdos abordados são organizados de acordo com a figura 1:

Figura 1 - Referencial Teórico do Estudo



Fonte: Elaborado pelos Autores, 2017.

## 2.1 COMÉRCIO ELETRÔNICO (*E-COMMERCE*)

Para Kotler e Keller (2006, p. 438-439), *e-commerce* é a prática que utiliza *websites* para realizar transações e também facilitar a venda de produtos e serviços *online*. Ainda, para os autores, “os varejistas *online* competem em três aspectos principais de uma transação: (1) interação do cliente com o site, (2) entrega e (3) capacidade de resolver problemas quando ocorrem”.

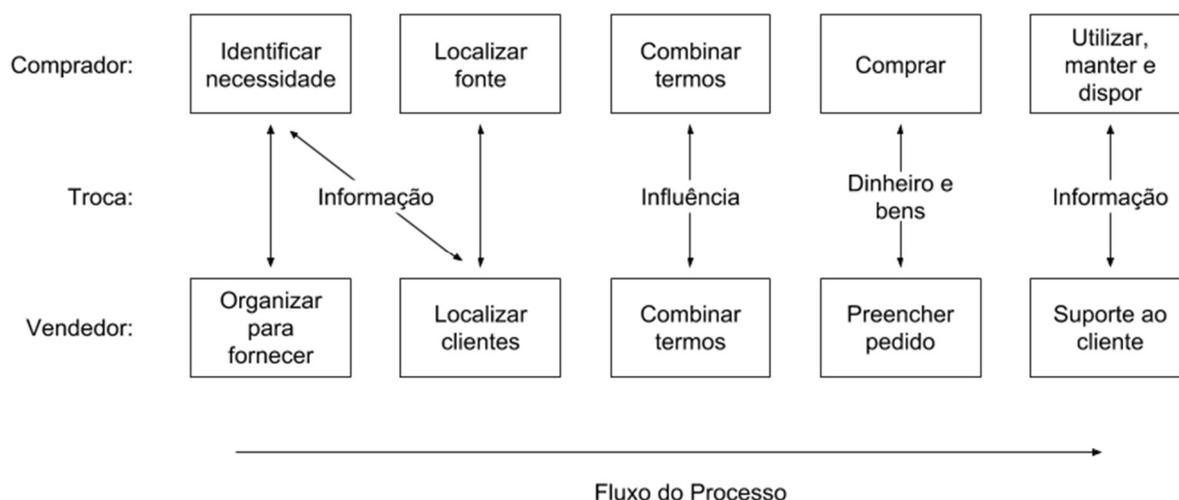
Desde o final dos anos 90, segundo Turchi (2012), a internet fez com que grandes corporações planejassem e integrassem suas estratégias de marketing à tecnologia, elaborando um modelo de maior performance, começando a realizar transações comerciais via *web*. As bases do *e-commerce* foram criadas e conseqüentemente, algum tempo depois, suas variações como o comércio por meio de dispositivos móveis (*m-commerce*), comércio por meio da televisão (*t-commerce*), comércio via redes sociais (*social commerce*) e mais especificamente o comércio via *Facebook* (*f-commerce*).

De acordo com Rosenbloom (2015), o comércio eletrônico (*e-commerce*) faz parte do termo “canais eletrônicos de marketing”, que na prática engloba outros termos semelhantes como compra pela internet, no ciberespaço, na *web*, compra virtual, distribuição eletrônica ou apenas compra *online*. Conforme Albertin (2010, p. 3), “o comércio eletrônico é a realização de toda a cadeia de valor dos processos de negócio num ambiente eletrônico, por meio da aplicação intensa das tecnologias de comunicação e de informação, atendendo aos objetivos de negócio”.

Cameron (1997) cita que o comércio eletrônico (*e-commerce*) é definido por qualquer negócio que transacione de maneira eletrônica, e tais transações ocorrem entre dois parceiros de negócio ou entre um negócio e seus respectivos clientes.

Segundo Bloch, Pigneur e Segev (1996, p. 2), o comércio eletrônico é definido como “a compra e venda de informações, produtos, e serviços via redes de computadores”. Os autores também colocam o *e-commerce* como “o suporte para qualquer tipo de transação de negócio através de uma estrutura digital”, como uma forma de fornecer informação para seus consumidores, tanto como ferramenta de marketing, canal de vendas e até como uma linha de suporte ao consumidor. Sob a perspectiva comprador-vendedor, o *e-commerce* pode ser entendido da mesma forma que transação comercial comum, como mostra a Figura 02 abaixo:

Figura 2 - Modelo Integrado de Comércio Eletrônico



Fonte: adaptado de Bloch, Pigneur e Segev, 1996.

### 2.1.1 Tipos de Comércio Eletrônico (*e-commerce*)

Para Andrade (2014, p. 97), “o *e-commerce* pode assumir diferentes formas, dependendo de quão transformado o processo de negociação foi do físico ao digital”. De acordo com Potter e Turban (2005 apud MENDES, S/D), existem diversos tipos de Comércio Eletrônico (*e-commerce*), sendo os mais gerais:

- B2B – *Business-to-Business*: se trata da negociação eletrônica entre organizações. É uma modalidade comum, e a que mais movimentava valores monetários.
- B2C – *Business-to-Consumers*: entre empresas e consumidores, representando a compra e venda *online*. Aqui, a principal diferença é que as pessoas escolhem e pagam os produtos pela internet.
- C2B – *Consumers-to-Business*: negociação entre consumidores e empresas (o contrário do modelo B2C). Acontece quando consumidores vendem para empresas.
- C2C – *Consumer-to-Consumer*: esta modalidade acontece entre consumidores e é bastante comum, com muitas negociações de valores pequenos.

No entanto, Limeira (2003) afirma que existem, além dos modelos mais comuns como B2B e B2C, transações entre consumidores e governo (*Consumer-to-*

*Government – C2G*), trocas entre governo e empresas, (*Business-to-Government – B2G*) e transações entre órgãos de governo (*Government-to-Government – G2G*).

### **2.1.2 Cenário do Comércio Eletrônico (*e-commerce*)**

Em pesquisa realizada pela E-Bit (*Webshoppers*, 2017), com relatório demonstrativo sobre a situação do comércio eletrônico (*e-commerce*) no Brasil no ano de 2016, o volume de transações relativas ao período foi igual ao ano anterior, registrando um total de 106 milhões de pedidos. O relatório cita que o número só não foi menor devido às vendas acontecidas no segundo semestre do ano, principalmente em decorrência da Black Friday e outras datas sazonais. Conforme apresentado na introdução deste estudo (Seção 1), a pesquisa mostra que 48 milhões de brasileiros realizaram pelo menos uma compra virtual no período. Além disso, também é detalhado o crescimento do segmento e a estimativa de faturamento para 2017, com um crescimento nominal de 12% ao atingir um total de R\$ 49,7 bi, como mostrado no gráfico 01 deste estudo.

Para se ter um panorama geral sobre o cenário, um estudo realizado pelo Sebrae Nacional (2016) aponta as categorias mais vendidas pelas empresas brasileiras no *e-commerce* em 2015, conforme pode ser observado na lista abaixo:

- Moda e acessórios (14%)
- Eletrodomésticos (13%)
- Telefonia / celulares (11%)
- Cosméticos e perfumaria / cuidados pessoais (10%)
- Livros / assinaturas e revistas (9%)
- Casa e decoração (9%)
- Informática (7%)
- Eletrônicos (6%)
- Esporte e lazer (4%)
- Brinquedos e games (4%)

De acordo com Torres (2012), no contexto de *e-commerce* também se faz necessário um investimento em TI, ainda que isto não se torne uma vantagem

competitiva sustentável, já que as aplicações vão se tornando mais praticáveis. No entanto, para as empresas que atuam *online*, a fim de se manterem competitivas devem manter investimento contínuo em TI, acompanhando tudo o diz respeito a tecnologias de informação.

Na maioria dos casos, empresas de *e-commerce* lidam com dados estruturados e não-estruturados, segundo Akter e Wamba (2016), e o desafio é lidar com ambos os tipos de dados, a fim de gerar insights significativos para aumentar as conversões. Koirala (2012) cita que devido à necessidade de se manter no topo, empresas de *e-commerce* estão entre os principais grupos a aderirem ao uso de *Big Data Analytics* (BDA).

## 2.2 MARKETING DIGITAL

Para auxiliar o entendimento do conceito de marketing digital, se faz necessária a leitura e entendimento de autores da área.

Segundo Kotler (2010), a disseminação do digital gerou impacto no comportamento de consumidores e marcas, e, portanto, no marketing como um todo. As facilidades no acesso à informação possibilitaram a comparação entre produtos e serviços, removendo os hiatos da relação marca-consumidor, tanto geográficos quanto temporais.

De acordo com Gabriel (2010), tecnologias digitais estão cada vez mais presentes em todos os aspectos da vida humana: social, profissional e pessoal. Uma das consequências mais importantes das tecnologias digitais é seu auxílio no processo da chamada inversão do vetor de marketing, que traz uma grande mudança na relação marca-consumidor, fazendo com que hoje os consumidores busquem a marca onde, quando, e da forma que desejarem, indo na direção contrária do conhecido conceito tradicional do marketing em que “as ações de promoção e relacionamento acontecem no sentido da empresa para o consumidor”.

Segundo Kotler, Kartajaya e Setiawan (2016), a segmentação do marketing tradicional é substituída na economia digital, onde os consumidores estão conectados uns com os outros por meio de comunidades formadas naturalmente dentro dos limites definidos por eles próprios. As comunidades são os novos segmentos, e estão imunes ao *spamming* e publicidade irrelevante. Conforme os autores, as ferramentas

do mix de marketing tradicional – os 4 P's (produto, preço, praça e promoção) – evoluíram para atender uma maior participação dos consumidores, e deve ser redefinido através dos 4 C's – cocriação (*co-creation*), moeda (*currency*), ativação comunitária (*communal activation*) e conversa (*conversation*) propostos. A cocriação (*co-creation*) se trata da nova estratégia de desenvolvimento de produto, permitindo consumidores a customizar e personalizar seus bens e serviços em conjunto com a marca, enquanto a moeda (*currency*), substitui a precificação tradicional e é uma definição para o que é chamado de precificação dinâmica - que é o preço que flutua dependendo da demanda de mercado que cada produto possui, e assim as companhias podem otimizar sua rentabilidade através da cobrança diferenciada para cada indivíduo. Os comércios eletrônicos, por exemplo, coletam grandes quantidades de dados que os permitem realizar uma análise de *Big Data* (*BDA*) e a partir disso oferecem uma precificação única para cada cliente, com base em seu histórico de compra, proximidade em relação às lojas e outros aspectos de perfil de consumo. O entendimento de ativação comunitária (*communal activation*) representa uma mudança em relação aos canais de distribuição (praça), que cada vez mais se apresentam como distribuições *peer-to-peer* (ponta-a-ponta), provendo o acesso fácil a produtos e serviços que não necessariamente são seus, mas de outros clientes – como fazem *Airbnb* e *Uber*, por exemplo. Já o conceito de conversa (*conversation*), por fim, substitui o que é conhecido por promoção, que tradicionalmente sempre apresentou uma relação unilateral com as empresas enviando mensagens sobre seus produtos a possíveis clientes.

Para Cintra (2010, p. 8), “o marketing *online* é um processo destinado a vender produtos e serviços para um público-alvo usuário de internet e sistemas de informação *online*, junto com ferramentas e serviços *online*, de forma estratégica e coerente com o programa de marketing da empresa”.

Desta forma, segundo Torres (2009, p. 25), “não há mais separação entre produtor e consumidor. Não há mais exclusividade de produção nem na mídia nem no software. E o mais importante: não há mais distinção entre informação, entretenimento e relacionamento”.

### 2.2.1 Ferramentas de Marketing Digital

Conforme Gabriel (2010) afirma, no universo do marketing digital existem diversas plataformas e tecnologias digitais que podem embasar estratégias de marketing. Para o autor, as principais plataformas e tecnologias de marketing digital são: páginas digitais, e-mail marketing, realidades mistas, *mobile*, *social media*, busca, *games*, *displays* e entretenimento *online*.

Torres (2009), por sua vez, também levanta algumas ferramentas e estratégias para o marketing digital, sendo: marketing de conteúdo, mídias sociais, e-mail marketing, marketing viral, publicidade *online* e pesquisa *online*.

Se tratando especificamente de busca, conforme Jansen e Resnick (2006), os motores de busca da *web* geralmente apresentam pelo menos duas categorias de resultados de pesquisa: a primeira é um conjunto de links não-patrocinaados (ou seja, chamados de resultados orgânicos), que o próprio mecanismo determina usando seu algoritmo de correspondência. Já a segunda é composta por links patrocinados (pagos) que aparecem porque um indivíduo ou uma empresa adquiriu as palavras-chave que o pesquisador utilizou na busca.

### 2.3 PUBLICIDADE ONLINE

Ainda conforme Torres (2009), as possibilidades da publicidade online com as tecnologias disponíveis atualmente são variadas, e o modelo é completamente distinto do padrão de publicidade tradicional.

A internet, que hoje interliga milhões de computadores, *smartphones* e *desktops* de pessoas e organizações (Chiavenato, 2011), propiciou ao mundo das marcas uma série de oportunidades e desafios na relação com a publicidade e conseqüentemente para a publicidade no contexto digital. Conforme Pinho (2004, p. 96), a publicidade pode ser definida como “qualquer forma remunerada de apresentar ou promover produtos, serviços e marcas, feita por um patrocinador claramente identificado e veiculada nos meios de comunicação”.

Em relação à publicidade online, Torres (2009) cita que esta teve início com os banners publicados em sites, que eram modelos de anúncios comuns em mídia exterior, como outdoor e mídia impressa. Com o passar do tempo esses banners

passaram a ter novos recursos como animações, interações, sons e vídeos e a publicidade online passou a crescer muito além dos banners tradicionais. Para Pinho (2004, p. 119), “a flexibilidade da publicidade on-line é total, pois uma campanha pode ser lançada, atualizada ou cancelada imediatamente”.

Segundo Kendzerski (2009), a publicidade online possui características como a possibilidade de produzir ações segmentadas e específicas ao público-alvo, além da importante capacidade de avaliar os resultados das ações. Por outro ângulo, Tang, Zhang e Wu (2014) cita que o objetivo da publicidade online é de anunciar produtos, serviços ou bens públicos e, por fim, obter lucros ou consciência social. De acordo com Gabriel (2010), o rápido avanço da tecnologia e das plataformas digitais favorece a diversificação das ações de marketing, bem como a perspectiva de mensuração que o ambiente digital disponibiliza.

Dentre os demais objetivos da publicidade online, segundo o *Interactive Advertising Bureau* (IAB, 2016), pode-se separar em dois grupos principais: baseado em impressões e baseado em performance. O primeiro diz respeito ao valor gasto em publicidade pelo número total de “impressões”, enquanto o segundo é correspondente ao custo por clique, vendas, custo por aquisição, geração de *lead*, aplicação (ex.: pedido de cartão de crédito) ou participação direta na receita (por exemplo, % de comissão paga mediante venda).

Para fins de melhor compreensão deste estudo, também se faz necessário conceituar o que é uma impressão de publicidade online. Segundo Pinho (2004, p. 326), impressão se trata da “exposição de um banner em uma página *web*. Cada banner visto é contado como impressão”.

### **2.3.1 Tipos de Conteúdo**

Para Pinho (1990), a publicidade e seu conteúdo podem ser classificados das seguintes formas:

- a) Publicidade de Produto – divulga um produto, levando o consumidor ao conhecimento e posterior compra.

