



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS QUIXADÁ**  
**CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**ANTÔNIO ANDSON DA SILVA**

**IDENTIFICAÇÃO DO PERFIL DO CONSUMIDOR: UM ESTUDO DE CASO NA LOJA  
AND STORE**

**QUIXADÁ**  
**2023**

ANTÔNIO ANDSON DA SILVA

IDENTIFICAÇÃO DO PERFIL DO CONSUMIDOR: UM ESTUDO DE CASO NA LOJA  
AND STORE

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação do Campus Quixadá da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Ciência da Computação.

Orientador: Prof. Dr. Lucas Ismaily Bezerra Freitas.

Coorientador: Prof. Dr. Antônio Joel Ramiro de Castro.

QUIXADÁ

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S578i Silva, Antônio Andson da Silva.  
Identificação do perfil do consumidor: um estudo de caso na loja And Store / Antônio Andson da Silva  
Silva. – 2023.  
37 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Quixadá,  
Curso de Ciência da Computação, Quixadá, 2023.

Orientação: Prof. Dr. Lucas Ismail Bezerra Freitas.

Coorientação: Prof. Dr. Antônio Joel Ramiro de Castro.

1. Comportamento do consumidor. 2. Agrupamento. 3. K-means. 4. Weka. I. Título.

CDD 004

---

ANTÔNIO ANDSON DA SILVA

IDENTIFICAÇÃO DO PERFIL DO CONSUMIDOR: UM ESTUDO DE CASO NA LOJA  
AND STORE

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Ciência da Computação  
do Campus Quixadá da Universidade Federal  
do Ceará, como requisito parcial à obtenção do  
grau de bacharel em Ciência da Computação.

Aprovada em: \_\_\_/\_\_\_/\_\_\_\_\_

BANCA EXAMINADORA

---

Prof. Dr. Lucas Ismaily Bezerra Freitas (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Antônio Joel Ramiro de  
Castro (Coorientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Me. Roberto Cabral Rabêlo Filho  
Universidade Federal do Ceará (UFC)

---

Prof. Me. Carlos Igor Ramos Bandeira  
Universidade Federal do Ceará (UFC)

Por você, Iza.

## **AGRADECIMENTOS**

Agradeço meus avós Izabel Tomaz Melo e Francisco Raimundo de Aquino Melo por todo apoio, carinho e amor que fez com que eu me tornasse quem sou hoje, muito obrigado.

As minhas irmãs, Amanda e Beatriz que sempre estiveram do meu lado, motivando e me inspirando ser uma pessoa ainda melhor.

Aos meus cachorros, Tobias e And Maria, que chegaram para encher meu coração de amor e alegria.

Aos professores da Universidade Federal do Ceará, Campus Quixadá, que contribuíram para minha formação, em especial aos professores Joel e Ismaily, muito obrigado por todo aprendizado repassado.

“O sonho é que leva a gente para frente. Se a gente for seguir a razão, fica aquietado, acomodado.”

(Ariano Suassuna)

## RESUMO

Este estudo investigou o comportamento do consumidor na loja And Store, aprofundando-se na compreensão do seu público-alvo, com o objetivo de identificar os perfis dos consumidores. Adotou-se uma abordagem que combinou a coleta e o tratamento de dados com a aplicação da técnica estatística K-means, a fim de oferecer uma visão detalhada e segmentada dos comportamentos de compra dos consumidores. A técnica estatística K-means foi aplicada no Weka, uma plataforma de análise de dados com interface simples e intuitiva. Os resultados apresentados, evidenciaram a identificação de 9 agrupamentos distintos, cada um caracterizando um perfil específico de consumidor. A conclusão apontou uma base sólida para futuras estratégias de negócios.

**Palavras-chave:** comportamento do consumidor; agrupamento; k-means; weka.



## **ABSTRACT**

This study investigated consumer behavior at the And Store store, giving a deeper understanding its target audience, with the aim of identifying consumer profiles. An approach was adopted that combined data collection and processing with the application of the K-means statistical technique, in order to offer a detailed and segmented view of consumers' purchasing behaviors. The K-means technique was applied using Weka, a data analysis platform with a simple and intuitive interface. The results presented showed the identification of 9 distinct clusters, each one characterizing a specific consumer profile. The conclusion showed a solid and promise business strategies for the future.

**Keywords:** consumer behavior; clustering; k-means; weka.

## LISTA DE FIGURAS

Figura 1 – Conjunto de dados com 3 agrupamentos . . . . .	16
Figura 2 – Atributos são apresentados no Weka . . . . .	20
Figura 3 – Iniciando análise de agrupamentos no Weka . . . . .	21
Figura 4 – Aplicando o K-means . . . . .	21

## LISTA DE TABELAS

Tabela 1 – <i>Clustered Instances 8%</i> . . . . .	23
Tabela 2 – <i>Clustered Instances 13%</i> . . . . .	24
Tabela 3 – <i>Clustered Instances 13%</i> . . . . .	25
Tabela 4 – <i>Clustered Instances 10%</i> . . . . .	25
Tabela 5 – <i>Clustered Instances 9%</i> . . . . .	26
Tabela 6 – <i>Clustered Instances 7%</i> . . . . .	26
Tabela 7 – <i>Clustered Instances 19%</i> . . . . .	26
Tabela 8 – <i>Clustered Instances 6%</i> . . . . .	27
Tabela 9 – <i>Clustered Instances 15%</i> . . . . .	27

## LISTA DE CÓDIGOS-FONTE

Código-fonte 1 – Atributos ARFF . . . . .	32
Código-fonte 2 – Alguns dados ARFF . . . . .	33
Código-fonte 3 – Tratando dados . . . . .	34