- b) Publicidade de Serviços – divulga um serviço, tais como bancos, seguradoras, etc. É preciso prestar atenção no tipo de conteúdo, pois se assemelha muito ao modelo propaganda institucional.
- c) Publicidade de Varejo – produtos anunciados por intermediários (varejistas). O varejo é um canal de distribuição e é caracterizado por se especializar na venda ao consumidor final. Esta modalidade também pode-se chamar de publicidade comercial.
- d) Publicidade Comparativa – aqui, são feitas alusões aos produtos e serviços concorrentes, evidenciando que as qualidades do anunciante são superiores às dos concorrentes.
- e) Publicidade Cooperativa – anúncio ou campanha para a venda realizada conjuntamente pelo fabricante com um ou mais lojistas.
- f) Publicidade Industrial – aplicada em um contexto industrial onde os bens são comercializados para intermediários que encaminharão os produtos ao consumidor final. Aqui, o objetivo é facilitar, otimizar e reduzir custos do processo de venda.
- g) Publicidade de Promoção – deve ser entendida como apoio às ações de promoção de vendas, utilizando meios de massa como: rádio, tv, jornal, revista, etc.

### 2.3.2 Formatos de Publicidade Online

Menezes (2003) afirma que os meios mais populares de se atingir o público-alvo na internet são: banner, e-mail, *streaming media* e mala direta eletrônica. Segundo levantamento do *Interactive Advertising Bureau* (IAB, 2016), as definições dos formatos de publicidade online e suas aplicações estão contidos na tabela 1:

Tabela 1 - Definições dos formato de publicidade e modelos de precificação

Formato	Modelo de precificação
Banner	Anunciante paga uma empresa online pelo espaço em uma ou mais páginas para mostrar um banner estático ou clicável ou um logo.

Patrocínio	<p>Anunciante paga por conteúdo e/ou experiências personalizadas, que podem ou não incluir elementos de anúncio como exibir publicidade, logomarcas, publiceditorial, ou vídeo pre-roll. Patrocínios são divididos em algumas categorias:</p> <ul style="list-style-type: none"> <li>• <i>Spotlight</i>: são páginas feitas sob-medida incorporando uma marca e abrigando uma coleção de conteúdos geralmente em torno de um tema relacionado.</li> <li>• <i>Advergaming</i>: podem variar de um anunciante comprando todos os blocos de anúncios em torno de um jogo ou um “patrocinado por” para criar uma experiência de jogo personalizada.</li> <li>• <i>Content &amp; Section</i>: acontecem quando um anunciante patrocina exclusivamente uma seção particular de um site ou e-mail (geralmente com conteúdo existente) repaginado com a identidade visual do patrocinador.</li> <li>• <i>Sweepstakes &amp; Contests</i>: podem variar de sorteios de marcas em sites até um concurso completo com inscrições e jurados.</li> </ul>
E-mail	<p>Banners, links ou patrocínios que aparecem em newsletter de e-mails, campanhas de e-mail marketing e outras comunicações comerciais. Isto inclui tanto anúncios dentro de um e-mail ou o e-mail inteiro.</p>
Busca	<p>Anunciante paga uma empresa digital para listar e/ou <i>linkar</i> seu site com palavras ou frases específicas de busca (incluindo buscas pagas). As categorias de busca englobam:</p> <ul style="list-style-type: none"> <li>• <i>Anúncios Pagos</i>: pagamentos feitos por cliques em links de texto que aparecem no topo ou ao lado dos resultados de busca para palavras-chave específicas. Quanto mais se paga, mais alta fica sua posição nas buscas. Os profissionais de marketing pagam somente quando um usuário clica no link do texto.</li> <li>• <i>Busca por Contexto</i>: pagamentos realizados para cliques em links de texto que aparecem em um artigo baseado no contexto do conteúdo, em vez de um usuário submetido por palavra-chave. O pagamento só ocorre quando o link é clicado.</li> <li>• <i>Inclusão Paga</i>: pagamentos acontecem para garantir que a URL do anunciante seja indexada pela ferramenta de busca. Nesta aplicação, o anunciante não paga por cliques.</li> <li>• <i>Otimização de Site</i>: anunciante paga para otimizar seu site a fim de melhorar sua posição nos resultados do motor de busca das páginas.</li> </ul>
Geração de Lead	<p>Taxas pagas pelos anunciantes para empresas que indicam ou encaminham consumidores potenciais qualificados ou fornecem informações sobre consumidores (demográficas, comportamentais e contato), por onde os consumidores optam por serem contatados pelo anunciante (e-mail, correio, telefone ou fax). Estes métodos são precificados baseados em performance (ex.: custo por ação, lead ou pesquisa) e podem incluir aplicações do próprio usuário, surveys, concursos ou registros.</p>
Classificados e Leilões	<p>Anunciante paga para uma empresa online para listar produtos ou serviços específicos (ex.: vagas de emprego, anúncios de imóveis, anúncios de automóveis, anúncios de leilão, páginas amarelas).</p>
<i>Rich Media</i>	<p>Exibir anúncios que carregam algum item de interatividade. Os anúncios de <i>Rich media</i> frequentemente incluem códigos Flash ou JavaScript e podem permitir usuários a ver e interagirem com produtos e serviços (ex.: rolando ou clicando no anúncio, abre-se uma multimídia do produto, seja uma descrição, expansão, animação, vídeo ou até mesmo um “test-drive” virtual dentro do anúncio). Os anúncios conhecidos como <i>Interstitials</i> se consolidaram dentro da categoria <i>Rich media</i> e representam um página total ou parcial de texto e imagem que aparecem na transição entre duas páginas de conteúdo. Os <i>Interstitials</i> podem incluir as seguintes variações:</p>

	<ul style="list-style-type: none"> <li>• <i>Splash screens</i>: uma página preliminar que antecede a página inicial comum de um site, sendo programada para se mover sobre a página inicial do site após um curto período de tempo.</li> <li>• <i>Pop-up / Pop-under</i>: um anúncio que aparece em uma página separada que carrega automaticamente sobre um conteúdo existente, sem banner.</li> <li>• <i>Daughter windows</i>: anúncio que é executado em uma janela separada associado a um banner apresentado simultaneamente. O conteúdo e o banner são normalmente mostrados primeiro, seguidos da <i>daughter window</i>.</li> <li>• <i>Superstitials</i>: anúncios que são distintos de <i>Interstitials</i> por conta da maior qualidade do anúncio, e eles são mostrados instantaneamente. and that they play instantly (ads are fully downloaded before they are displayed).</li> </ul>
Áudio Online	<p>Publicidade por áudio parcial ou total com suporte de programação disponível para consumidores em uma base de <i>streaming</i> entregue via conexão à internet mobile. Isso inclui uma grande variedade de serviços, como os seguintes:</p> <ul style="list-style-type: none"> <li>• Transmissão de áudio online em estações de rádio;</li> <li>• Estações de rádio com DJs profissionais ou amadores;</li> <li>• Serviços de áudio personalizado e sob-demanda que criam listas de reprodução com base nas preferências do usuário de artistas, faixas ou gêneros.</li> <li>• Música ou conteúdo falado entregue em um site ou aplicativo diferente. Por exemplo: serviços de música inserido dentro de um jogo.</li> </ul>
Vídeo Online	<p>Publicidade que aparece antes, durante ou após o conteúdo de vídeo em um player de vídeo (ou seja, pré-roll, mid-roll, post-roll video ads). Os anúncios de vídeo online incluem anúncios veiculados na TV e também podem aparecer em transmissões ao vivo ou em um vídeo para download. Anúncios conhecidos como vídeo <i>overlay</i> também são compreendidos como publicidade online em vídeo, incluindo pequenos anúncios que aparecem no topo de um vídeo.</p>
Publicidade Mobile	<p>Publicidade adaptada e entregue através de dispositivos móveis sem fio, como <i>smartphones</i> e <i>tablets</i>. Tipicamente, assumindo a forma de anúncios de exibição de mídia estática ou <i>rich media</i>, anúncios de mensagens de texto, anúncios de pesquisa ou locais de áudio / vídeo, essa publicidade geralmente aparece em sites otimizados para visualização em dispositivos móveis, aplicativos móveis, serviços de mensagens de texto (ou seja, SMS, MMS) ou em resultados de busca.</p> <p>Os formatos de publicidade móvel incluem: busca, banners, vídeo, áudio, patrocínios e <i>rich media</i>, além de outros anúncios publicitários para dispositivos móveis.</p>
Mídias Sociais	<p>Publicidade entregue em plataformas sociais, incluindo redes sociais, sites e aplicativos de jogos sociais, em todos os tipos de dispositivos, incluindo desktop, laptop, smartphone e tablet.</p>

Fonte: adaptado de *Interactive Advertising Bureau*, 2016.

### 2.3.3 Mídia Programática

De acordo com Watts (2016), é importante saber a diferença entre as formas tradicionais de publicidade online e a mídia programática, cujo termo é utilizado para descrever o processo de compra de publicidade englobando técnicas de segmentação

comportamental, *Real Time Bidding* (RTB) e *retargeting*. Desta maneira, a mídia programática abrange uma ampla gama de tecnologias que automatizam a compra, a colocação e otimização de publicidade. Para o autor, a crescimento da mídia programática tem a ver com tecnologia, sistemas e métodos mais eficientes em relação à coleta e análise de dados pessoais dos consumidores. Além disso, o *Big Data* permite que os profissionais de marketing também utilizem mídia programática para alcançar a pessoa certa, no momento certo e no contexto certo.

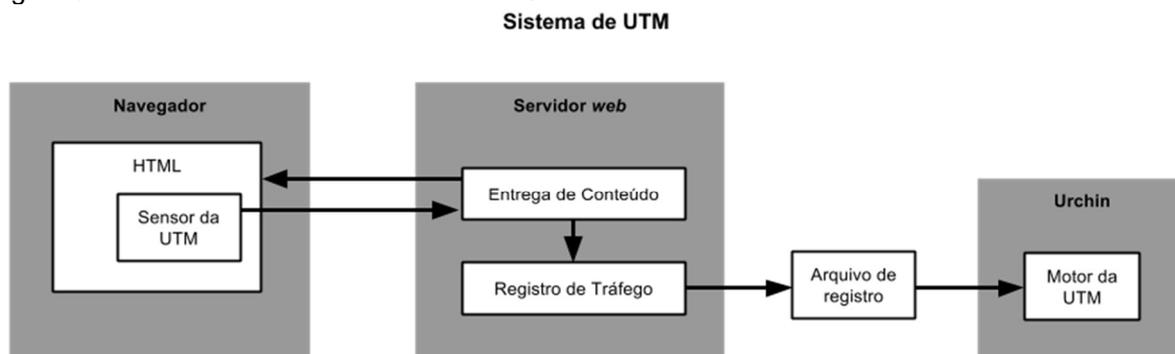
Para Almeida (2016, p. 62), “mídia programática é uma maneira de se comprar mídia de maneira programada e dinâmica, sem a necessidade de intermediação humana ou mesmo de publicação manual veículo a veículo de peças”.

#### 2.4 UTM'S (*URCHIN TRACKING MODULES*)

Segundo Delalana (2017), o termo *Urchin Tracking Module* (UTM) possui essa denominação porque a empresa desenvolvedora, adquirida pelo Google em 2005 e que deu origem ao Google Analytics, desenvolveu a tecnologia para ser utilizada inicialmente na ferramenta de análise de tráfego *Urchin Software*. As UTM's operam como marcadores, que quando incorporados à uma determinada URL conseguem rastrear os cliques, apontando a sua origem. Assim, é possível saber de onde e como o visitante chegou ou qual o formato de imagem está gerando mais conversões e visita em uma página ou post específico.

Para o Google (S/D), existem dois componentes no sistema de UTM: o sensor e o motor da UTM. O primeiro é um módulo instalado dentro do conteúdo específico, enquanto o segundo é a parte de processamento dos registros. O sensor UTM permite a coleta de dados do lado do cliente, que são combinados com os dados existentes do lado do servidor pelo motor da UTM, visando fornecer uma imagem mais precisa e completa da atividade do site. Na figura abaixo é possível compreender melhor o funcionamento do sistema de UTM.

Figura 3 - Funcionamento de um sistema de UTM



Fonte: Adaptado pelos Autores, 2017.

Segundo Greco (2016), com muitos profissionais de marketing digital utilizando o Google Analytics (GA), a forma sobre como um rastreamento de campanha online é feito na plataforma praticamente se tornou o padrão do mercado. O método do GA utiliza um conjunto de parâmetros UTM para identificar a origem, o termo, o conteúdo e a campanha. Tais parâmetros normalmente são ocorridos na URL e analisados pelo Google Analytics (GA) para preencher os relatórios analíticos mais apropriados.

Mitova (2017) cita que o acompanhamento de links via UTM ajuda os profissionais de marketing a responder importantes questões (exemplo: quantos usuários chegaram a página de registro de um determinado evento através de um link incluído no boletim semanal versus um *tweet*; ou quantas pessoas que vieram deste boletim realmente se cadastraram para o evento, etc.), e assim os insights que são coletados como resultado deste rastreamento ajudam a medir a eficácia de promoções e canais. O autor ainda cita que existem dois tipos de links existentes: regulares e longos, como mostrados no Quadro 01 abaixo:

Quadro 1 - Diferenças entre tipos de links

Links Regulares	Links Longos
“ <i>www.linkregular.com/produtos</i> ”	“ <i>www.linklongo.com/produtos/?utm_medium=email&amp;utm_source=email_provider&amp;utm_campaign=15Q1_Horizontal_B</i> ”

Fonte: Adaptado de Mitova, 2017.

Quando o código UTM é adicionado ao URL de uma página da *web*, o resultado é um link longo que contém informações de rastreamento, que permitem com que as ferramentas de análise que monitoram a página possam analisá-la e exibi-la em relatórios específicos ou relatórios personalizados. Segundo o Google (S/D), existem 5 parâmetros que podem ser adicionados aos URLs, listados na tabela 2:

Tabela 2 - Parâmetro de UTM

<b>Parâmetro</b>	<b>Objetivo</b>	<b>Exemplos</b>
utm_source (obrigatório)	Identifica qual site enviou o tráfego, e é um parâmetro obrigatório.	utm_source=Google
utm_medium	Identifica que tipo de link foi usado, como o custo por clique ou e-mail.	utm_medium=cpc
utm_campaign	Identifica uma campanha específica de produto ou uma campanha estratégica.	utm_campaign=spring_sale
utm_term	Identifica termos de busca.	utm_term=running+shoes
utm_content	Identifica o que especificamente foi clicado para trazer o usuário para o site, como um banner ou link de texto. É geralmente usado para teste A/B e anúncios segmentados por conteúdo.	utm_content=logolink or utm_content=textlink

Fonte: adaptado do Google pelos autores, 2017.

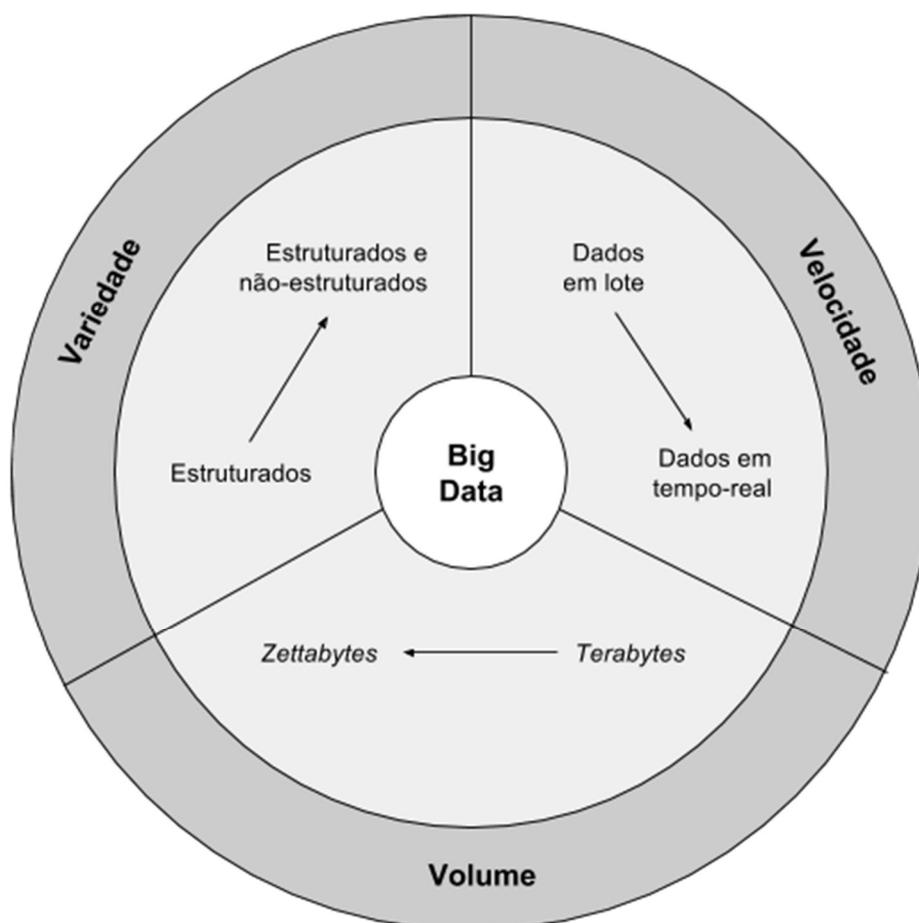
## 2.5 BIG DATA NO CONTEXTO DE MARKETING

De acordo com Rigby, Reicheld e Schefter (2002 apud Verhoef, Kooge e Walk 2016), os dados sempre estiveram à nossa volta por décadas. No entanto, nos anos 90, algumas empresas começaram a investir em grandes bases de dados de clientes, o que resultou na criação de registros de milhões de consumidores nos quais constavam informações armazenadas sobre comportamentos de compra, contatos e outras características de consumo. Conforme Verhoef, Kooge e Walk (2016), a propagação da internet e das mídias sociais permitiu que houvesse uma explosão de dados que se tornaram disponíveis para muitas companhias, sendo diariamente ou até mesmo em tempo real.

O volume de dados e a multiplicidade de suas fontes sofrem um crescimento exponencial, criando novas técnicas e desafios de aplicação. Os dados vêm de toda parte: desde sensores usados para condições climáticas, informações de voos e trânsito, *posts* em mídias sociais, e até vídeos ou fotos digitais (Bello-Organ, Jung e

Camacho, 2015). Para Laney (2001), existem três dimensões na gestão de dados, sugerido pelo modelo dos três V's - Volume, Variedade e Velocidade. Em "Understanding Big Data", Zikopoulos et al. (2012) descreve o modelo dos três V's através da figura 4:

Figura 4 - Modelo do 3 V's: volume, variedade e velocidade



Fonte: adaptado de Zikopoulos et al., 2012.

Zaslavsky et al. (2012) conceitua o modelo dos três V's da seguinte forma:

- **Volume:** Refere-se ao tamanho dos dados como terabytes (TB), petabytes (PB), zettabytes (ZB), etc.
- **Variedade:** Significa os tipos de dados. Além disso, diferentes fontes como sensores, dispositivos, redes sociais, a internet, telefones celulares, etc, irão produzir big data.
- **Velocidade:** Isso significa o quão frequentemente os dados são gerados. Por exemplo a cada milissegundo, segundo, minuto, hora, dia, semana, mês, ano,

etc. A frequência do processamento também pode diferir de acordo com os requerimentos do usuário, sendo que alguns dados precisam ser processados em tempo real e outros apenas quando necessário.

Segundo Gandomi e Haider (2015), o modelo de três V's se tornou a forma mais comum de descrever o que hoje se conhece por *Big Data*. A IBM (S/D) cita que os dados de *Big Data* estão sendo gerados a todo momento e que “muitos deles chegam em uma forma não-estruturada, tornando difícil colocá-los em tabelas estruturadas com linhas e colunas. Para extrair *insights* destes dados complexos, projetos de *Big Data* muitas vezes dependem de análises de ponta envolvendo ciência de dados e *machine learning*”. *Machine Learning* se trata do aprendizado de máquina, que é basicamente um modo de análise de dados que é capaz de automatizar a elaboração de modelos analíticos.

No contexto do marketing, muitas oportunidades podem ser vislumbradas com a utilização de técnicas de *Big Data*. De acordo com Li e Malthouse (2017), os pontos de contato dos consumidores com uma marca no mundo digital têm uma relação simbiótica com o *Big Data*: interações digitais criam o *Big Data*, que então pode ser usado para informar decisões subsequentes sobre os pontos de contato. Portanto, o escopo do *Big Data* pode ser entendido como pesquisas dos ambientes digitais onde ocorrem cada ponto de contato de um consumidor com uma marca, ou que irão ocorrer em um futuro próximo. A tabela 3 sintetiza oportunidades com *Big Data* no contexto de marketing.

Tabela 3 - Sumário de Oportunidades com *Big Data*

Oportunidade	Descrição
Dados não-estruturados	Desenvolvimento de aplicativos para a rede, textos, imagens ou vídeo para coleta de dados e identificar insights e mensurar ambientes de publicidade.
Dados de Painel	Métodos de dados de painel para inferir causalidade, mensurar efeitos de publicidade, testar teorias.
Métodos de Aprendizagem Estatística	Análise preditiva e sistemas de recomendação personalizada, decisões de publicidade programadas, etc.

Montagem de Quebra-cabeça	Adquirir e juntar conjuntos de dados para obter uma imagem mais completa dos clientes.
Parcerias com os Anunciantes	Obter acesso a conjuntos de dados novos e exclusivos, teorias de teste de campo e abordagens de otimização, <i>link</i> para resultados financeiros.
Pesquisa Interdisciplinar	Parceria com cientistas de dados / informática, cientistas de marketing, econométristas para desenvolver algoritmos para otimizar a execução de anúncios e as teorias de teste.
Dispositivos Móveis	Compreender a experiência do consumidor com a tecnologia móvel, otimizar as interações dos clientes através de canais móveis.
Marcas à escuta	Monitore o que os clientes estão dizendo nas mídias sociais sobre a marca, como eles usam produtos com a Internet das coisas, responda conforme apropriado.
Dados contextuais	Use dados sobre o contexto atual em que o cliente se encontra para personalizar mensagens e melhorar a eficácia.
Evitação de Anúncios	Estudar os comportamentos de evitação do anúncio, quantificar o valor do anúncio deslocado no tempo.
Publicidade como um processo interativo	Compreenda como os relacionamentos se desenvolvem ao longo do tempo através de interações de marca de dois sentidos.
Endereçamento de TV e <i>streaming</i>	Entregue a mensagem de TV personalizada certa aos consumidores certos no momento certo no dispositivo certo e pague o valor certo.
Resultados Financeiros	Demonstrar efeitos da propaganda sobre os resultados financeiros.

Fonte: Li e Malthouse, 2017.

Em relação ao *e-commerce*, Akter e Wamba (2016) citam que algumas empresas como Amazon e Netflix já capturam vários tipos de dados (por exemplo: pedidos, carrinhos, visitas, usuários, links, palavras-chave, catálogo de navegação, dados sociais), que podem ser colocados em quatro categorias: (a) dados de transação / atividade comercial, (b) dados de fluxo de clique, (c) dados de vídeo e (d) dados de voz. No comércio eletrônico, de acordo com os autores, dados são essenciais para rastrear comportamentos de compra – que são coletadas ao longo do tempo usando a navegação do consumidor e os pontos de transação – a fim de personalizar as próximas ofertas.

### 2.5.1 *Big Data Analytics (BDA)* – Análise de *Big Data*

A movimentação em torno de *Big Data* e análise de *Big Data* provavelmente foi gerada principalmente pela web e comunidades de *e-commerce*, de acordo com Chen, Chiang e Storey (2012). Os autores citam que significantes transformações no mercado foram realizadas pelos principais fornecedores de comércio eletrônico, como a Amazon e eBay através de suas plataformas inovadoras e altamente escaláveis e com sistemas de recomendação de produtos.

Para Verhoef, Kooge e Walk (2016), atualmente existem três fatores fundamentais que estão mudando a forma de analisar dados – do tradicional para o *Big Data Analytics (BDA)*. Primeiro, os novos tipos de dados e especificamente dados não-estruturados estão sendo analisados; segundo, no ambiente digital e do *Big Data*, surgiram novos desafios e questões, como avaliar os investimentos em novas ferramentas de publicidade online (anúncios em mecanismos de pesquisa e afiliados); terceiro, estão sendo desenvolvidas novas técnicas analíticas que podem explicar o enorme fluxo contínuo de dados, e como consequência um enorme campo a ser explorado por analistas.

Segundo Li e Malthouse (2017), os grandes dados são normalmente armazenados em alguma espécie de base de dados. Para preparar e usar os conjuntos de dados de *Big Data*, os autores citam ser importante, como um ponto inicial, o conhecimento sobre bancos de dados relacionais e linguagem de consulta estruturada (*query language* ou SQL). Muitos outros tipos de plataformas e bases de dados são utilizadas para guardar tipos específicos de *Big Data* (ex. Hadoop), e ter conhecimento sobre elas é especialmente importante para a execução das mensagens que irão existir em anúncios.

## 2.6 EFICIÊNCIA DE PUBLICIDADE ONLINE

No contexto deste estudo, eficiência é um critério utilizado para a mensuração e avaliação do desempenho da publicidade *online* e seus resultados. Segundo Chiavenato (2011, p. 149), eficiência “é a razão entre o esforço e o resultado, entre a despesa e a receita, entre o custo e o benefício resultante”. Assim, o que garante a

eficiência da publicidade *online* é a sua mensuração entre a despesa e a receita gerada pela ação.

De acordo com Maximiano (2000, p. 115), “eficiência realiza tarefas de maneira inteligente, com o mínimo de esforço e com o melhor aproveitamento possível dos recursos”.

### **2.6.1 Indicadores de Desempenho de Marketing**

Para Gabriel (2010), determinar indicadores que serão utilizados para mensurar resultados é essencial para averiguar se os objetivos de um plano de marketing foram atendidos ou não.

O conceito base de uma métrica, segundo Farris et al (2012) é um conjunto sistemático de medidas, válidas utilizadas em qualquer disciplina, que buscam mensurar, quantificar fenômenos em uma tendência, dinâmica ou característica. O conceito de métrica de Marketing, segundo Kotler e Keller (2012) é um “conjunto de mensurações que ajudam a quantificar, comparar e interpretar a performance de Marketing”. Indo além, Farris et al (2012) coloca a que o profissional de Marketing não está imune, comparando, por exemplo, às ciências exatas e engenharia, à ter que fazer avaliações quantitativas.

### **2.6.2 Métricas para Eficiência Online em E-Commerce**

Serão expostas a seguir as principais métricas online e seus conceitos. Numa tabela anexa abaixo, seguem as equações para chegar em cada métrica.

- **Exposições, hits e pageviews:** Segundo Farris *et al* (2012), monitora-se o tráfego de um site a partir número de vezes em que o site foi acessado (*pageviews*). Trata-se da oportunidade, portanto, se um usuário web visualizar determinado conteúdo. Já hits se refere aos arquivos fornecidos aos usuários web. Uma determinada página pode ter mais de um arquivo, portanto, essa métrica leva em conta quantos arquivos tem determinada página.
- **Tempo de exibição em rich media:** “Rich media é um termo utilizado para mídia interativa que permite que os consumidores se envolvam mais ativamente com um painel, uma propaganda de TV ou até mesmo com uma propaganda tradicional de internet” (Farris *et al*, 2012). Utilizar esta métrica ajuda a monitorar por quanto tempo os anúncios estão prendendo a atenção dos consumidores.
- **Taxa de interação de rich media:** Segundo Farris *et al* (2012), taxa de interação de rich media é a avaliação de em até qual ponto o consumidor em potencial se envolve com a propaganda. Tal preocupação ocorre para mensurar a taxa de envolvimento dos consumidores.
- **Taxa de cliques:** Para Farris *et al* (2012), a taxa de cliques é a porcentagem de exposições comparado ao número de cliques de um anúncio online. Os sites têm o propósito de incentivar uma ação, e isso é o que está sendo medido.
- **Custo por exposição, custo por clique e custo por pedido:** Farris *et al* (2012) coloca estas métricas como sendo o custo médio da exposição, dos cliques e dos clientes que efetivamente realizaram uma compra.
- **Visitas, visitantes e abandono:** Farris *et al* (2012) coloca visitas como sendo a quantidade de vezes que os indivíduos requisitam ao servidor uma página pela primeira vez. Esta métrica é também conhecida como “sessões”. Visitantes, também conhecido como visitantes únicos, são quantos indivíduos requisitam as páginas do servidor em determinado período pela única vez.

- **Custo por Mil Exposições:** razão do valor gasto em publicidade pelo número total de exposições (chamados de "impressões" neste estudo) em milhares.
- **Número de compras:** o número de compras efetuadas no *e-commerce* é um indicador de eficiência.

Tabela 4 - Equações para métricas online

Métrica	Equação
Exposições, hits e pageviews	$Hits = Pageviews (n^{\circ}) \cdot N^{\circ} \text{ de arquivos da página}$
Tempo de exibição em rich media	$\frac{\text{Tempo de exibição em rich media } (n^{\circ})}{\text{Tempo total de exibição em rich media}} = \frac{\text{Total de impressões em rich media}}{\text{Total de impressões em rich media}}$
Tempo de Interação de rich media	$\frac{\text{Taxa de interação em rich media } (\%) \cdot \text{Total de impressões de rich media com interação}}{\text{Total de impressões em rich media}}$
Taxa de cliques	$\text{Taxa de cliques} = \frac{\text{Cliques } (n^{\circ})}{\text{Exposições } (n^{\circ})}$
Custo por exposição, custo por clique, custo médio por pedido.	$\text{Custo por exposição} = \frac{\text{Custo do anúncio } (\$)}{\text{Quantidade de exposições } (n^{\circ})}$ $\text{Custo por clique} = \frac{\text{Custo do anúncio } (\$)}{\text{Quantidade de cliques } (n^{\circ})}$ $\text{Custo por pedido} = \frac{\text{Custo do anúncio } (\$)}{\text{Pedidos } (n^{\circ})}$
Visitas, visitantes e abandono.	Sem equação (por definição)
Custo por mil exposições	$\text{Custo por mil exposições (CME)} = \frac{\text{Custo do comercial } (\$)}{\text{Exposições geradas } (n^{\circ} \text{ em milhares})}$

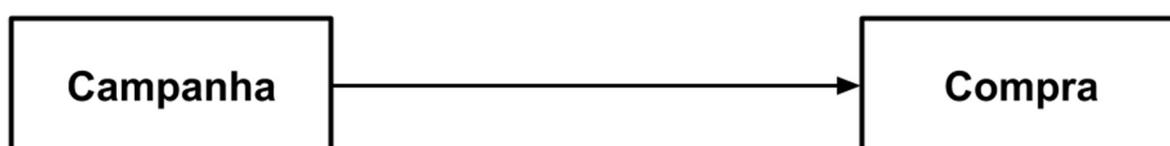
Fonte: Farris *et al*, 2012.