## LISTA DE ABREVIATURAS E SIGLAS

ARFF *Attribute-Relation File Format*

WEKA *Waikato Enviroment for Knowledge Analysis*

## **LISTA DE SÍMBOLOS**

UFC            Universidade Federal do Ceará

## SUMÁRIO

1	<b>INTRODUÇÃO</b>	14
2	<b>FUNDAMENTAÇÃO TEÓRICA</b>	15
2.1	<b>Loja And Store</b>	15
2.2	<b>Estatística</b>	15
2.2.1	<i>Agrupamento (Clustering)</i>	15
2.2.2	<i>K-means</i>	16
2.3	<b>Weka</b>	17
3	<b>TRABALHOS RELACIONADOS</b>	18
4	<b>COLETA E ANÁLISE DE DADOS</b>	19
4.1	<b>Coleta de Dados</b>	19
4.2	<b>Tratamento dos Dados</b>	19
4.3	<b>Aplicação da Técnica Estatística</b>	20
5	<b>RESULTADOS E DISCUSSÃO</b>	23
6	<b>CONCLUSÃO E TRABALHOS FUTUROS</b>	29
	<b>REFERÊNCIAS</b>	30
	<b>APÊNDICE A – CÓDIGO-FONTE ATRIBUTOS</b>	32
	<b>APÊNDICE B – CÓDIGO-FONTE DADOS</b>	33
	<b>APÊNDICE C – CÓDIGO-FONTE TRATAMENTO</b>	34
	<b>ANEXO A – QUESTIONÁRIO UTILIZADO</b>	35

## 1 INTRODUÇÃO

Em um cenário comercial altamente competitivo e em constante evolução, entender o comportamento e as preferências do consumidor é de suma relevância para qualquer empresa. A compreensão aprofundada do consumidor possibilita a criação de estratégias mais eficazes, que podem envolver desde a personalização de produtos e serviços até a segmentação de mercado mais precisa (SCHIFFMAN *et al.*, 2010).

As preferências individuais das pessoas variam significativamente. No trabalho de (KOTLER, 2015) expõe que os especialistas em marketing empregam a estratégia da segmentação para examinar as variações nas preferências individuais. Isso lhes permite identificar os grupos de consumidores que apresentam as melhores oportunidades. Para cada segmento de mercado identificado, as empresas elaboram estratégias de marketing direcionadas, com o objetivo de criar uma percepção distinta na mente dos consumidores, realçando os principais benefícios de seus produtos ou serviços.

Este estudo busca inserir-se nesse contexto desafiador ao propor uma abordagem estatística para identificar os perfis dos consumidores da loja And Store. Assim que este trabalho conseguir detalhar e segmentar grupos após a análise de dados, o objetivo estará concluído.

A loja And Store participou desta trabalho reconhecendo sua capacidade de fornecer *insights* valiosos sobre as preferências e padrões de compra de seus clientes.

Neste contexto, o presente trabalho está estruturado da seguinte forma: no Capítulo 2, é apresentada a fundamentação teórica; no Capítulo 3, são apresentados os trabalhos relacionados; no Capítulo 4, são apresentadas a coleta e análise de dados; no Capítulo 5, são mostrados os resultados; por fim, no Capítulo 6, temos a conclusão e sugestões para trabalhos futuros.



## 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, introduzimos os conceitos necessários ao longo do texto. Ele está dividido da seguinte forma: na Seção 2.1, apresentamos a Loja And Store, empresa que será utilizada neste estudo. Em seguida, na Seção 2.2, definimos alguns conceitos de Estatística. Finalmente, na Seção 2.3, apresentamos o Weka, a ferramenta utilizada para analisar os resultados gerados pela técnica estatística empregada neste estudo.

### 2.1 Loja And Store

A And Store é uma loja especializada em camisas esportivas, a loja surgiu em 2019 e oferece uma variedade de produtos de diferentes clubes brasileiros, internacionais e seleções. A loja possui um ponto comercial em Quixadá-CE, proporcionando aos consumidores um ambiente de compra prático e confiável. A loja conta ainda com a opção de compra online, que envia seus produtos para todo o Brasil (SILVA, 2023).

### 2.2 Estatística

A estatística desempenha um papel essencial na interpretação e análise de dados, fornecendo ferramentas valiosas para compreender padrões, realizar inferências e tomar decisões (KOTZ *et al.*, 2004).

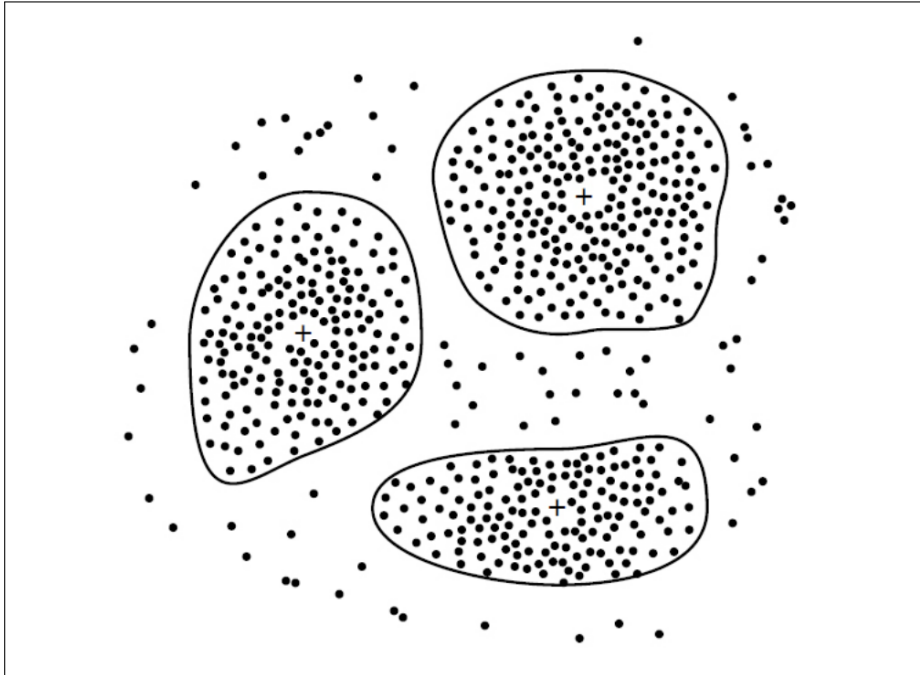
#### 2.2.1 Agrupamento (*Clustering*)

A atividade de agrupamento tem como objetivo reconhecer e aproximar itens semelhantes.

A tarefa de agrupamento visa identificar e aproximar os registros similares. Um agrupamento (ou *cluster*) é uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. (CAMILO; SILVA, 2009, p. 9).

As aplicações das tarefas de agrupamento são as mais variadas possíveis: pesquisa de mercado, reconhecimento de padrões, processamento de imagens, análise de dados, segmentação de mercado, taxonomia de plantas e animais, pesquisas geográficas, classificação de documentos da Web, detecção de comportamentos atípicos (fraudes), entre outras. (CAMILO; SILVA, 2009, p. 10).