Dependendo do objetivo da publicidade *online*, a quantidade de venda (compras feitas por consumidores no e-commerce em questão) pode ser considerada uma métrica de eficiência, baseada em performance (visto em "Publicidade Online", seção 2.3 deste estudo).

### 3 METODOLOGIA

Partindo do referencial teórico, é possível começar a compreender a elaboração da hipótese do estudo, que abrange os conceitos de campanha e compra. A hipótese é de que usuários com origem em campanha têm propensão significativamente maior de realizar uma compra em relação a usuários provenientes de origem orgânica:

Figura 5 - Modelo gráfico de hipótese



Fonte: Elaborado pelos Autores, 2017.

Apoiada no conceito de que um dos objetivos da publicidade online é de anunciar produtos e obter lucros, e no indicador de eficiência de número de compras, a hipótese foi formatada.

Para este estudo, foram utilizados dados secundários. Dados secundários, segundo Malhotra (2012, p. 36), são "dados coletados para uma finalidade diferente daquela do problema em pauta". Ou seja, para esta pesquisa, os dados não foram coletados pelos pesquisadores.

No total foram quatro bases de dados, sendo uma a descrição de número de acessos, dias e horários de visitas, a segunda uma descrição de custos e quantidade de impressões de publicidade que foram veiculadas, a terceira com a descrição de dados das sessões de navegação e a quarta uma compilação (a partir da base de sessões de navegação) individualizada de cada usuário.

As bases foram fornecidas por uma empresa de serviços de análise de dados que não poderá ser identificada neste estudo e será chamada pelo codinome Empresa Y, sendo esta uma prestadora de serviços de um *e-commerce* (também com lojas físicas) no setor de cosméticos e produtos de beleza, sendo esta tratada como Empresa X neste estudo.

Este estudo foi realizado com dados entre os dias 15 a 31 de agosto de 2017. Tais datas foram escolhidas já que viabilizaram a pesquisa. O número total de acessos

neste período foi de 336.377, sendo 323.674 indivíduos que realizaram os acessos divididos entre 267.459 visitas únicas (visitas que ocorreram ao menos uma vez dentro de um dos dias). O número de visitas sempre será maior ou igual ao número de indivíduos, assim como o número de indivíduos será sempre maior ou igual ao número de visitas únicas. Isto ocorre porque um determinado indivíduo pode acessar mais de uma vez por dia (contando somente como uma visita única), assim como uma pessoa pode acessar mais de uma vez em dias diferentes (contando como mais de uma visita única).

### 3.1 BASE DE DADOS SOBRE NÚMEROS, DIAS E HORÁRIOS DE VISITAS

Esta base de dados foi fornecida diretamente pelo sistema da Empresa Y, que forneceu um usuário e uma senha para acesso da equipe deste estudo.

Dentro deste sistema, foi selecionado o filtro de cliente (Empresa X) e um filtro de período de tempo, conforme já citado, entre os dias 15 e 31 de agosto de 2017.

A partir daí, o sistema disponibiliza alguns dados descritivos que foram utilizados nesta pesquisa, tais como: visitantes únicos e visitantes. A diferença dos dois dados encontra-se na definição, segundo a própria Empresa Y:

- Visitas únicas: "Os usuários que realizaram pelo menos uma sessão no período selecionado. Inclui usuários novos e recorrentes".
- Visitas: "Número total de visitas no período. Uma visita é um período que um usuário fica ativamente engajado com seu *website*"

É possível filtrar estes dados a partir dos dias do mês, dias da semana e horários, todos compreendidos dentro do filtro de datas inicialmente programado. Por exemplo: quando o filtro "Audiência por dias da semana" foi selecionado, com a seleção "Visitas", apareceram todos os dias da semana e a quantidade de visitas efetuadas nestes dias.

A partir destes dados, foram criadas quatro tabelas no Excel a fim de demonstrar o comportamento de acesso nos dias do mês, da semana e nos períodos de horas do dia.

A primeira tabela criada compreendeu as seguintes variáveis (colunas): Dia (do mês), Dia da semana, Nº de visitas, Nº Visitas únicas e Nº Impressões. A última

variável é um cruzamento feito com dados da base de dados que será explicada na seção 3.2. Foram preenchidas manualmente as variáveis, já que o sistema da Empresa Y não permite a exportação destes dados em formato de tabela. Em seguida, um gráfico de linhas foi gerado, tendo as quatro variáveis com linhas em cores diferentes.

A segunda tabela criada compreendeu as seguintes variáveis: Dia (do mês), Dia da semana, Orçamento/dia (também retirado da base que será explicada na seção 3.2), Número de impressões e Custo por mil (CPM). O gráfico foi de linha do orçamento gasto aos dias.

A terceira tabela criada compreendeu as seguintes variáveis: Dia (da semana), Nº de visitas e Nº de visitas únicas. Foi gerado um gráfico de colunas, com duas colunas por dia da semana, cada uma indicando, respectivamente, o Nº de visitas e o Nº de visitas únicas.

A última e quarta tabela compreendeu as seguintes variáveis: Horário, Nº de visitas e Nº de visitas únicas. A variável horário era correspondente a faixas de horário de 1h ao longo do dia. Foi gerado um gráfico de colunas mistas, sendo a primeira parte da coluna correspondente ao Nº de visitas únicas e a segunda o Nº de visitas, em relação às faixas de horário.

### 3.2 BASE DE DADOS DE CUSTOS DE PUBLICIDADE E QUANTIDADE DE IMPRESSÕES

A segunda base de dados contém dados de impressão de publicidade e os custos envolvidos. Ela foi fornecida pela Empresa Y e a figura 6 demonstra as variáveis que constam nesta base de dados:

Figura 6 - Estrutura básica da base de dados de impressão de publicidade e custos.

start_date	end_date	advertiser	campaign_name	campaign_budget	impressions	total_spent
15/08/17	15/08/17	Empresa X	[HKN] Marcas Agosto	R\$ 2.000,00	1464	R\$ 4,63
16/08/17	16/08/17	Empresa X	[HKN] Marcas Agosto	R\$ 2.000,00	1722	R\$ 6,14
17/08/17	17/08/17	Empresa X	[HKN] Marcas Agosto	R\$ 2.000,00	1564	R\$ 5,01
18/08/17	18/08/17	Empresa X	[HKN] Marcas Agosto	R\$ 2.000,00	1804	R\$ 7,07
19/08/17	19/08/17	Empresa X	[HKN] Marcas Agosto	R\$ 2.000,00	0	R\$ 0,00
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Fonte: Base de dados fornecida pela Empresa Y.

A data final solicitada desta base de dados é o dia 31 de agosto de 2017, conforme consta. O significado de cada variável, segundo a Empresa Y, é:

- start\_date: data de início na qual aquela linha de dados representa;
- end\_date: data final na qual aquela linha de dados representa;
- advertiser: anunciante selecionado;
- campaign\_name: nome da campanha daquele período;
- campaign\_budget: orçamento máximo de campanha;
- impressions: número de impressões de publicidade efetivados no período (não inclui envios de *E-mail Marketing*);
- total\_spent: valor total gasto no período selecionado.

Esta base de dados teve importância descritiva, pois demonstrou dados financeiros e de impressão, os quais são demonstrados no futuro. Foi possível, inclusive, estimar o custo médio em publicidade de sobre cada compra efetivada.

### 3.3 BASE DE DADOS DAS SESSÕES DE NAVEGAÇÃO

Os dados de navegação vieram inicialmente em formato *JavaScript Object Notation* (JSON) e foram posteriormente exportados para o formato *Comma Separated Value* (CSV). Os dados vieram em JSON e foram abertos no programa RStudio (Assistente de programação para a linguagem de programação R) utilizando uma biblioteca específica RJSON. Não houve envolvimento dos pesquisadores nestas conversões, pois foram feitos por uma especialista em Estatística e Análise de Dados que prestou suporte neste estudo.

A base de dados utilizada possui sete tipos de variáveis brutas, apresentadas na sequência. Segundo a Empresa Y que forneceu os dados:

- Data: data de acesso a sessão;
- Hora: horário de acesso a sessão (precisão em segundos);
- PageTitle: título da página atual;
- PageUrl: URL referente a página atual;
- ReferrerUrl: URL anterior a página atual;
- SessionId: identificação da sessão da linha. A duração de cada sessão é de no máximo 30 minutos para cada página acessada. Uma nova sessão será gerada quando o usuário clicar em uma nova página.
- VisitorId: identificação única de cada visitante

A figura 7 exemplifica, um trecho de linha da base de dados e a estrutura de links encontrados. Frisa-se que uma linha pode ser referente a mais do que um consumidor, já que este pode ter navegado por diversas páginas da Empresa X enquanto procurava pelos seus produtos. O desenho é ilustrativo e não representa linhas reais da base de dados.

Figura 7 - Estrutura básica da base de dados de sessões.

Data	Hora	PageTitle	PageUrl	ReferrerUrl	SessionId	VisitorId
15/08/2017	00:00:30	Página Principal	www.empresax...	google.com.br/	1	1
15/08/2017	00:01:00	Produto X	www.empresax...	www.empresax...	2	1
15/08/2017	00:00:40	Página Principal	www.empresax...	outlook.live.com/	3	2
15/08/2017	00:01:15	Produto Y	www.empresax...	www.empresax...	4	2
15/08/2017	00:01:00	Produto Z	www.empresax...	NA	5	3

•           •           •           •           •           •           •  
 :           :           :           :           :           :           :  
 •           •           •           •           •           •           •

Fonte: adaptado pelos autores a partir da base de dados.

Conforme pode-se observar na figura 7, uma única linha será sempre referente a um dado de navegação de um único usuário. Contudo, um mesmo usuário pode representar mais do que uma linha. Também pode-se observar que o ReferrerUrl referente ao primeiro acesso pode ser desde a página de uma ferramenta de busca (tal como o Google), um e-mail (neste caso, presume-se que o usuário acessou a partir de uma campanha disparada ao seu e-mail), ou simplesmente nada (ou "NA"), referindo-se ao usuário que acessou o site diretamente.

Tendo em vista que com esta base de dados, já foi possível saber quantas sessões de navegação foram abertas no período. Foi possibilitado, então, já calcular o número de sessões médias por indivíduo e o número de sessões médias por visita. Para isso, bastou dividir o número total de linhas desta base de dados, subtraindo um (a primeira linha é reservada ao nome das variáveis), pelo número de visitas e pelo número de indivíduos.

### **3.3.1 Identificação de Sessões Provenientes de Campanhas (Variável Independente)**

A partir da descrição feita anteriormente, é possível entender se a origem do usuário é a partir de campanha ou não. A variável "PageUrl" carrega essa informação a partir da origem de UTM, conforme foi descrito no referencial teórico. Assim como o "ReferrerUrl" carrega a informação se houve ou não a compra por aquele usuário, a partir de informações de transação: visto que, no momento que ocorre a transação, o usuário é redirecionado a página principal do *e-commerce*. Sendo assim, a informação da compra sempre estará presente no "ReferrerUrl" através de algumas palavras-chave fornecidas pela própria Empresa X.

Portanto, a partir disso, já é possível extrair algumas informações: se o usuário veio ou não de campanha e também se ele efetuou ou não uma compra. Para realizar o registro disso, abriu-se a base de dados no Excel.

O Instituto Nacional de Pesquisas Espaciais, INPE, possui um tutorial ilustrado de como importar arquivos CSV para o Excel. Não foi utilizado todo o tutorial para o caso a seguir. Ao abrir o Excel, importou-se a base de dados então foi aberto, em seguida, um assistente de importação de texto para colunas.

- Na primeira página do assistente de importação de texto para colunas, clique-se em "Delimitado - Caracteres como vírgulas ou tabulações separam cada campo. Avançou-se para seguinte.
- O delimitador que foi utilizado é "Vírgula", portanto qualquer outro delimitador (Lembrando que o delimitador pode ser outro de acordo com a base de dados) foi desmarcado. Clicou-se em avançar. A partir daí, concluído.

Ao abrir os dados corretamente, criou-se uma coluna extra ao lado do "PageUrl". Essa coluna serviu para criar uma fórmula para localizar usuários que vieram de campanhas.

Para isso, foi utilizada a função "LOCALIZAR", do Excel. Conforme consta no manual do Suporte do Microsoft Office, tal função serve para localizar cadeias de texto dentro de textos maiores. A sintaxe da função é "=LOCALIZAR(texto\_procurado;no\_texto;[núm\_inicial])". Para a localização de cadeias de texto com "utm\_campaign" (portanto, provenientes de campanhas), foi utilizado da forma "=LOCALIZAR("utm\_campaign";D2)", sendo D2 a linha inicial do "PageUrl". Tal fórmula foi lançada a todas as células abaixo. Os valores como resultados foram números correspondentes ao caractere onde o texto "utm\_campaign" se encontra. Caso não encontre, o Excel gerará um resultado "#VALUE!", é um erro que demonstra valores não encontrados. A fim de corrigir o texto com erros, criou-se uma nova coluna e nela foi copiado o conteúdo da coluna com os erros, porém no formato "Somente Valores". Essa nova coluna continha todos os erros "#VALUE!" em formato de texto, o que facilita para posteriormente a correção do erro.

Para criar uma variável corrigindo todos os erros citados anteriormente, criou-se uma nova coluna, e foi utilizada a função "SE" do Microsoft Excel. Esta função é uma condicional que permite fazer comparações lógicas, de acordo com o Suporte do Microsoft Office. A sintaxe básica da função é: "=SE(teste\_lógico;[Valor se verdadeiro];[Valor se falso])". A partir do ponto que todas as células anteriormente criadas contém ou um número maior do que 1, ou um texto "#VALUE!", gerou-se então a seguinte fórmula: "=SE(E2>0;1;0)". Lembrando que a coluna criada anteriormente é a "E" do Excel. Então, foi feita uma nova coluna, somente com "0" e "1", onde 0 indica a não existência de "utm\_campaign" e, portanto, aquela sessão não teve origem em campanha e 1 indica que aquela sessão teve origem em campanha. Finalizando, aí, a variável descritiva de sessões que contém usuários provenientes de campanhas.

### **3.3.2 Identificação de Sessões que Finalizaram em Compra (Variável Dependente)**

O procedimento para detecção de usuários que realizaram uma compra é muito similar ao procedimento anterior. Contudo, desta vez interessam somente algumas faixas de texto presentes no "ReferrerUrl". Os textos são "transaction" (indicam um link onde ocorreu a transação financeira) ou "Paypal", que indicam que o usuário realizou uma transação financeira via o serviço Paypal e realizou a compra. Tais textos se encontram no "ReferrerUrl", pois o usuário é redirecionado a página principal no momento que realiza a compra, deixando somente um rastro. Tais conclusões ocorreram sob orientação da Empresa Y.

Para tal, repetiu-se o procedimento anterior para o "ReferrerUrl". Criou-se uma coluna com o título "Compra1" e buscou-se o texto "transaction" pelo procedimento de "=LOCALIZAR("transaction";G2)", sendo G2 o primeiro dado da coluna "ReferrerUrl". Tal fórmula é lançada para todas as células abaixo. O resultado foi, na coluna H, o número da ordem a qual se encontra o caractere inicial de "transaction" ou o "#VALUE!", quando nada é encontrado. Em seguida, criou-se uma nova coluna, a coluna I, com o título "Paypal1" e buscou-se o texto "paypal", desta forma, sem diferenciação maiúscula ou minúscula. Novamente, foi gerado o erro "#VALUE!" para a coluna que não possui compras via Paypal ou o número da ordem do caractere onde o texto inicia. Em seguida, transformou-se a coluna da variável "Compra1" em texto e o mesmo se faz com a coluna da variável "Paypal1". Para isso, criou-se novas colunas para estas variáveis, chamadas "Compra2" e "Paypal2", copiou-se todo o conteúdo de "Compra1" e "Paypal1" para o "Compra2" e "Paypal2", respectivamente, só que desta vez no formato de "Somente Valores".

Para, então, transformar em "0" e "1" (0 para "Não comprou" e 1 para "Comprou"), criaram-se as colunas "Compra3" e "Paypal3", e a lógica "SE" foi utilizada novamente para cada uma delas, no formato "=SE(L2>0;1;0)", sendo L a coluna de "Compra2". Utilizou-se a lógica para todas as demais células abaixo. A coluna da variável "Compra3" então recebe um "0" para sessões onde não houve compra e "1" para sessões onde houve compra. O mesmo procedimento repete-se para a coluna da variável "Paypal3".

Feito isso, em ambas as colunas "Compra3" e "Paypal3", houve a identificação de sessões com compra e não compra. Porém, foi necessário unificá-las a uma única

variável a qual indica se há ou não há compra. Para isso, utilizou-se a lógica "ou", do Excel. De acordo com o Suporte Microsoft Office, a função lógica "ou" retorna "Verdadeiro" caso um dos elementos seja "Verdadeiro". Criou-se, então, uma coluna com a variável "Compra0", no qual foi utilizada a fórmula "=OU(M2;N2)", sendo, respectivamente, as colunas "M" e "N" as variáveis "Compra3" e "Paypal3". Utilizou-se essa fórmula para todas as células abaixo. A coluna da variável "Compra0" encontrava-se na coluna "O". As saídas na coluna "Compra0" serão "Verdadeiro" ou "Falso" ("Verdadeiro" para o caso de naquela sessão haver compra e "Falso" para o caso de não ter havido compra).

Em seguida, criou-se uma variável "CompraFinal" na coluna "Q". Esta coluna tem como objetivo transformar "Verdadeiro" e "Falso" em "1" e "0" respectivamente, mantendo o padrão. Para isso, utilizou-se "=SE(O=1;1;0)". Repete-se essa fórmula para todas as células abaixo. A coluna "CompraFinal" é a variável que indica todas as compras que foram realizadas nas sessões analisadas.

### 3.4 BASE DE DADOS INDIVIDUAIS

Para então realizar a análise de dados de usuários únicos, portanto, criou uma nova base na qual cada linha representa um usuário, em vez de uma sessão de navegação. Foi utilizada a base da seção 3.3 para esta transformação.

Nesta conversão, foi utilizado linguagem de programação R, sendo feita com auxílio externo de uma pessoa especialista na área. A nova formatação de dados utilizada possui as seguintes variáveis:

- V1: identificação numérica após a conversão via R;
- PageTitle: título da página onde se encontra o segmento "utm\_campaign";
- PageUrl: url da "PageTitle";
- ReferrerUrl: URL da página de referência de localização dos segmentos de compra;
- SessionId: identificação da sessão onde se encontra o segmento de compra;
- VisitorId: identificação do visitante;
- Data: data da visita àquela sessão;

- Hora: horário de entrada da primeira sessão;
- Id: combinação de SessionId e VisitorId no momento da compra e do "utm\_campaign"
- count: contagem da quantidade de páginas navegadas por aquele usuário;
- cp: identificação se o usuário teve origem de campanha;
- buy: identificação se o usuário comprou ou não.

Como forma de ilustrar a nova base de dados, sem que tenham dados reais, segue a figura 8:

Figura 8 - Estrutura da base de dados por usuário.

Data	Hora	PageTitle	PageUrl	ReferrerUrl	SessionId
15/08/2017	00:00:30	Página Principal	www.empresax...	google.com.br/	1500
15/08/2017	00:01:00	Produto X	www.empresax...	www.empresax...	2000
15/08/2017	00:00:40	Página Principal	www.empresax...	outlook.live.com/	2500
15/08/2017	00:01:15	Produto Y	www.empresax...	www.empresax...	3000
15/08/2017	00:01:00	Produto Z	www.empresax...	NA	3500

▪	▪	▪	▪	▪	▪
▪	▪	▪	▪	▪	▪
▪	▪	▪	▪	▪	▪

VisitorId	Id	count	cp	buy
1	1500 - 1	5	1	1
2	2000 - 2	2	0	0
3	2500 - 3	4	0	1
4	3000 - 4	1	1	0
5	3500 - 5	2	0	0

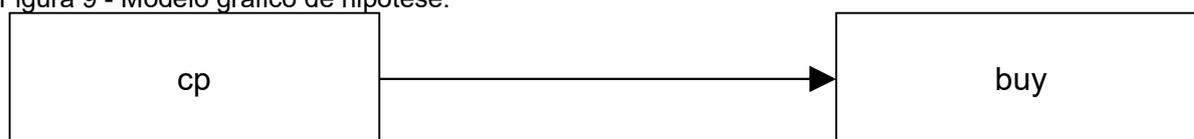
▪	▪	▪	▪	▪
▪	▪	▪	▪	▪
▪	▪	▪	▪	▪

Fonte: Elaborado pelos autores a partir da base de dados.

Esta nova configuração de base possibilitou que fossem feitas comparações de usuários orgânicos (aqueles que não foram proveniente de anúncios) e usuários provenientes de anúncios em relação ao comportamento de compra. Tais comparações são uma das bases deste estudo. Além de ter melhorado a questão da contagem de dados.

Tendo isso, foi levantada a hipótese de que usuários com origem em campanha têm propensão significativamente maior de realizar uma compra em relação a usuários provenientes de origem orgânica, ou seja, aqueles os quais não houve esforços de comunicação para atraí-los. A figura 9 demonstra o modelo demonstrado:

Figura 9 - Modelo gráfico de hipótese.



Fonte: Elaborado pelos Autores, 2017.

Antes, criou-se, com auxílio do software SPSS, uma tabela cruzada entre as variáveis "buy" e "cp", a fim de verificar se há cruzamentos entre elas. Para verificar a hipótese, foi utilizado uma regressão logística, também com auxílio do SPSS, na qual a variável dependente é "buy" e a variável independente é "cp". Ambas variáveis categóricas binárias.

### 3.4.1 Contagem de Casos

Para fazer a contagem, utilizou-se a função "CONT.SE" do Excel. Conforme o Suporte Microsoft Office indica, a função "CONT.SE" conta o número de células que atendem determinado critério. A sintaxe da função é "`=CONT.SE([intervalo];[critério])`". Essa função foi utilizada para realizar a contagem do número de compras e do número de sessões provenientes de campanha.

Para a contagem do número de sessões que originaram compra, usou-se a fórmula "`=CONT.SE(Q:Q;1)`". Ou seja, a fórmula contou quantos casos na coluna "CompraFinal" são igual a "1", isto é, as células na qual foi detectada a compra.

Já para a contagem do número de sessões que possuem origem em campanha, a fórmula é "`=CONT.SE(F:F;1)`". Lembrando que a variável que está na coluna "F" é "utm\_campaign". O retorno dessa fórmula é a contagem de quantas sessões foram provenientes de campanha.

A contagem de usuários únicos foi mais simples. Para isso, selecionou-se a coluna "VisitorId", copiou-se em uma outra coluna, chamada "Usuarios\_Unicos". Segundo o Suporte Microsoft Office, pode-se remover dados duplicados de uma coluna selecionando-a e, a partir da aba "Dados", clica-se em "Remover Duplicatas". Esta nova coluna só conterá visitantes únicos. Para realizar a contagem, bastou checar a numeração da última linha onde se encontram estes usuários, e subtrair 1, pois a primeira linha é utilizada para nomear as variáveis.

### **3.4.2 Contagem de Compras por horário do dia**

Para realização da contagem de compras por faixa de horário do dia, abriu-se a base usuários únicos no Excel e criou-se um filtro na faixa de horário. Os horários estavam no formato "hh:mm:ss" ("hora":"minuto":"segundo"). Reconfigurou-se, então, para o formato "hh:mm" ("hora":"minuto"). Para isso, a coluna dos horários foi selecionada e criou-se uma formatação de células customizada no formato "hh:hh". Depois dessa configuração, as faixas correspondentes aos segundos desapareceram, permanecendo somente o horário e o minuto.

Após a criação do filtro, criou-se 24 novas abas no Excel, cada uma correspondendo a uma faixa de 1h do dia, nomeando-as no formato "00h - 01h", "01h - 02h", e assim por diante. Na realidade, "00h - 01h" seria "00h - 00h 59 min e 59 seg.", porém, a fim de reduzir o tamanho da variável e facilitar a visualização, se optou pela primeira forma.

Após isso, o processo foi manual: na base de dados principal, filtrou-se o horário no formato "00:", "01:", e assim por diante. Ao colocar os ":" após a primeira hora de interesse na faixa de horário, selecionou-se todos os horários que começam com dessa forma, impossibilitando a entrada de qualquer outro horário. Após criado o filtro para cada hora, copiou-se a coluna de horários e "buy" para a aba correspondente criada anteriormente. Repetiu-se o processo para as 24 faixas de horário.

Em seguida, em cada aba, foi realizada uma contagem de compras realizadas no horário. Utilizou-se, numa célula ao lado da variável "buy" com a seguinte fórmula: "=CONT.SE(B:B;1)", sendo B a coluna de "buy". Com isso, contou-se a quantidade de compras na faixa de horário correspondente. Repetiu-se o processo para as 24 faixas de horário.

Depois, criou-se uma nova aba, com o título "Compilação". Titulou-se a coluna "A" como "Horário" e a coluna "B" como "Nº de Compras". Titulou-se as células "A2" até a célula "A25" com as faixas de horário e a célula "A26" com "Total". Preencheu-se as células da variável "Nº de Compras" com o número de compras de acordo com a contagem feita. Depois disso, somou-se todo o conteúdo na linha "B26". Um histograma foi gerado para auxiliar a apresentação dos dados.

### **3.4.3 Contagem de Compras por dia do mês**

Basicamente o mesmo processo explicado na seção 3.4.2 foi aplicado para esta análise. A diferença principal é que ao contrário dos horários, as datas estavam em um formato que não necessita nenhuma alteração ("dd/mm/aaaa" - "dia/mês/ano"). Bastou, então, criar o filtro nelas e gerar uma aba do Excel para cada um dos 16 dias.

Para cada data filtrada, foi copiado e colado na aba correspondente às colunas de data e compra. E então, foi utilizada a fórmula de contagem "=CONT.SE(B:B;1)" para realizar a contagem. A variável "buy" estava localizada na coluna B.

Então, após realizada a contagem para todos os dias analisados, criou-se uma compilação em uma única tabela, utilizando uma nova aba do Excel, com as colunas das datas analisadas e o número de vendas realizada no dia. Na última linha, foi calculado, então, o total de pedidos.

Contudo, também foi destacado o investimento diário, calculando, aí, pela divisão do orçamento diário pelo número de vendas, o Custo por pedido (Conforme seção 2.6.2 do Referencial Teórico). Depois, então, na última linha, foi calculada a média do Custo por Pedido por cliente proveniente de campanha.

### **3.4.4 Análise do tipo de anúncio utilizado**

Conforme consta na seção 2.3.1 da fundamentação teórica deste estudo, existem vários tipos de conteúdo publicitário. Ao analisar o tipo de anúncio utilizado pela Empresa X (Exemplo na figura 10), foi possível classificá-la como publicidade de varejo, ou seja, "produtos anunciados por intermediários (varejistas). O varejo é um canal de distribuição e é caracterizado por se especializar na venda ao consumidor final. Esta modalidade também pode-se chamar de publicidade comercial." (Seção 2.3.1). A marca original foi coberta e substituída.

Figura 10 - Exemplo do tipo de anúncio.



Fonte: Empresa X – Marca coberta e substituída pelos Autores, 2017.

Os demais anúncios que foram utilizados, com a marca substituída por "Empresa X", encontram-se no Apêndice A deste estudo.

## 4 ANÁLISE DOS RESULTADOS

### 4.1 ANÁLISE DESCRITIVA

Conforme já colocado na metodologia, o período selecionado da Empresa X teve 336.377 visitas, sendo 323.674 indivíduos que realizaram os acessos divididos entre 267.459 visitas únicas. É importante ressaltar estas informações porque elas são a base dos resultados.

A tabela 5 é o primeiro cruzamento de dados feito na etapa 3.1 da metodologia. A seção 2.3.3 conceitua impressão de publicidade online como "Exposição de um *banner* em uma página *web*. Cada *banner* visto é contato como impressão".

Tabela 5 - Cruzamento entre os dias do mês, dias da semana, nº de visitas, nº de visitas únicas e impressões de publicidade

<b>Dia</b>	<b>Dia da Semana</b>	<b>Nº de Visitas</b>	<b>Nº Visitas Únicas</b>	<b>Nº Impressões</b>
15/08	Terça-feira	15862,00	11758,00	1464,00
16/08	Quarta-feira	15251,00	12167,00	1722,00
17/08	Quinta-feira	15313,00	11890,00	1564,00
18/08	Sexta-feira	16704,00	13789,00	1804,00
19/08	Sábado	15402,00	12412,00	0,00
20/08	Domingo	15343,00	13640,00	0,00
21/08	Segunda-feira	19243,00	15862,00	1374,00
22/08	Terça-feira	16347,00	12576,00	1355,00
23/08	Quarta-feira	16672,00	12691,00	12392,00
24/08	Quinta-feira	17473,00	13497,00	23095,00
25/08	Sexta-feira	25009,00	18960,00	22944,00

26/08	Sábado	20397,00	16727,00	0,00
27/08	Domingo	12349,00	9667,00	0,00
28/08	Segunda-feira	41714,00	35216,00	27412,00
29/08	Terça-feira	19062,00	13700,00	19283,00
30/08	Quarta-feira	38000,00	29681,00	18789,00
31/08	Quarta-feira	16236,00	13226,00	21739,00
<b>Total</b>		<b>336377,00</b>	<b>267459,00</b>	<b>154937,00</b>

Fonte: Dados da Empresa X fornecidos pela Empresa Y. Elaborado pelos Autores, 2017.

O gráfico 2 destaca as informações da tabela 5 cruzando-as em ao longo dos dias:

Gráfico 2 - Cruzamento dos dados da Tabela 5.



Fonte: Tabela 5. Elaborado pelos Autores, 2017.

Neste gráfico, fica bem visível picos de investimento em publicidade online que ocorreram a partir da segunda metade do período, e, não somente isso, mas também picos de acesso que ocorreram no final do mês. A Empresa Y não esclareceu o porquê

de ter mantido os investimentos baixos até o dia 22 de agosto, porém esclareceu que não há impressões de publicidade aos finais de semana, conforme o gráfico e os dados demonstram.