Figura 1 – Conjunto de dados com 3 agrupamentos



Fonte: adaptado pelo autor (2023).

### 2.2.2 *K-means*

Dado que o objetivo deste trabalho é identificar agrupamentos de consumidores com características semelhantes, tornou-se essencial utilizar uma técnica de agrupamento que revele semelhanças entre padrões e agrupe padrões semelhantes em categorias ou *clusters*.

De acordo com Fonseca e Beltrame (2010), o algoritmo de clusterização K-means é amplamente empregado na tarefa de agrupar dados, destacando-se não apenas por sua eficácia, mas também por sua alta funcionalidade. O algoritmo K-means é preferido por sua facilidade de implementação e sua complexidade computacional  $\theta(n)$ , onde  $n$  representa o tamanho da entrada.

K-means é uma técnica que usa o algoritmo de agrupamento de dados por K-médias (*K-means clustering*). O objetivo deste algoritmo é encontrar a melhor divisão de  $P$  dados em  $K$  grupos  $C_i, i = 1, \dots, K$ , de maneira que a distância total entre os dados de um grupo e o seu respectivo centro, somada por todos os grupos, seja minimizada. (PIMENTEL *et al.*, 2003, p. 497).

O algoritmo inicialmente atribui aleatoriamente os  $P$  pontos de dados a  $K$  agrupamentos e calcula as médias dos vetores em cada agrupamento. Posteriormente, cada ponto de dados é realocado para o agrupamento correspondente ao vetor médio do qual está mais próximo. Com este novo arranjo dos pontos de dados em  $K$  agrupamentos, novos vetores médios são recalculados. O processo de realocação de pontos de dados para novos agrupamentos, cujos

vetores médios estão mais próximos deles, continua até que seja alcançado um estado onde todos os pontos de dados já pertencem aos agrupamentos de seus vetores médios mais próximos (SILVA *et al.*, 2001).

### 2.3 Weka

A *Waikato Environment for Knowledge Analysis* (WEKA) é uma plataforma de software de código aberto bastante utilizada em aprendizado de máquina e mineração de dados. Seu nome é uma referência à Universidade de Waikato, na Nova Zelândia, onde o projeto teve origem. Essa plataforma oferece um conjunto de bibliotecas e ferramentas para análise de dados, pré-processamento, classificação, regressão, clusterização, associação, seleção de atributos e visualização (HALL *et al.*, 2009).

Weka é formado por um conjunto de pacotes, que são: *attribute selection* (seleção de atributos), *classifiers* (implementação de algoritmos de aprendizagem supervisionados), *clustering* (implementação de algoritmos de aprendizagem não supervisionados), *association rules* (generalização de regras de associação usando taxonomias), *filters* (seleção de instância de uma base) e *estimators* (estimadores). (SMIDERLE; OLIVEIRA, 2005, p. 3).

Todas as aplicações da plataforma podem ser acessadas através de uma interface simples e intuitiva. Os algoritmos podem ser aplicados de forma direta aos conjuntos de dados, eliminando a necessidade de escrever código manualmente. Além disso, oferece extensa documentação on-line, sustenta uma comunidade de entusiastas dedicados a manter o projeto atualizado e em constante evolução (WITTEN *et al.*, 2005).

### 3 TRABALHOS RELACIONADOS

A seguir são apresentados os trabalhos que auxiliaram na construção deste estudo.

Os autores Heinzen e Martins (2018) realizam um estudo utilizando análise de agrupamento (clusterização) para identificar o perfil de consumidores em ambientes alimentícios. Semelhante a este trabalho, buscou compreender o comportamento e as preferências dos consumidores, além de ter utilizado o algoritmo de agrupamento K-means.

O trabalho de Santana e Pontes (2020) foi uma referência inspiradora na elaboração deste trabalho. Santana e Pontes (2020) propõem o desenvolvimento de um sistema de recomendação de produtos com base nos padrões de compra dos clientes, por meio de modelagem em uma empresa distribuidora. Nessa metodologia, utiliza-se o algoritmo de clusterização K-means para realizar agrupamentos estratégicos dos clientes.

A análise de agrupamentos é utilizada em diversas áreas, como no trabalho de Carvalho *et al.* (2022), que tem como objetivo analisar o perfil sociodemográfico, laboral e a intensidade das cargas de trabalho de trabalhadores da enfermagem de hospitais universitários. Esse estudo resultou em implicações práticas no cotidiano desses profissionais. Realizou-se a análise de agrupamento e, ao identificar os perfis, tornou-se possível o planejamento de intervenções direcionadas a minimizar as cargas de trabalho de acordo com as particularidades de cada grupo.

Pires *et al.* (2020) conduziram um estudo investigando os perfis de óbitos infantis em Santa Maria/RS no período de 2000 a 2017, empregando regras de associação e redes neurais artificiais. O intuito era detectar padrões nas causas desses óbitos. Assim como Pires *et al.* (2020), iremos usar a ferramenta Weka para identificar os perfis. A partir desse trabalho, vale destacar a ferramenta Weka, visto que ela pode ser utilizada em diversos trabalhos, gerenciando vários tipos de algoritmos e métodos.

## 4 COLETA E ANÁLISE DE DADOS

Os procedimentos adotados neste estudo são divididos em três etapas: (i) coleta de dados, (ii) tratamento dos dados e (iii) aplicação da técnica estatística.

### 4.1 Coleta de Dados

A coleta de dados desempenha um papel central em qualquer pesquisa, e o questionário emerge como uma ferramenta valiosa para explorar percepções e opiniões (CRESWELL; POTH, 2016).

Os dados coletados para o estudo foram obtidos por meio de um questionário aplicado pela loja And Store aos seus clientes. Ressaltamos que a loja realizou a coleta de dados exclusivamente com o objetivo de compreender melhor seus consumidores. As perguntas do questionário foram elaboradas pelo time de *marketing* da loja. O questionário esteve disponível para os clientes do dia 1 de setembro de 2023 até o dia 15 de setembro de 2023.