Visando analisar os aspectos monetários, a tabela 6 cruza os dias e dias da semana do período com as informações orçamentárias fornecidas. O custo por mil (CPM) é uma métrica de eficiência explicada na seção 2.6.2 (Métricas Online), com o nome "Custo por mil exposições":

Tabela 6 - Cruzamento entre os dias do mês e os dias da semana, com o orçamento gasto no dia, número de impressões e custo por mil (CPM)

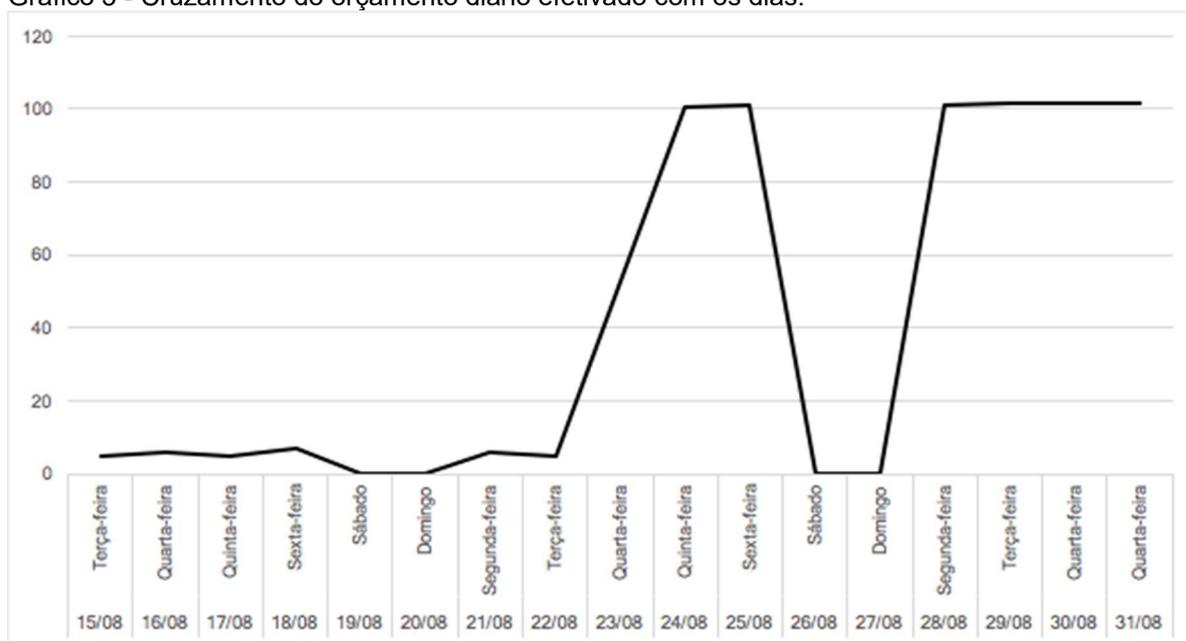
<b>Dia</b>	<b>Dia da Semana</b>	<b>Orçamento/dia</b>	<b>Número de Impressões</b>	<b>CPM</b>
15/08	Terça-feira	R\$4,63	1464,00	R\$3,16
16/08	Quarta-feira	R\$6,14	1722,00	R\$3,57
17/08	Quinta-feira	R\$5,01	1564,00	R\$3,20
18/08	Sexta-feira	R\$7,07	1804,00	R\$3,92
19/08	Sábado	R\$0,00	0,00	R\$0,00
20/08	Domingo	R\$0,00	0,00	R\$0,00
21/08	Segunda-feira	R\$5,81	1374,00	R\$4,23
22/08	Terça-feira	R\$5,08	1355,00	R\$3,75
23/08	Quarta-feira	R\$52,70	12392,00	R\$4,25
24/08	Quinta-feira	R\$100,71	23095,00	R\$4,36
25/08	Sexta-feira	R\$100,92	22944,00	R\$4,40
26/08	Sábado	R\$0,00	0,00	R\$0,00
27/08	Domingo	R\$0,00	0,00	R\$0,00

28/08	Segunda-feira	R\$100,88	27412,00	R\$3,68
29/08	Terça-feira	R\$101,35	19283,00	R\$5,26
30/08	Quarta-feira	R\$101,37	18789,00	R\$5,40
31/08	Quarta-feira	R\$101,36	21739,00	R\$4,66
<b>Total</b>		<b>R\$693,03</b>	<b>154937,00</b>	

Fonte - Dados da Empresa X fornecidos pela Empresa Y. Elaborado pelos Autores, 2017.

O gráfico 3 cruza o orçamento diário efetivado com os dias do mês e os dias da semana.

Gráfico 3 - Cruzamento do orçamento diário efetivado com os dias.



Fonte: Elaborado pelos Autores, 2017.

Devido ao fato de ser mídia programática, conforme a Empresa Y destaca, por mais que o orçamento tenha se mantido praticamente constante entre os dias 28 de agosto e 31 de agosto, o número de impressões caiu até o dia 30 de agosto. Isto ocorre porque houve aumento no custo da publicidade neste período.

Entende-se que no nível a partir do dia 22/08, a publicidade online passa a apresentar resultados bem mais razoáveis em termos de acesso, portanto, devem ser

mantidos neste nível. E os investimentos aos finais de semana, interrompidos, poderiam ser mantidos a fim de gerar um resultado constante.

O terceiro cruzamento de dados descrito na seção 3.1 da metodologia é dos dias da semana com o número de visitas e número de visitas únicas, conforme a tabela 7.

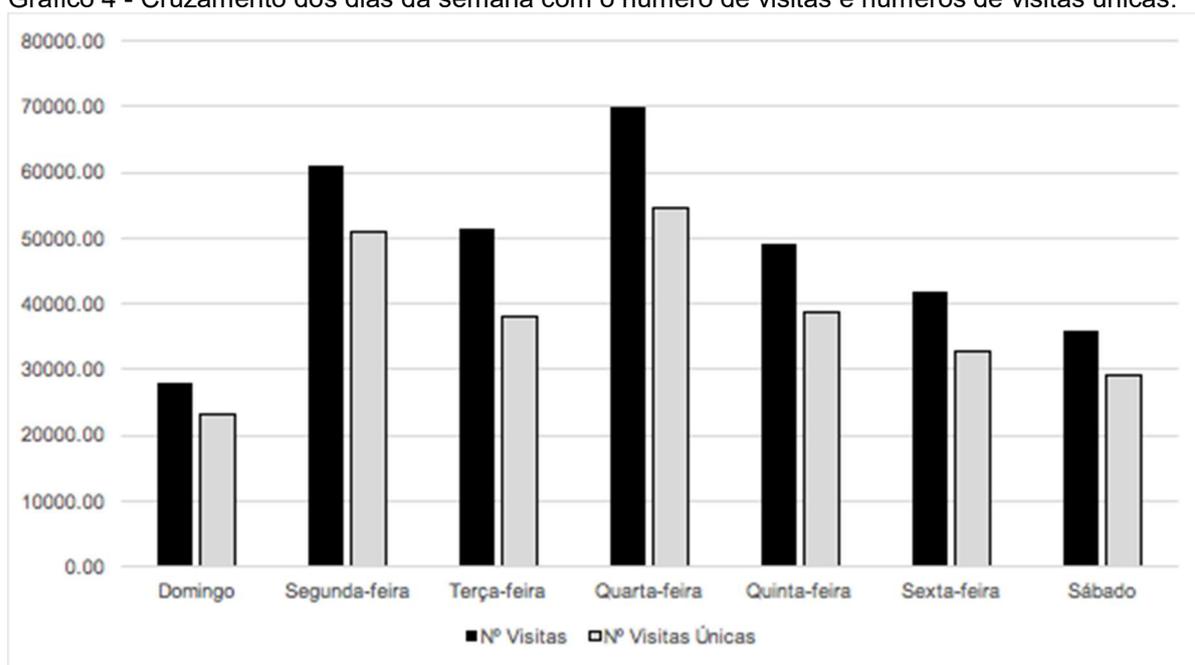
Tabela 7 - Cruzamento entre os dias de semana, número de visitas e número de visitas únicas.

<b>Dias da Semana</b>	<b>Nº Visitas</b>	<b>Nº Visitas Únicas</b>
Domingo	27692,00	23307,00
Segunda-feira	60957,00	51078,00
Terça-feira	51271,00	38034,00
Quarta-feira	69923,00	54539,00
Quinta-feira	49022,00	38613,00
Sexta-feira	41713,00	32749,00
Sábado	35799,00	29139,00
<b>Total</b>	<b>336377,00</b>	<b>267459,00</b>

Fonte: Elaborado pelos Autores, 2017.

O gráfico 4 cruza os gráficos da tabela 7, para melhor visualização:

Gráfico 4 - Cruzamento dos dias da semana com o número de visitas e números de visitas únicas.



Fonte: Elaborado pelos Autores, 2017.

A partir destes dados, conclui-se que existe um pico de visitas nos dias de segunda a sexta-feira em relação aos sábados e domingos. Porém, entre os dias 15 e 31 de agosto, as semanas não estão completas, pois o período é composto por três terças, quartas e quintas-feiras, enquanto o restante dos dias tem apenas duas amostras de cada. Portanto, existe a tendência que a segunda-feira seja o dia mais acessado do *e-commerce* da Empresa X. Contudo, presume-se que a tendência também seja que não haja grande diferença entre as visitas ao longo da semana.

A quarta análise descritiva buscou entender a diferença entre as visitas ao longo das horas do dia, conforme a tabela 8:

Tabela 8 - Cruzamento entre as 24h do dia, com o número de visitas e número de visitas únicas.

Horário	Nº Visitas	Nº Visitas Únicas
00h - 01h	8245,00	6526,00
01h - 02h	4783,00	4524,00
02h - 03h	2741,00	2453,00
03h - 04h	1953,00	1700,00

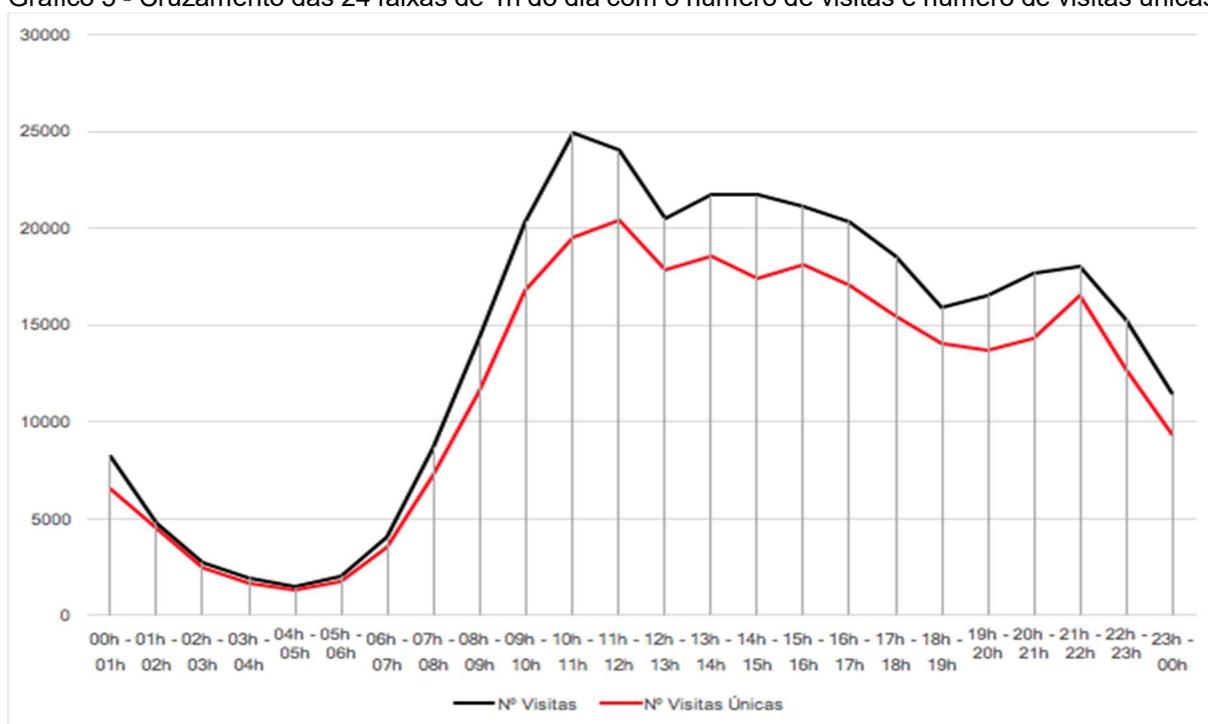
04h - 05h	1460,00	1322,00
05h - 06h	2021,00	1744,00
06h - 07h	4031,00	3504,00
07h - 08h	8660,00	7216,00
08h - 09h	14299,00	11540,00
09h - 10h	20363,00	16740,00
10h - 11h	24963,00	19539,00
11h - 12h	24027,00	20413,00
12h - 13h	20520,00	17834,00
13h - 14h	21722,00	18563,00
14h - 15h	21691,00	17428,00
15h - 16h	21122,00	18135,00
16h - 17h	20353,00	17055,00
17h - 18h	18523,00	15447,00
18h - 19h	15899,00	14011,00
19h - 20h	16526,00	13705,00
20h - 21h	17669,00	14262,00
21h - 22h	18047,00	16473,00
22h - 23h	15289,00	12749,00

23h - 00h	11470,00	9385,00
<b>Total</b>	<b>336377,00</b>	<b>282268,00</b>

Fonte: Elaborado pelos Autores, 2017.

No gráfico 5 pode-se observar os cruzamentos dos horários do dia com os dados de visitas e visitas únicas demonstradas na tabela 8:

Gráfico 5 - Cruzamento das 24 faixas de 1h do dia com o número de visitas e número de visitas únicas.



Fonte: Elaborado pelos Autores, 2017.

Percebe-se que o pico de acesso ficou entre 10h e 11h. Contudo, apesar da tendência descendente durante a tarde, há um novo pico de acesso, no período da noite, entre as 21h e 22h. O período entre 22h e 5h é descendente, sendo o início da tendência de subida a partir das 6h.

Outra observação que se deve fazer a partir destes dados é o fato de a somatória dos visitantes únicos, quando divididos por horário, estar diferente da somatória dos usuários únicos nos demais gráficos. Isso ocorre porque, caso um determinado usuário esteja acessando o site às 19h59, por exemplo, e continue navegando a partir das 20h, ele passará a ser contabilizado como um novo usuário único no período das 20h às 21h. Só não ocorrerá a recontagem de caso o usuário

feche, por exemplo, às 21h05 e reabra as 21h30: ele continuará sendo apenas um visitante único. Por esta razão, ao dividir o dia entre usuários únicos em períodos de 1h, a contagem será maior.

Conforme seção 3.3 da metodologia, sobre a base de dados das sessões de navegação, extraiu-se que foram abertos, no período selecionado neste estudo, 1.208.426 sessões. A partir deste, tendo a informação de que foram houve 336.337 visitas de 323.674 indivíduos, pode-se chegar a dois resultados: a média de sessões navegadas por visita é 3,59 e a médias de sessões acessadas por pessoa é de 3,73.

Foram efetivadas 344 compras no período analisado. Distribuídas, ao longo do mês. Frisa-se que o número de compras é métrica de eficiência neste estudo, conforme explicado na seção 2.6.2 neste trabalho.

O número de compras em relação ao horário do dia, conforme feito na seção 3.4.2. da Metodologia, gerou a tabela 9 (tabela de nº de compras versus horário do dia):

Tabela 9 - Número de compras por faixa de horário.