A loja And Store comunicou que o questionário foi divulgado em suas redes sociais e por *e-mails*. Foram obtidas um total de 125 respostas, totalizando 649 compras. No questionário (ver Anexo A), temos as perguntas que foram consideradas no estudo. Os dados recebidos podem ser visualizados em forma de planilha ou gráficos.

### 4.2 Tratamento dos Dados

Após o recebimento dos dados, inicia-se a etapa de tratamento. Cada linha da planilha representa a resposta de um cliente, por exemplo: "Masculino", "Entre 20 e 29 anos", "Fortaleza-CE", "Entre 2 e 4 salários mínimos", "Mais de 5 camisas", "Sim, apenas em promoção", "Fortaleza, Brasil", "Ambos", "Outro", "Comprei na loja física", "5". Essa resposta precisa ser configurada para o padrão de entrada da ferramenta Weka. O *Attribute-Relation File Format* (ARFF) é o padrão de formato de arquivo utilizado pelo Weka, sendo um formato de arquivo de texto que descreve conjuntos de dados, especificando os atributos e suas relações.

A partir das perguntas do questionário, surgem os atributos (ver Apêndice A). Na pergunta: "Você comprou apenas quando os produtos estavam em promoção:", foi criado o atributo "promocao" com os parâmetros "sim", "não" e "ambos". Para as respostas, instâncias são geradas (ver Apêndice B) fornecendo os valores específicos.

Cada instância corresponde a um atributo específico; assim, no exemplo da resposta

acima, é necessário lidar com o caso "Fortaleza, Brasil", visto que o atributo "time" só pode ser atribuído a um parâmetro. Nessa situação, duas instâncias são iniciadas (ver Apêndice C), repetindo os demais dados. Perguntas com a possibilidade de respostas variadas geralmente precisam ser tratadas.

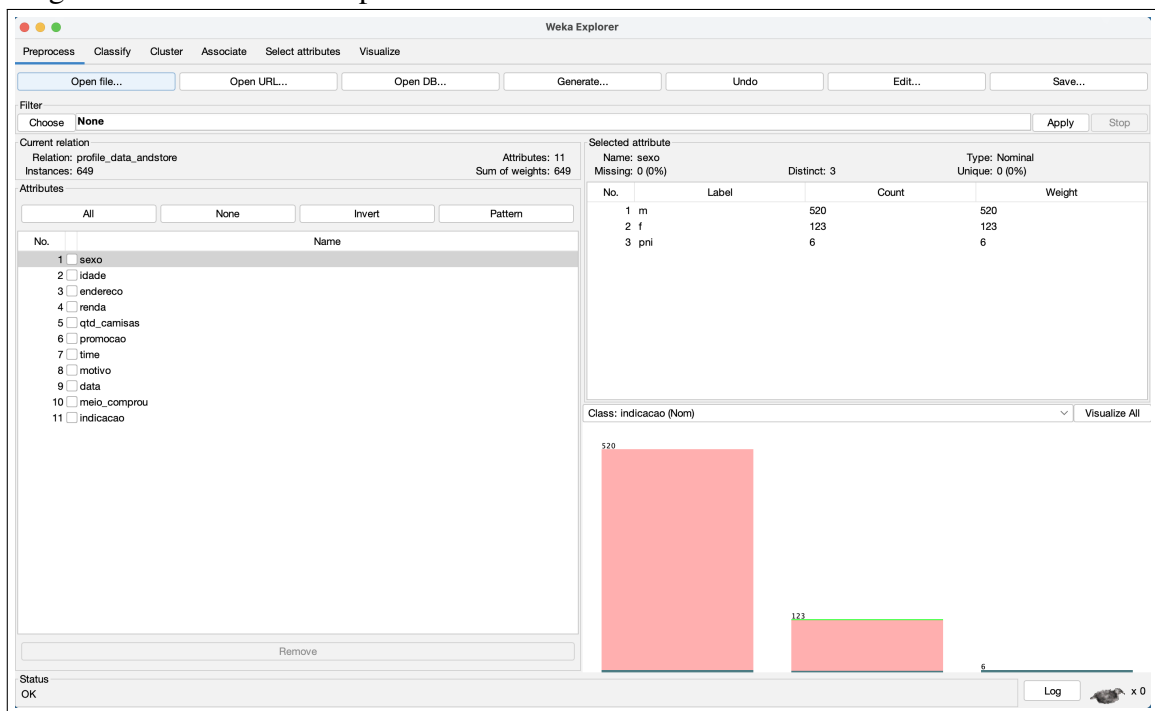
Por último, tratamos da segmentação das vendas, distinguindo entre vendas locais e não locais, utilizando o ponto comercial em Quixadá-CE como ponto de referência. A classificação foi realizada com base nas informações de endereço fornecidas pelos clientes.

### 4.3 Aplicação da Técnica Estatística

Nesta etapa, empregaremos a técnica estatística no Weka para analisar os dados. Essa fase é essencial para a compreensão dos dados e a identificação dos perfis.

Antes da aplicação da técnica, foi realizado o tratamento dos dados, como mencionado na seção anterior, resultando na criação do arquivo ARFF. Com o ambiente Weka devidamente configurado, o arquivo ARFF contendo os dados relevantes foi carregado. Os atributos são mostrados no Weka conforme a Figura 2.

Figura 2 – Atributos são apresentados no Weka

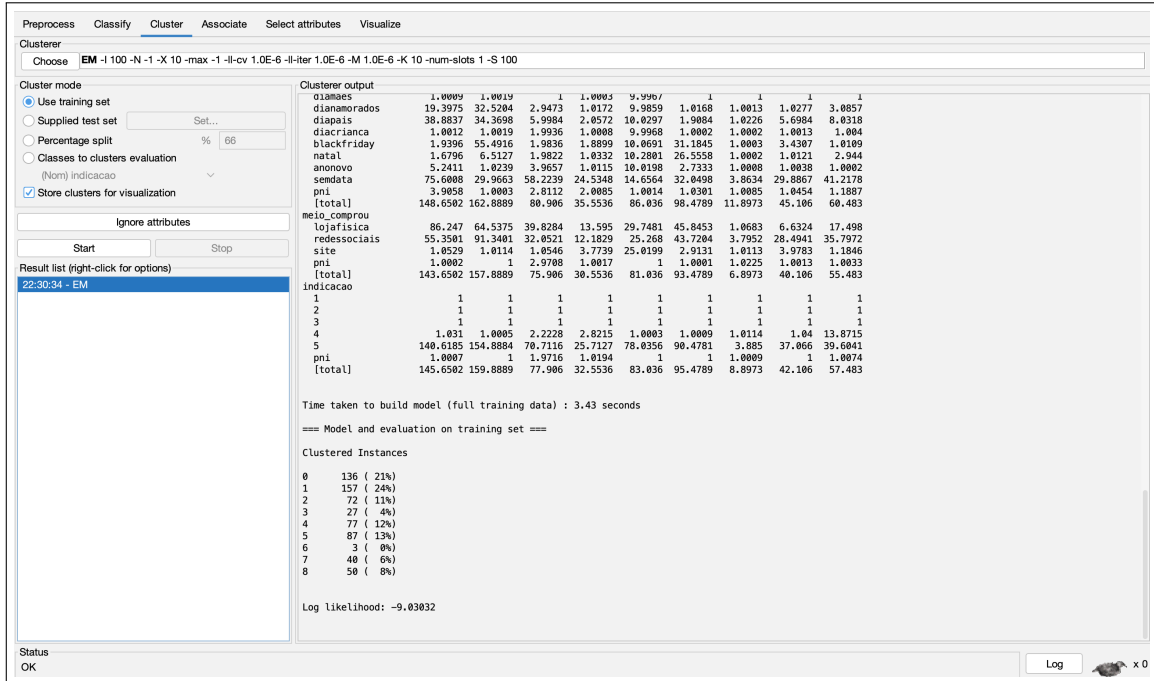


Fonte: elaborado pelo autor (2023).

Na sequência, a próxima ação consiste em aplicar a técnica estatística, selecionando a opção "Cluster" no Weka e iniciar a análise para determinar quantos clusters foram identificados.

Na Figura 3, é possível visualizar o resultado alcançado, revelando a identificação de 9 clusters distintos.

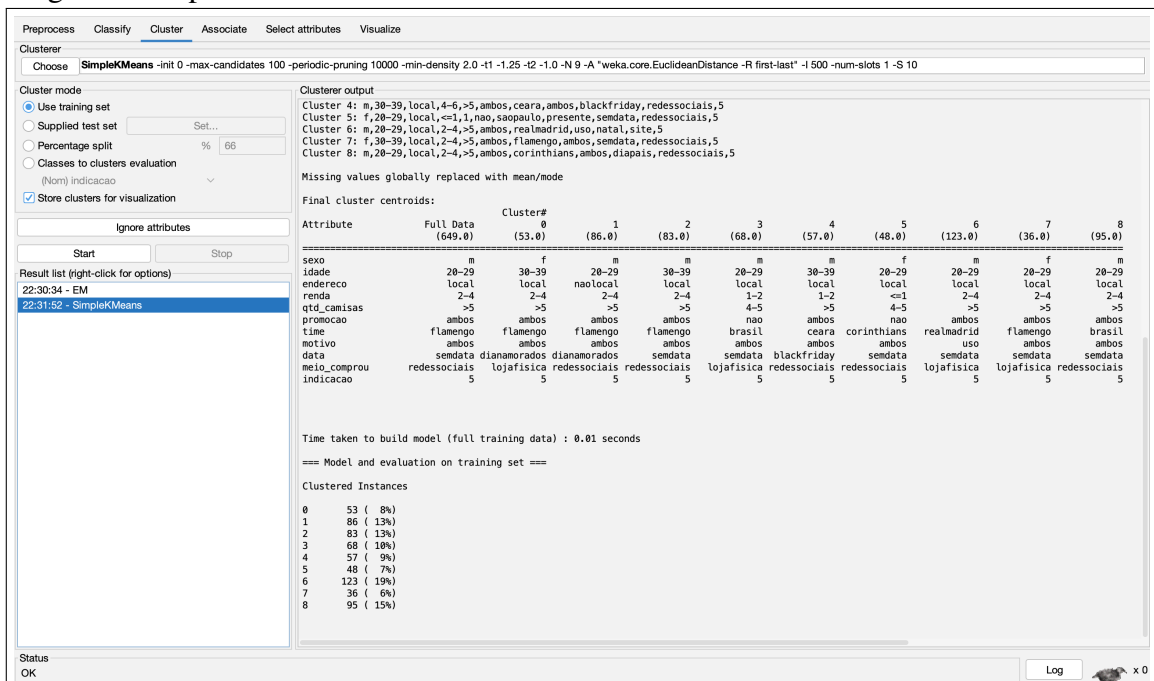
Figura 3 – Iniciando análise de agrupamentos no Weka



Fonte: elaborado pelo autor (2023).

Na fase final da aplicação, realiza-se a seleção do K-means, conforme evidenciado na Figura 4.

Figura 4 – Aplicando o K-means



Fonte: elaborado pelo autor (2023).

Essa etapa representa um ponto-chave no processo, indicando a escolha do algoritmo específico para análise de agrupamento. A figura fornece uma representação visual que sintetiza o momento crucial em que o método selecionado é aplicado, iniciando assim a segmentação dos dados. Este passo é fundamental para revelar padrões, desempenhando um papel crucial na compreensão mais profunda da natureza dos dados analisados.



## 5 RESULTADOS E DISCUSSÃO

A fase de resultados deste trabalho, dedicou-se à realização de uma análise de agrupamento por meio do algoritmo K-means, utilizando a plataforma Weka. O propósito principal era descobrir padrões nos dados coletados, e os resultados revelaram uma segmentação significativa, evidenciando agrupamentos distintos.

Após a preparação dos dados, a aplicação do K-means revelou a formação de agrupamentos identificáveis. A Figura 4 visualiza os resultados, proporcionando uma representação dos agrupamentos identificados. A distribuição dos dados entre os agrupamentos destaca a natureza distintiva de cada agrupamento.

Ao realizar uma análise de agrupamento, a métrica '*Clustered Instances*' mostra quantas instâncias do conjunto de dados total foram atribuídas a esses agrupamentos. Essa métrica é um indicador da eficácia do algoritmo em agrupar as observações e da abrangência da segmentação nos dados.