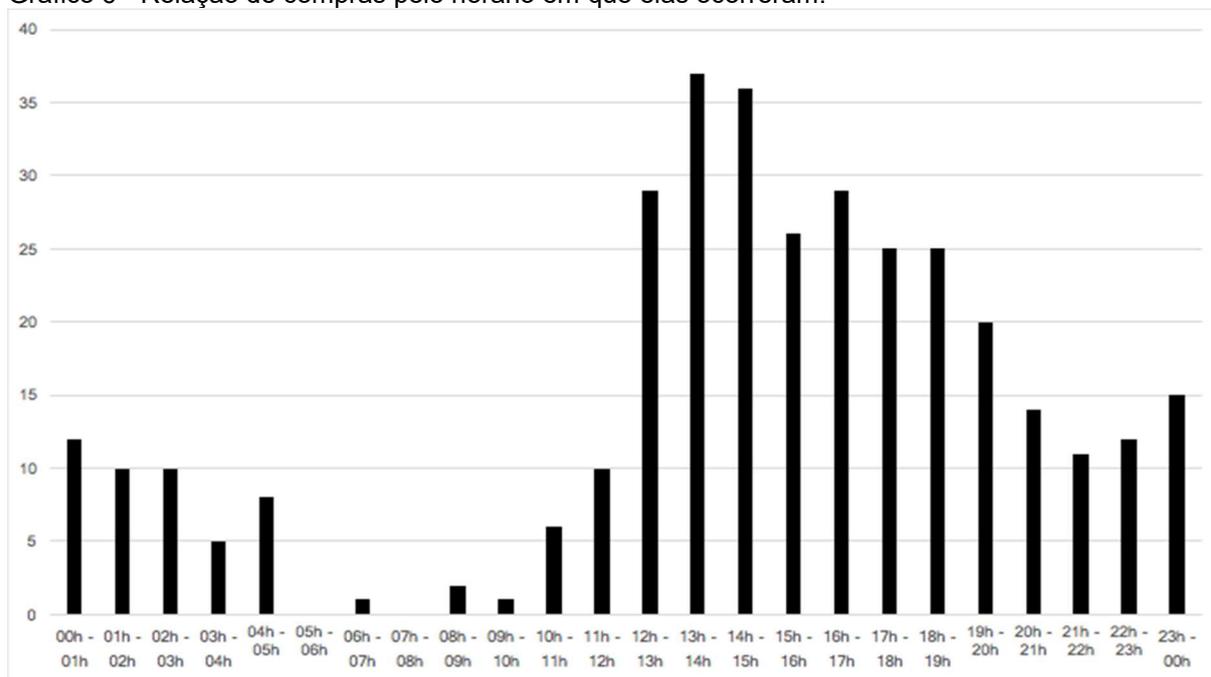
<b>Horário</b>	<b>Nº de Compras</b>
00h - 01h	12
01h - 02h	10
02h - 03h	10
03h - 04h	5
04h - 05h	8
05h - 06h	0
06h - 07h	1
07h - 08h	0
08h - 09h	2

09h - 10h	1
10h - 11h	6
11h - 12h	10
12h - 13h	29
13h - 14h	37
14h - 15h	36
15h - 16h	26
16h - 17h	29
17h - 18h	25
18h - 19h	25
19h - 20h	20
20h - 21h	14
21h - 22h	11
22h - 23h	12
23h - 00h	15
<b>Total</b>	<b>344</b>

Fonte: Elaborado pelos Autores, 2017.

Para facilitar a visualização, o gráfico X foi elaborado a partir da tabela 9:

Gráfico 6 - Relação de compras pelo horário em que elas ocorreram.



Fonte: Elaborado pelos Autores, 2017.

Percebe-se que existe uma relação entre as compras e a quantidade de acessos nos períodos do dia, porém, não sendo uma relação perfeita (Ex.: o horário de maior acesso possui apenas 6 compras no período estudado).

Depois disso, conforme foi analisada a quantidade de compras efetuadas pelos clientes em relação aos dias do mês. Foi adicionado o orçamento publicitário investido naquele dia, que também foi cruzado, podendo gerar o custo por pedido. É possível observar esta relação na tabela 10:

Tabela 10 - Cruzamento entre dias do mês, número de compras, orçamento investido no dia e custo por pedido.

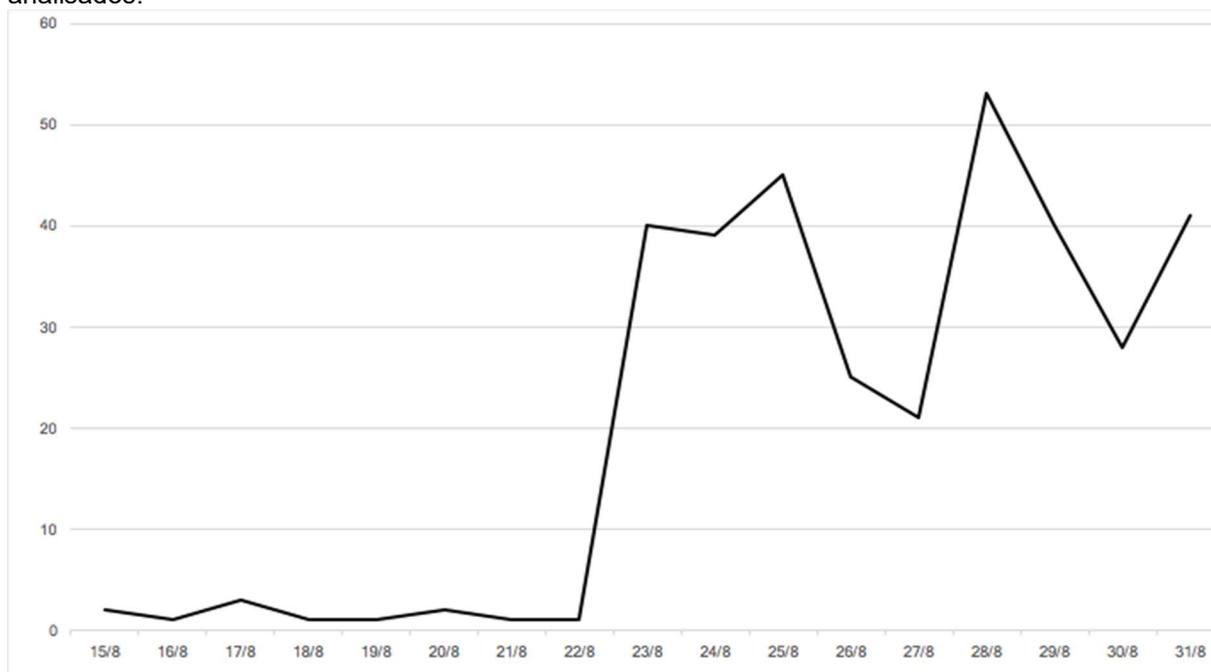
Dia	Nº de Compras	Orçamento/Dia	Custo por Pedido
15/08/17	2	R\$4,63	R\$2,32
16/08/17	1	R\$6,14	R\$6,14
17/08/17	3	R\$5,01	R\$1,67

18/08/17	1	R\$7,07	R\$7,07	
19/08/17	1	R\$0,00	R\$0,00	
20/08/17	2	R\$0,00	R\$0,00	
21/08/17	1	R\$5,81	R\$5,81	
22/08/17	1	R\$5,08	R\$5,08	
23/08/17	40	R\$52,70	R\$1,32	
24/08/17	39	R\$100,71	R\$2,58	
25/08/17	45	R\$100,92	R\$2,24	
26/08/17	25	R\$0,00	R\$0,00	
27/08/17	21	R\$0,00	R\$0,00	
28/08/17	53	R\$100,88	R\$1,90	
29/08/17	40	R\$101,35	R\$2,53	
30/08/17	28	R\$101,37	R\$3,62	
31/08/17	41	R\$101,36	R\$2,47	
<b>Total</b>	<b>344</b>	<b>R\$693,03</b>	<b>Média</b>	<b>R\$2,63</b>

Fonte: Elaborado pelos Autores, 2017.

O gráfico 7 demonstra a relação entre os dias do mês e a quantidade de compras efetivadas, conforme a tabela anterior.

Gráfico 7 - Relação entre a quantidade de compras efetuadas pelos clientes e os dias do mês analisados.



Fonte: Elaborado pelos Autores, 2017.

#### 4.2 ANÁLISE DA RELAÇÃO ENTRE EXPOSIÇÃO POR CAMPANHA E EFETIVAÇÃO DA COMPRA

A tabela 11 é um cruzamento a relação da exposição dos clientes por esforços de campanha e as compras. A variável "buy" indica se houve ou não compra naquela linha de usuário individual e cp indica se aquele consumidor em específico foi impactado ou não por alguma ação de campanha publicitária. Ambas as variáveis são binárias.

Tabela 11 - Tabulação Cruzada entre as variáveis "buy" e "cp".

		buy				
		0		1		
cp	0	198411	99,89%	214	0,1077%	<b>198625</b>
	1	124920	99,90%	130	0,1040%	<b>125050</b>
Total		<b>323331</b>		<b>344</b>		<b>323675</b>

Fonte: Elaborado pelos Autores, 2017.

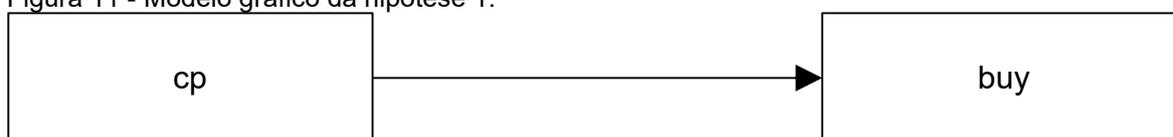
A tabela indica é que houveram 323.675 indivíduos que acessaram o *e-commerce* da Empresa X. Do todo, 198.625 vieram de forma orgânica, sendo que 214 realizaram uma compra e 125.050 vieram por campanha e, destes, 130 realizaram uma compra.

A partir disso, seguindo o modelo já demonstrado na seção 3.4 da metodologia deste estudo, apresentou-se a seguinte hipótese:

- Hipótese 1 (h1): Usuários com origem em campanha têm propensão significativamente maior de realizar uma compra em relação a usuários provenientes de origem orgânica.

O modelo gráfico, então, da hipótese 1 pode ser resumido conforme a figura 11:

Figura 11 - Modelo gráfico da hipótese 1.



Fonte: Elaborado pelos Autores, 2017

Após realizado a regressão logística, a qual "cp" foi a variável independente e "buy" a variável dependente, logo no teste Omnibus (tabela 12) de modelo de coeficientes, a hipótese 1 foi refutada, já que não foi significativo enquanto o teste Hosmer e Lemeshow (tabela 13) foi significativo.

Tabela 12 - Testes de Omnibus de Modelos de Coeficientes.

		Chi-square	df	Sig.
<b>Step 1</b>	<b>Step</b>	0,104	1	0,747
	<b>Block</b>	0,104	1	0,747
	<b>Model</b>	0,104	1	0,747

Fonte: Elaborado pelos Autores, via teste no SPSS, 2017.

Tabela 13 - Teste de Hosmer e Lameshow.

Step	Chi-Square	df.	Sig.
1	0,000	0	.

Fonte: Elaborado pelos Autores, via teste no SPSS, 2017.

Portanto, vale a hipótese nula, conforme descrita abaixo:

**Hipótese nula (h<sub>0</sub>):** Usuários com origem em campanha não têm propensão significativamente maior de realizar uma compra em relação a usuários provenientes de origem orgânica.

Por fim, foi calculado um dado descritivo extra, a partir do dado exato dos clientes que compraram vindos via campanha. O custo médio de pedido destes compradores foi de R\$ 5,33, o que só pode ser viável dependendo do *ticket médio* dos produtos.

## 5 CONSIDERAÇÕES FINAIS

### 5.1 APLICAÇÕES TEÓRICAS

O primeiro grande desafio deste estudo foi lidar com uma quantidade tão alta de dados, o que gera um conhecimento teórico muito grande de como limpar e analisar bases de *Big Data*.

A partir da análise de UTM's, foi possível identificar exatamente quais consumidores vieram de campanha e quais vieram de forma orgânica. E não somente isso, foi possível também analisar fatores de seus comportamentos que seriam impossíveis de serem visualizados através de um *software* ao exemplo do oferecido pela Empresa Y.

Dentre as vantagens destas análises, estão o aumento do escopo de informações relevantes que podem ser extraídas de uma base de dados que existe e está armazenada.

Ter como visualizar estes dados também permitiu analisar uma métrica de indicação de desempenho de eficiência extremamente importante: custo por pedido. Analisada da maneira clássica, ela se afasta muito do seu valor real, conforme já destacado na introdução destas considerações finais.

Ainda destacando a métrica acima, considerada a mais importante deste estudo, foi possível, com as análises de UTM's e contagem de usuários, chegar ao número do custo exato de pedido por cliente proveniente de publicidade.

E, por fim, o estudo mais importante, que foi a verificação se consumidores oriundos de campanhas de publicidade *online* são mais propensos a comprar. Isso é impossível de ser feito sem uma análise de UTM's e gerou conhecimento para que sejam aplicados em estudos futuros os quais seja importante a conhecer clientes que compraram.

### 5.2 APLICAÇÕES GERENCIAIS

A proporção em relação à de visitantes que vêm de forma orgânica e compram é extremamente parecida com a quantidade de pessoas que vêm via campanhas. Foi testado e provado que não há diferença. Contudo, pode-se observar que os clientes que vêm de forma orgânica, apesar de não serem mais propensos a compra, costumam

menos e geram resultados tão bom quantos. A propensão a compra de ambos é a mesma no caso da Empresa X, porém, se faz necessária a análise caso a caso para então entender o perfil do consumidor e assim gerar decisões gerenciais de qualidade. No caso da Empresa X, deve-se entender a importância dessa conclusão em termos de implicações gerenciais: se não há diferenças, quanto menos investir para atrair visitantes, melhor.

Com isso, ao entender essa diferença, é possível gerar estratégias de atração de visitantes que sejam mais baratas em relação a publicidade, por exemplo, a fim de gerar vendas. O ponto-chave é entender a como o comprador de origem orgânica e o comprador de origem de campanha se comportam em relação ao consumo. Entendendo essa relação e fazendo um balanceamento de custos das estratégias, é possível baixar o custo por pedido, um dos principais indicadores de eficiência.

Outra grande possibilidade ao se calcular a origem de UTMs e visualizar, em separado, a origem dos compradores é a possibilidade de calcular com mais precisão o custo por pedido individual do cliente não-orgânico. Por exemplo, observou-se a diferença entre o custo médio por pedido ao calcular todos os clientes (R\$ 2,63) e somente os clientes que vieram por campanha (R\$ 5,33). A diferença demonstrou que o custo real do pedido de clientes que vieram via campanha é 102% superior à média anterior, geralmente utilizada. Não é um erro aceitável.

O grande ponto neste estudo é: *e-commerces* como a Empresa X não têm o acompanhamento das vendas provenientes de campanhas separadamente das vendas de compradores que vieram de forma orgânica. A administração destes dados ainda é genérica e necessita melhorias, através de análises mais profundas, tal como a análise de UTM's.

Mas, ao, analisar o comportamento de compra ao longo das horas do dia, é possível ver que os períodos de maior venda e de maior acesso são diferentes. Neste estudo, não foi possível criar uma aplicação teórica a fim de compreender a razão disto, contudo, é possível entender de que, mantendo-se este padrão, deve-se procurar estratégias focadas nestes horários, a fim de potencialização dos resultados, ou mesmo observar a possibilidade de estratégias em horários de pouca venda, a fim de otimizar a performance.

### 5.3 LIMITAÇÕES E SUGESTÕES DE ESTUDOS FUTUROS

É preciso, primeiramente, entender que não haviam dados como valores das compras para serem utilizados em demais análises, dificultando a criação diversas informações sobre eficiência. Em futuras análises é importante que esses dados sejam analisados e incorporados no processo.

Quanto à coleta de dados, é importante salientar a confiabilidade dos dados obtidos, mas existiram limitações e dificuldades no processo e obtenção dos dados o que geraram contratempos para as análises. Como, por exemplo, a limitação do presente estudo que diz respeito à ausência do *ticket* médio total e *ticket* médio dos produtos vendidos, impossibilitando análises desses fatores. Como também a ausência de *tickets* de clientes que vieram via campanha ou de forma orgânica. Dificuldades que prejudicaram as análises do estudo a respeito dessas variáveis. Diferenças no consumo entre clientes provenientes de campanha e clientes de busca orgânico impossibilitando a utilização de métricas de ROI nas análises. Para estudos futuros, deve-se incluir uma base de dados que contenha esses dados, fatores relevantes ao modelo de estudo.