Vamos examinar os agrupamentos identificados:

Agrupamento 1: Este agrupamento (Tabela 1) retrata uma cliente do sexo feminino, que demonstra uma propensão a adquirir camisas do Flamengo como presente no Dia dos Namorados. Ela está na faixa etária de 30 a 39 anos, possui uma renda mensal entre 2 e 4 salários mínimos. Reside em Quixadá ou nas proximidades da região, sendo uma cliente estabelecida da loja, como indicado pelo histórico de múltiplas compras. Além disso, observa-se que ela adquiriu camisas para si mesma, com ou sem promoção.

Tabela 1 – *Clustered Instances* 8%

Atributo	Cluster
sexo	f
idade	30-39
endereco	local
renda	2-4
qtd_camisas	>5
promocao	ambos
time	flamengo
motivo	ambos
data	dianamorados
meio_comprou	lojafisica
indicacao	5

Fonte: elaborado pelo autor (2023).

Agrupamento 2: Este agrupamento (Tabela 2) evidencia um cliente do sexo mascu-

lino que, no Dia dos Namorados, tende a adquirir camisas do Flamengo como presente. Este cliente encontra-se na faixa etária de 20 a 29 anos, com uma renda mensal entre 2 e 4 salários mínimos. Reside a uma distância considerável do ponto físico da loja, no entanto, é um cliente consolidado, como indicado por seu histórico de compras significativo. Ele adquiriu camisas para si mesmo, com ou sem promoção. Devido à sua localização distante, as transações são realizadas principalmente via redes sociais.

Tabela 2 – *Clustered Instances 13%*

Atributo	Cluster
sexo	m
idade	20-29
endereco	naolocal
renda	2-4
qtd_camisas	>5
promocao	ambos
time	flamengo
motivo	ambos
data	dianamorados
meio_comprou	redessociais
indicacao	5

Fonte: elaborado pelo autor (2023).

Agrupamento 3: Este agrupamento (Tabela 3) revela um cliente do sexo masculino que, sem uma data específica, demonstra uma propensão a adquirir camisas do Flamengo, seja para si mesmo ou para presentear alguém. Este cliente está na faixa etária de 30 a 39 anos, com uma renda mensal entre 2 e 4 salários mínimos. Ele reside em Quixadá ou nas proximidades, sendo um cliente estabelecido da loja, conforme evidenciado por seu histórico substancial de compras, tanto com quanto sem promoção. Além disso, destaca-se que as transações são predominantemente conduzidas via redes sociais.

Agrupamento 4: Este agrupamento (Tabela 4) apresenta um cliente do sexo masculino que, sem uma data específica, revela interesse em adquirir camisas do Brasil, seja para si mesmo ou para presentear alguém. Este cliente está na faixa etária de 20 a 29 anos, com uma renda mensal entre 1 e 2 salários mínimos. Reside em Quixadá ou nas proximidades, e realizou entre 4 e 5 compras de camisas, sem promoção, na loja física. Esse padrão de comportamento sugere uma afinidade específica por camisas do Brasil e uma preferência por realizar compras presenciais na loja física.

Agrupamento 5: Este agrupamento (Tabela 5) destaca um cliente do sexo masculino que, durante a *Black Friday*, manifesta interesse em adquirir camisas do Ceará, seja para si

Tabela 3 – *Clustered Instances 13%*

Atributo	Cluster
sexo	m
idade	30-39
endereço	local
renda	2-4
qtd_camisas	>5
promocao	ambos
time	flamengo
motivo	ambos
data	semdata
meio_comprou	redessociais
indicacao	5

Fonte: elaborado pelo autor (2023).

Tabela 4 – *Clustered Instances 10%*

Atributo	Cluster
sexo	m
idade	20-29
endereço	local
renda	1-2
qtd_camisas	4-5
promocao	nao
time	brasil
motivo	ambos
data	semdata
meio_comprou	lojafisica
indicacao	5

Fonte: elaborado pelo autor (2023).

mesmo ou para presentear alguém. Este cliente está na faixa etária de 30 a 39 anos, com uma renda mensal entre 1 e 2 salários mínimos. Reside em Quixadá ou nas proximidades, e realizou mais de 5 compras de camisas, tanto com quanto sem promoção. Notavelmente, suas transações são predominantemente conduzidas nas redes sociais, indicando uma preferência por esse canal para suas compras durante períodos promocionais, como a *Black Friday*.

Agrupamento 6: Neste agrupamento (Tabela 6), apresenta uma cliente do sexo feminino, com idade entre 20 e 29 anos e residente em Quixadá ou proximidades, com uma renda de até 1 salário mínimo, já adquiriu entre 4 e 5 camisas do Corinthians, sem aproveitar valores promocionais. Suas transações, realizadas predominantemente nas redes sociais, não seguem uma data específica, indicando uma preferência por compras ao longo do tempo. Suas aquisições abrangem tanto compras para si mesma quanto para presentear outras pessoas.

Agrupamento 7: Neste agrupamento (Tabela 7), retrata um cliente do sexo masculino, na faixa etária de 20 a 29 anos e residente em Quixadá ou proximidades, com uma renda de 2 a 4

Tabela 5 – *Clustered Instances 9%*

Atributo	Cluster
sexo	m
idade	30-39
endereço	local
renda	1-2
qtd_camisas	>5
promocao	ambos
time	ceara
motivo	ambos
data	blackfriday
meio_comprou	redessociais
indicacao	5

Fonte: elaborado pelo autor (2023).

Tabela 6 – *Clustered Instances 7%*

Atributo	Cluster
sexo	f
idade	20-29
endereço	local
renda	<=1
qtd_camisas	4-5
promocao	nao
time	corinthians
motivo	ambos
data	semdata
meio_comprou	redessociais
indicacao	5

Fonte: elaborado pelo autor (2023).

salários mínimos, já adquiriu várias camisas, com destaque para as do Real Madrid, destinadas ao uso pessoal. Suas compras envolvem tanto a aquisição com promoções quanto sem promoções específicas, e não seguem uma data específica, indicando uma preferência por compras ao longo do tempo. Importante destacar que suas compras são realizadas na loja física.

Tabela 7 – *Clustered Instances 19%*

Atributo	Cluster
sexo	m
idade	20-29
endereço	local
renda	2-4
qtd_camisas	>5
promocao	ambos
time	realmadrid
motivo	uso
data	semdata
meio_comprou	lojafisica
indicacao	5

Fonte: elaborado pelo autor (2023).