A base de dados não contém os dados sociodemográficos de consumidores, como localização, gênero, idade etc, apresentando-os apenas com uma sequência de números e letras gerada aleatoriamente para cada usuário. Dados faltando que não possibilitam saber a região em que o consumidor acessou o site ou até mesmo comparar as compras de diferentes regiões do país, comparações do gênero que mais realizou compras ou qual é a idade média dos clientes que visitaram e realizaram compras no *e-commerce*. Falta de dados que impossibilitando qualquer análise em relação a esses fatores. A análise seria mais completa se a base tivesse tais dados, sendo uma limitação da própria base de dados. Tratando de ser uma sugestão de incluir esses dados em futuros estudos como uma maneira de superar os obstáculos encontrados neste estudo.

Outra limitação encontrada durante as análises foi não ser possível comparar de forma precisa os acessos durante a semana, ou seja, separar os acessos por dia da semana e nem realizar comparações do número de compras em relação ao dia da semana. Isto acontece porque entre os dias 15 e 31 existe apenas um domingo, uma segunda-feira, uma sexta-feira e um sábado em relação com os demais dias da semana, impossibilitando análises mais precisas sobre o período. Como sugestão

uma para estudos futuros, considerar a inclusão de no mínimo dois pares de dias para comparações e análises.

E, também, conforme citado nas implicações gerenciais, não foi possível entender a razão de porque os picos de visitas diários não representam os picos de vendas no dia. E, a partir disso, é interessante que seja feito um estudo futuro com a análise do comportamento dos consumidores de *e-commerce* nos diferentes períodos do dia.

## REFERÊNCIAS

- AKTER, Shahriar; WAMBA, Samuel Fosso. **Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets**, [S.L.], v. 26, n. 2, p. 173–194, mar. 2016
- ALBERTIN, Alberto Luiz. **Comércio Eletrônico: Modelo, aspectos e contribuições de sua aplicação**. Edição 6. São Paulo: Atlas, 2010.
- ALMEIDA, Ricardo. **Algoritmos & robôs: os novos públicos-alvo das marcas**. Clube de Autores, 2016.
- ANDRADE, Mairum Ceoldo. **Fundamentos de sistemas de informação**. [S.L.]: Universidade Estácio de Sá, 2014.
- BELLO-ORGAZ, Gema; JUNG, Jason J.; CAMACHO, David. **Social big data: Recent achievements and new challenges. Information Fusion**, Spain, v. 28, p. 45-59, mar. 2016.
- BLOCH, Michael; PIGNEUR, Yves; SEGEV, Arie. **On the road of electronic commerce: a business value framework, gaining competitive advantage and some research issues. CITM Working Paper**, [S.L.], mar. 1996.
- CAMERON, D. **Electronic commerce: the new business platform of the Internet**. Charleston: Computer Technology Research Corp., 1997
- CHEN, Hsinchun; CHIANG, Roger H. L.; STOREY, Veda C.. **Business Intelligence and Analytics: from Big Data to Big Impact. MIS Quarterly**, v. 36, n. 4, 2012.
- CHIAVENATO, Idalberto. **Introdução à Teoria Geral da Administração**. 8. ed. Rio de Janeiro: Elsevier, 2011.
- CINTRA, Flavia Cristina. **Marketing Digital: a era da tecnologia on-line. Investigação**, São Paulo, v. 10, n. 1, 2010.

DELALANA, Jessica. **O que são os parâmetros UTM do Google Analytics? Entenda a Importância e Como Começar a Utilizar.** Disponível em: <<https://automacaodevendas.com/o-que-sao-os-parametros-utm-do-google-analytics-entenda-a-importancia-como-comecar-a-utilizar/>>. Acesso em: 18 de set. 2017.

FARRIS, P. W. et al. **Métricas de marketing: O Guia Definitivo de Avaliação do Desempenho do Marketing.** 2ª ed. Porto Alegre: Bookman, 2012.

GABRIEL, Martha. **Marketing na era digital: conceitos, plataformas e estratégias.** 1. ed. São Paulo: Novatec, 2010.

GANDOMI, Amir; HAIDER, Murtaza. **Beyond the hype: Big data concepts, methods, and analytics.** *International Journal of Information Management*, v. 35, p. 137-144, 2015.

GOOGLE. **Campaign URL Builder.** Disponível em: <<https://ga-dev-tools.appspot.com/campaign-url-builder/>>. Acesso em: 27 de out. 2017.

GOOGLE. **Urchin Tracking Module (UTM).** Disponível em: <<https://support.google.com/urchin/answer/28307?hl=en>>. Acesso em: 27 de out. 2017.

GRECO, Adam. **Using UTM Campaign Parameters In Adobe Analytics.** Disponível em: <<https://community.digitalanalyticsassociation.org/blogs/adam-greco/2016/05/24/using-utm-campaign-parameters-in-adobe-analytics>>. Acesso em: 27 de out. 2017.

HAWKINS, del I.; MOTHERSBAUGH, David L.. **Consumer Behavior: Building Marketing Strategy.** 11. ed. New York: Mcgraw-Hill/Irwin, p. 778, 2010.

HOFACKER, Charles F.; MALTHOUSE, Edward Carl; SULTAN, Fareena. **Big Data and consumer behavior: imminent opportunities.** *Journal of Consumer Marketing*, v. 33, n. 2, p. 89-97, 2016.

IAB, Brasil. **Infográfico IAB brasil - Pesquisa Digital ADSPEND 2017**. Disponível em: <<http://iabbrasil.net/artigo/infografico-iab-brasil---pesquisa-digital-adspend-2017>>. Acesso em: 17 de set. 2017.

IAB, Brasil. **Números de Investimentos 2017**. Disponível em: <[http://iabbrasil.net/assets/upload/boas\\_praticas/1490794914.pdf](http://iabbrasil.net/assets/upload/boas_praticas/1490794914.pdf)>. Acesso em: 17 de set. 2017.

IAB, Brasil. **WEBSHOPPERS 35ª EDIÇÃO 2017**. Disponível em: <<http://iabbrasil.com.br/wp-content/uploads/2017/08/PESQUISA-EBIT-WEBSHOPPERS-35%C2%AA-EDI%C3%87%C3%83O-2017.pdf>>. Acesso em: 17 de set. 2017.

IBM. **What is big data?** Disponível em: <<https://www.ibm.com/analytics/us/en/big-data/>>. Acesso em: 17 de ago. 2017.

INPE. **Instituto Nacional de Pesquisas Espaciais**. Disponível em: <<http://sonda.ccst.inpe.br/infos/importcsv.html>>. Acesso em: 24 de out. 2017.

JANSEN, B.J; RESNICK, M. **An examination of searchers: perceptions of non-sponsored and sponsored links during ecommerce web searching**. Journal of the American Society for Information Science and Technology, Vol. 57, No. 14, p.1949–1961, 2006.

KENDZERSKI, Paulo Roberto. **Web marketing e comunicação digital: bem vindo ao mundo digital**. Porto Alegre: Wbi Brasil, 2009.

KOIRALA, P. **What is Big Data Analytics and its Application in E-Commerce?** Disponível em: <[www.venturecity.com](http://www.venturecity.com)>. Acesso em: 19 de out. de 2017.

KOTLER, Philip; KARTAJAYA, Hermawan; SETIAWAN, Iwan. **Marketing 4.0: moving from traditional to digital**. 1 ed. Nova Jersey: John Wiley & Sons, 2016.

KOTLER, Philip; KELLER, Kevin Lane. **Marketing Managment**. 14. ed. New Jersey: Pearson Prentice Hall, p. 657, 2012.

KOTLER, Philip. **Administração de Marketing**. 10ª Edição, 7ª reimpressão. Tradução Bazán Tecnologia e Lingüística; revisão técnica Arão Sapiro. São Paulo: Prentice Hall, 2000.

LANEY, Doug. **3D Data Management: Controlling Data Volume, Velocity, and Variety**. Disponível em: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>>. Acesso em: 15 de ago. 2017.

LIMEIRA, Tania M. Vidigal. **E-Marketing: O marketing na internet com casos brasileiros**. Edição 1. São Paulo: Saraiva, 2003.

MALHOTRA, Naresh K. **Pesquisa de Marketing: uma orientação aplicada**. 4ª ed. Porto Alegre. Bookman, 2006.

MALTHOUSE, Edward C.; LI, Hairong. **Opportunities for and Pitfalls of Using Big Data in Advertising Research**. *Journal of Advertising*, [S.L], v. 2, n. 46, p. 227-235, abr. 2017.

MAXIMIANO, Antonio César Amauri. **Teoria Geral da Administração: da Revolução Urbana à Revolução Digital**. São Paulo. Atlas, 2000.

MENDES, Marcos. **O comércio eletrônico no Brasil**. PUC-RS. Disponível em: <[http://www.cultura.ufpa.br/rcientifica/artigos\\_cientificos/ed\\_08/pdf/marcos\\_mendes3.pdf](http://www.cultura.ufpa.br/rcientifica/artigos_cientificos/ed_08/pdf/marcos_mendes3.pdf)>. Acesso em: 07 de nov. de 2017.

MENEZES, H. **Comércio eletrônico para pequenas empresas**. Florianópolis: Visual Books, 2003.

MITOVA, M. **Chapter 1: Your FREE UTM Builder Guide**. Disponível em: <<http://www.campaigntrackly.com/google-utm-builder-guide/>>. Acesso em: 27 de out. 2017.

PINHO, José Benedito. **Propaganda institucional: usos e funções da propaganda em relações públicas**. São Paulo: Summus Editorial, 1990.

PINHO, José Benedito. **Publicidade e vendas na internet: técnicas e estratégias**. São Paulo: Summus Editorial, 2004.

PRANGE, Eduardo. **Big Data: por que a tecnologia analítica é vital para o Marketing?** Disponível em: <<https://www.mundodomarketing.com.br/artigos/eduardo-prange/37559/big-data-por-que-a-tecnologia-analitica-e-vital-para-o-marketing.html>>. Acesso em: 25 de set. 2017.

ROSENBLOOM, Bert. **Canais de marketing: uma visão gerencial**. São Paulo: Atlas, 2015.

SAMARA, BEATRIZ S. e BARROS, JOSÉ CARLOS DE. **Pesquisa de Marketing: Conceitos e Metodologia**. 3ª Ed. São Paulo: Prentice Hall, 2002.

SEBRAE NACIONAL. **Conheça o panorama do e-commerce no Brasil**. Disponível em: <<https://www.sebrae.com.br/sites/PortalSebrae/artigos/conheca-o-panorama-do-e-commerce-no-brasil,66d975e0dc256510VgnVCM1000004c00210aRCRD>>. Acesso em 02 de nov. 2017.

Suporte Microsoft Office. **Como corrigir um erro #VALOR!**. Disponível em: <<https://support.office.com/pt-br/article/Como-corrigir-um-erro-VALOR-15e1b616-fbf2-4147-9c0b-0a11a20e409e>>. Acesso em 24 de out. 2017.

Suporte Microsoft Office. **COUN.SE (Função CONT.SE)**. Disponível em <<https://support.office.com/pt-br/article/CONT-SE-Fun%C3%A7%C3%A3o-CONT-SE-e0de10c6-f885-4e71-abb4-1f464816df34>>. Acesso em 24 de out. 2017.

Suporte Microsoft Office. **Filtrar valores exclusivos ou remover valores duplicados**. Disponível em: <<https://support.office.com/pt-br/article/Filtrar-valores-exclusivos-ou-remover-valores-duplicados-ccf664b0-81d6-449b-bbe1-8daaec1e83c2>>. Acesso em 24 de out. 2017.

Suporte Microsoft Office. **Função SE**. Disponível em <<https://support.office.com/pt-br/article/Fun%C3%A7%C3%A3o-SE-69aed7c9-4e8a-4755-a9bc-aa8bbff73be2>> . Acesso em 24 de out. 2017.

Suporte Microsoft Office. **Funções Lógicas**. Disponível em <<https://support.office.com/pt-br/article/Fun%C3%A7%C3%B5es-l%C3%B3gicas-refer%C3%A2ncia-e093c192-278b-43f6-8c3a-b6ce299931f5?ui=pt-BR&rs=pt-BR&ad=BR>>. Acesso em 24 de out. 2017.

Suporte Microsoft Office. **LOCALIZAR, LOCALIZAR (Funções LOCALIZAR, LOCALIZARB)**. Disponível em: <<https://support.office.com/pt-br/article/LOCALIZAR-LOCALIZARB-Fun%C3%A7%C3%B5es-LOCALIZAR-LOCALIZARB-9ab04538-0e55-4719-a72e-b6f54513b495>>. Acesso em: 24 de out. 2017.

TANG, Jian; ZHANG, Ping; WU, Philip Fei. **Categorizing consumer behavioral responses and artifact design features: The case of online advertising**. *Information Systems Frontiers*, [S.L], v. 17, n. 3, p. 513–532, 2014.

TORRES, Claudio. **A bíblia do marketing digital: tudo o que você queria saber sobre marketing e publicidade na Internet e não tinha a quem perguntar**. São Paulo: Novatec, 2009.

TURCHI, Sandra R. Artigo: **Entenda a mídia programática e como ela impacta o consumo na internet**. 2015. Disponível em:<<http://sandraturchi.com.br/entenda-a-midia-programatica-e-como-ela-impacta-o-consumo-na-internet/>>. Acesso em 18 de out. 2017.

TURCHI, Sandra R. **Estratégias de marketing digital e e-commerce**. São Paulo Atlas, 2012.

VERHOEF, Peter; KOOGHE, Edwin; WALK, Natasha. **Creating Value with Big Data Analytics: Making Smarter Marketing Decisions**. Londres: Routledge, 2016.

WATTS, Mack. **Programmatic Advertising: Shaping Consumer Behavior or Invading Consumer Privacy?**. The Ohio State University, Ohio, 2016.

ZIKOPOULOS, P. C. et al. **Understanding big data**. McGraw-Hill, 2012.

## APÊNDICE A – ANÚNCIOS MODIFICADOS UTILIZADOS PELA EMPRESA X NO PERÍODO DE ANÁLISE

**EMPRESA X**

**BOSS**  
HUGO BOSS

Perfumes  
A PARTIR DE  
**R\$199**

**EMPRESA X**

**Dior**

Perfumes  
A PARTIR DE  
**R\$269**

**CONFIRA ▶**

**EMPRESA X**

**BOSS**  
HUGO BOSS

Perfumes  
A PARTIR DE  
**R\$199**

**CONFIRA ▶**

**EMPRESA X**

**Dior**

Perfumes  
A PARTIR DE  
**R\$269**

**CONFIRA ▶**

**EMPRESA X**

**BOSS**  
HUGO BOSS

Perfumes  
A PARTIR DE  
**R\$199**

**EMPRESA X**

**Calvin Klein**

Perfumes  
A PARTIR DE  
**R\$159**

**CONFIRA ▶**

**EMPRESA X**

**BOSS**  
HUGO BOSS

Perfumes  
A PARTIR DE  
**R\$199**

**CONFIRA ▶**

**EMPRESA X**

**Calvin Klein**

Perfumes  
A PARTIR DE  
**R\$159**

**CONFIRA ▶**

**EMPRESA X**

**BOSS**  
HUGO BOSS

Perfumes  
A PARTIR DE  
**R\$199**

**CONFIRA ▶**

**EMPRESA X**

**Calvin Klein**

Perfumes  
A PARTIR DE  
**R\$159**

**CONFIRA ▶**