Agrupamento 8: Neste agrupamento (Tabela 8), revela uma cliente do sexo feminino, na faixa etária de 20 a 29 anos e residente em Quixadá ou proximidades, com uma renda de 2 a 4 salários mínimos, já adquiriu mais de 5 camisas, com ênfase nas do Flamengo. Essas aquisições incluem camisas tanto para uso pessoal quanto para presentear outras pessoas. Compras tanto com promoções quanto sem promoções específicas, e não há uma data específica associada a essas transações. Preferência por realizar compras presenciais na loja física.

Tabela 8 – *Clustered Instances 6%*

Atributo	Cluster
sexo	f
idade	20-29
endereco	local
renda	2-4
qtd_camisas	>5
promocao	ambos
time	flamengo
motivo	ambos
data	semdata
meio_comprou	lojafisica
indicacao	5

Fonte: elaborado pelo autor (2023).

Agrupamento 9: Neste agrupamento (Tabela 9), destaca um cliente do sexo masculino, com idade entre 20 e 29 anos e residente em Quixadá ou nas proximidades, possui uma renda de 2 a 4 salários mínimos. Já adquiriu mais de 5 camisas, com destaque para as do Brasil. Estas compras englobam camisas tanto para uso pessoal quanto para presentear outras pessoas. O cliente realizou transações tanto com promoções quanto sem promoções específicas, sem uma data específica associada. Suas compras são efetuadas por meio das redes sociais.

Tabela 9 – *Clustered Instances 15%*

Atributo	Cluster
sexo	m
idade	20-29
endereco	local
renda	2-4
qtd_camisas	>5
promocao	ambos
time	brasil
motivo	ambos
data	semdata
meio_comprou	redessociais
indicacao	5

Fonte: elaborado pelo autor (2023).

Ao analisar os agrupamentos identificados, é possível notar a diversidade nos comportamentos de compra, representando uma riqueza de informações para estratégias de marketing e personalização de ofertas. Este processo de agrupamento revela-se uma ferramenta poderosa para desvendar padrões complexos nos dados, contribuindo para uma compreensão mais profunda do público-alvo e auxiliando na tomada de decisões embasadas em dados para melhor atender às necessidades específicas de cada segmento.

## 6 CONCLUSÃO E TRABALHOS FUTUROS

O presente trabalho conduziu uma análise detalhada do processo de coleta e manipulação de dados, empregando a técnica estatística conhecida como K-means, no cenário específico da loja And Store. Cada etapa, desde a coleta inicial de dados até o tratamento e aplicação da técnica estatística, proporcionou *insights* valiosos sobre os comportamentos dos consumidores. Essa metodologia desempenhou um papel crucial no aprofundamento da compreensão do público-alvo da loja, proporcionando uma análise abrangente e categórica.

A aplicação da técnica estatística no Weka, especificamente utilizando o algoritmo K-means, foi executada com sucesso. Os resultados apontaram para a identificação de 9 clusters distintos, cada um caracterizando um perfil específico de consumidor. Essa segmentação forneceu uma visão detalhada dos diferentes comportamentos de compra, desde preferências de produtos até canais de compra. As nuances nos padrões de compra, fornecem percepções importantes para estratégias de marketing e personalização nas ofertas.

Os resultados apresentados não apenas oferecem uma compreensão aguçada do cenário atual dos consumidores da And Store, mas também estabelecem uma base sólida para futuras estratégias de negócios e aprimoramento contínuo.

Este estudo, embora conclusivo em sua fase atual, sugere oportunidades para investigações futuras, como a aplicação de outras técnicas estatísticas, expansão do escopo da amostra e a consideração de variáveis adicionais para uma análise ainda mais aprofundada do comportamento do consumidor.

## REFERÊNCIAS

- CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. **Universidade Federal de Goiás (UFC)**, v. 1, n. 1, p. 1–29, 2009.
- CARVALHO, D. Pintanela de; ROCHA, L. P.; BRUM, A. N.; GUTIERRES, E. D.; NOGARIO, A. C. D.; BORDIGNON, S. S. Perfil de trabalhadores da enfermagem de hospitais universitários e as cargas de trabalho: análise por cluster. **Anna Nery School Journal of Nursing/Escola Anna Nery Revista de Enfermagem**, v. 26, 2022.
- CRESWELL, J. W.; POTTH, C. N. **Qualitative inquiry and research design: Choosing among five approaches**. [S. l.]: Sage publications, 2016.
- FONSECA, F. C. S.; BELTRAME, W. A. R. Aplicações práticas dos algoritmos de clusterização k-means e bisecting k-means. **UFES, Vitória**, 2010.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The weka data mining software: an update. **ACM SIGKDD explorations newsletter**, ACM New York, NY, USA, v. 11, n. 1, p. 10–18, 2009.
- HEINZEN, R.; MARTINS, P. D. S. Sistema para identificação de perfil de consumidores utilizando análise de agrupamento (clusterização). **Ciência da Computação-Tubarão**, 2018.
- KOTLER, P. Keller.(2016). **Marketing management**, v. 15, 2015.
- KOTZ, S.; BALAKRISHNAN, N.; JOHNSON, N. L. **Continuous multivariate distributions: Models and applications**. [S. l.]: John Wiley & Sons, 2004. v. 1.
- PIMENTEL, E. P.; FRANÇA, V. F. de; OMAR, N. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In: **Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)**. [S. l.: s. n.], 2003. v. 1, n. 1, p. 495–504.
- PIRES, M. O.; VIEIRA, S. A. G.; BACKES, D. S.; ROOS, M. de O.; ADAMES, N. Óbitos infantis entre os anos 2000 e 2017 em uma cidade do sul do brasil: técnicas de mineração de dados. **Research, Society and Development**, v. 9, n. 9, p. e550997489–e550997489, 2020.
- SANTANA, R. V.; PONTES, H. L. J. Aplicação da clusterização por k-means para criação de sistema de recomendação de produtos baseado em perfis de compra. **Navus: Revista de Gestão e Tecnologia**, Serviço Nacional de Aprendizagem Comercial (Senac), n. 10, p. 31, 2020.
- SCHIFFMAN, L.; KANUK, L.; WISENBLIT, J. **Consumer Behavior**. [S. l.]: Pearson Education/Prentice Hall, 2010. ISBN 9780135053010.
- SILVA, A. A. da. **Sobre Nós: And Store**. 2023. Disponível em: <https://www.lojaand.com.br/empresa>. Acesso em: 20 ago. 2023.
- SILVA, A. M. E.; PORTUGAL, M.; CECHIN, A. **Redes Neurais Artificiais e Análise de Sensibilidade: Uma aplicação à demanda de importações brasileira**. [S. l.]: Revecap, 2001.
- SMIDERLE, A.; OLIVEIRA, A. M. Mineração de dados: um estudo de caso de concessão de crédito explorando o software weka. In: **CNMAC–Congresso Nacional de Matemática Aplicada e Computacional. Santo Amaro–SP**. [S. l.: s. n.], 2005.



WITTEN, I. H.; FRANK, E.; HALL, M. A.; PAL, C. J.; DATA, M. Practical machine learning tools and techniques. In: ELSEVIER AMSTERDAM, THE NETHERLANDS. **Data mining**. [S. l.], 2005. v. 2, n. 4, p. 403–413.

**APÊNDICE A – CÓDIGO-FONTE ATRIBUTOS**

## Código-fonte 1 – Atributos ARFF

```
1 @attribute sexo {m, f, pni}
2 @attribute idade {<=19, 20-29, 30-39, 40-49, 50-59, >=60,
   pni}
3 @attribute endereco {local, naolocal}
4 @attribute renda {<=1, 1-2, 2-4, 4-6, >6, pni}
5 @attribute qtd_camisas {1, 2-3, 4-5, >5, pni}
6 @attribute promocao {sim, nao, ambos}
7 @attribute time {flamengo, palmeiras, corinthians, psg,
   fortaleza, usa, santos, arsenal, borussia, barcelona,
   brasil, ceara, saopaulo, juventus, ajax, argentina,
   vasco, intermiami, milan, italia, franca, holanda,
   bocajrs,.realmadrid, mancity, inglaterra, belgica,
   atleticomineiro, coreiadosul, tottenham, bayernmunique,
   alnassr, alhilal, croacia, portugal, nba, alemanha,
   manunited}
8 @attribute motivo {uso, presente, ambos}
9 @attribute data {diamaes, dianamorados, diapais, diacrianca
   , blackfriday, natal, anonovo, semdata, pni}
10 @attribute meio_comprou {lojafisica, redessociais, site,
   pni}
11 @attribute indicacao {1, 2, 3, 4, 5, pni}
```

**APÊNDICE B – CÓDIGO-FONTE DADOS**

## Código-fonte 2 – Alguns dados ARFF

```
1 @data
2 m, 20-29, naolocal, 1-2, 2-3, nao, flamengo, presente,
   semdata, redessociais, 5
3 m, 20-29, naolocal, >6, >5, ambos, palmeiras, ambos,
   semdata, redessociais, 5
4 m, 20-29, naolocal, >6, >5, ambos, corinthians, ambos,
   semdata, redessociais, 5
5 m, 20-29, naolocal, >6, >5, ambos, psg, ambos, semdata,
   redessociais, 5
6 f, 30-39, naolocal, <=1, pni, ambos, corinthians, ambos,
   pni, redessociais, 5
7 m, 30-39, local, 2-4, >5, nao, palmeiras, uso, semdata,
   lojafisica, 5
8 m, 20-29, local, <=1, 1, sim, flamengo, uso, semdata,
   redessociais, 5
```

**APÊNDICE C – CÓDIGO-FONTE TRATAMENTO**

## Código-fonte 3 – Tratando dados

```
1 @data
2 m, 20-29, local, 2-4, >5, sim, Fortaleza, ambos, outro,
   lojafisica, 5
3 m, 20-29, local, 2-4, >5, sim, Brasil, ambos, outro,
   lojafisica, 5
```

## ANEXO A – QUESTIONÁRIO UTILIZADO

O objetivo do questionário, é conhecer melhor os perfis de clientes da Loja And Store. PESQUISA TOTALMENTE EM ANONIMATO e os dados serão usados APENAS para o Trabalho de Conclusão de Curso. Não serão usados e nem divulgados em outros lugares. Esse questionário é para QUEM JÁ COMPROU, então caso nunca tenha comprado, por favor, não responder, visto que pode mudar a veracidade dos resultados.

Todas as perguntas são obrigatórias.

### Questão 1. SEXO:

- Masculino
- Feminino
- Outro
- Prefiro não informar

### Questão 2. IDADE:

- Até 19 anos
- Entre 20 até 29 anos
- Entre 30 até 39 anos
- Entre 40 até 49 anos
- Entre 50 até 59 anos
- 60 anos ou mais
- Prefiro não informar

### Questão 3. ENDEREÇO:

### Questão 4. RENDA:

- Até 1 salário mínimo
- Entre 1 e 2 salários mínimos
- Entre 2 e 4 salários mínimos
- Entre 4 e 6 salários mínimos
- Acima de 6 salários mínimos
- Prefiro não informar

### Questão 5. QUANTAS CAMISAS COMPROU:

- Apenas 1ª camisa
- Entre 2ª e 3ª camisas

- Entre 4ª e 5ª camisas
- Mais de 5ª camisas
- Prefiro não informar

**Questão 6.** VOCÊ COMPROU APENAS QUANDO OS PRODUTOS ESTAVAM EM PROMOÇÃO:

- Sim, apenas em promoção
- Não, nunca comprei em promoção
- Ambos, já comprei com e sem promoção

**Questão 7.** COMPROU CAMISAS DE QUAIS TIMES:

**Questão 8.** MOTIVO DA COMPRA:

- Para Uso
- Para Presente
- Ambos

**Questão 9.** EM QUAIS DESSAS DATAS VOCÊ JÁ COMPROU CAMISAS CONOSCO:

- Dia das Mães
- Dia dos Namorados
- Dia dos Pais
- Dia das Crianças
- Black Friday
- Natal
- Ano Novo
- Outro
- Prefiro não informar

**Questão 10.** VOCÊ COMPROU POR QUAL MEIO:

- Comprei na loja física
- Comprei pelas redes sociais (WhatsApp, Instagram e Facebook)
- Comprei no site
- Prefiro não informar

**Questão 11.** VOCÊ INDICARIA NOSSA LOJA: (SENDO 1 RUIM e 5 EXCELENTE)

- 1
- 2

- 3
- 4
- 5
- Prefiro não informar