Relatório de Desempenho da Loja em 2022

Consultor Responsável:

Marina Gabriela de Oliveira Cadete

Requerente:

Barbie

Brasília, 15 de agosto de 2025

ESTAT
consultoria estatística

ESTAT consultoria estatística

Conteúdo

1 Introdução

O objetivo deste relatório é fornecer uma compreensão mais profunda das informações relacionadas às vendas realizadas ao longo de 2022 pela loja de roupas da nossa cliente. A análise desempenha um papel de suma importância na avaliação do progresso das lojas ao longo do ano e na identificação de padrões no comportamento dos clientes, bem como possíveis oportunidades de crescimento.

Para atingir esse objetivo, foi conduzida uma análise exploratória e descritiva dos dados, explorando possíveis relações entre as variáveis. Além disso, quando necessário, testes de correlação e testes de hipóteses foram aplicados a um nível de significância de 5%, a fim de avaliar de maneira mais precisa as relações entre as variáveis e verificar as hipóteses apresentadas pelo gráfico.

Os dados utilizados foram fornecidos pela cliente e compreendem registros de vendas e devoluções. Os registros de vendas abrangem variáveis qualitativas ordinais, como a data da venda e o tamanho, além de variáveis qualitativas nominais, como o ID do usuário, ID do produto, nome do produto, marca, categoria de moda, cor e ID da venda. Destaca-se ainda a presença de variáveis quantitativas contínuas, como preço e avaliação, na base de dados de vendas. Já a base de devoluções contém o ID da venda e o motivo da devolução, que é qualitativa nominal. As bases de dados foram disponibilizadas de forma separada e, posteriormente, foram unidas e organizadas com base no id de venda, permitindo uma análise mais abrangente.

Além disso, a base de dados apresentava alguns registros de vendas duplicados que foram removidas, com base no ID de venda, para evitar possíveis impactos nos resultados.

Para realizar as análises, contei com o auxílio do software R na versão 4.3.1 (Beagle Scouts), uma ferramenta que facilitou a criação dos gráficos e na elaboração dos testes presentes neste relatório.

2 Referencial teórico

2.1 Frequência Relativa

A frequência relativa é utilizada para a comparação entre classes de uma variável categórica com c categorias, ou para comparar uma mesma categoria em diferentes estudos.

A frequência relativa da categoria j é dada por:

$$f_j = \frac{n_j}{n}$$

Com:

- $j = 1, ..., c$
- $n_j =$ número de observações da categoria j
- $n =$ número total de observações

Geralmente, a frequência relativa é utilizada em porcentagem, dada por:

$$100 \times f_j$$

2.2 Média

A média é a soma das observações dividida pelo número total delas, dada pela fórmula:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Com:

- $i = 1, 2, ..., n$
- $n =$ número total de observações

2.3 Mediana

Sejam as n observações de um conjunto de dados $X = X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ de determinada variável ordenadas de forma crescente. A mediana do conjunto de dados X é o valor que deixa metade das observações abaixo dela e metade dos dados acima. Com isso, pode-se calcular a mediana da seguinte forma:

$$med(X) = \begin{cases} X_{\frac{n+1}{2}}, \text{ para n ímpar;} \\ \frac{X_{\frac{n}{2}} + X_{\frac{n}{2}+1}}{2}, \text{para n par.} \end{cases}$$

2.4 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) é o conjunto que delimita os 25% menores valores, o segundo representa a mediana e é o valor que ocupa a posição central (ou seja, metade dos dados estão abaixo dela e a outra metade está acima) e o terceiro delimita os 25% maiores valores. Inicialmente deve-se calcular a posição do quartil:

- Posição do primeiro quartil P_1:

$$P_1 = \frac{n+1}{4}$$

- Posição da mediana (segundo quartil) P_2:

$$P_2 = \frac{n+1}{2}$$

- Posição do terceiro quartil P_3:

$$P_3 = \frac{3 \times (n+1)}{4}$$

Com n sendo o tamanho da amostra. Dessa forma, $X_{(P_i)}$ é a o valor do i-ésimo quartil, onde $X_{(j)}$ representa a j-ésima observação dos dados ordenados.

Se o cálculo da posição resultar em uma fração deve-se fazer a média entre o valor que está na posição do inteiro anterior e do seguinte ao da posição.

2.5 Variância

A variância é uma medida que avalia o quanto que os dados estão dispersos em relação à média, em uma escala ao quadrado da escala dos dados.

2.5.1 Variância Populacional

Para uma população, a variância é dada por:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Com:

- $X_i = i$-ésima observação da população
- $\mu =$ média populacional
- $N =$ tamanho da população

2.5.2 Variância Amostral

Para uma amostra, a variância é dada por:

$$S^2 = \frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}$$

Com:

- $X_i =$ i-ésima observação da amostra
- $\bar{X} =$ média amostral
- $n =$ tamanho da amostra

2.6 Desvio Padrão

O desvio padrão é a raiz quadrada da variância. Avalia o quanto os dados estão dispersos em relação à média.

2.6.1 Desvio Padrão Populacional

Para uma população, o desvio padrão é dado por:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}}$$

Com:

- $X_i =$ i-ésima observação da população
- $\mu =$ média populacional
- $N =$ tamanho da população

2.6.2 Desvio Padrão Amostral

Para uma amostra, o desvio padrão é dado por:

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{n-1}}$$

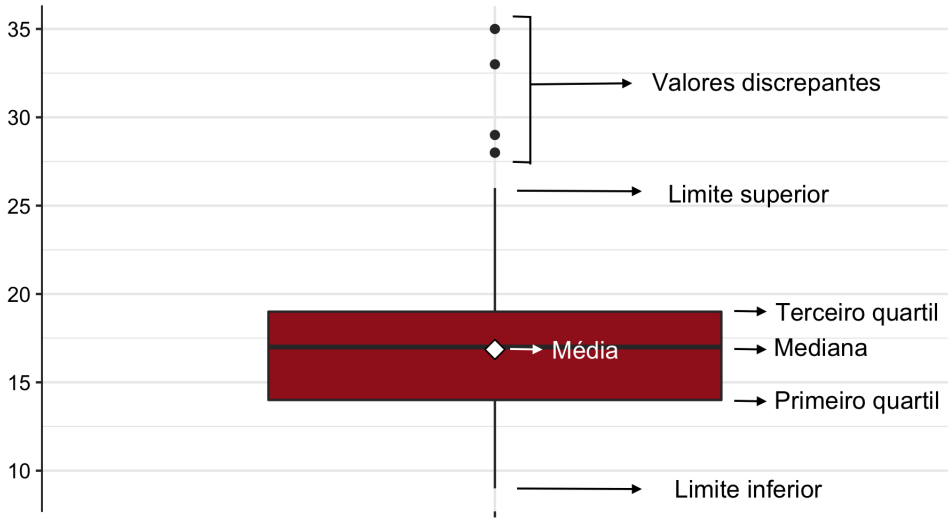
Com:

- $X_i =$ i-ésima observação da amostra
- $\bar{X} =$ média amostral
- $n =$ tamanho da amostra

2.7 Boxplot

O boxplot é uma representação gráfica na qual se pode perceber de forma mais clara como os dados estão distribuídos. A figura abaixo ilustra um exemplo de boxplot.

Figura 1: Exemplo de boxplot

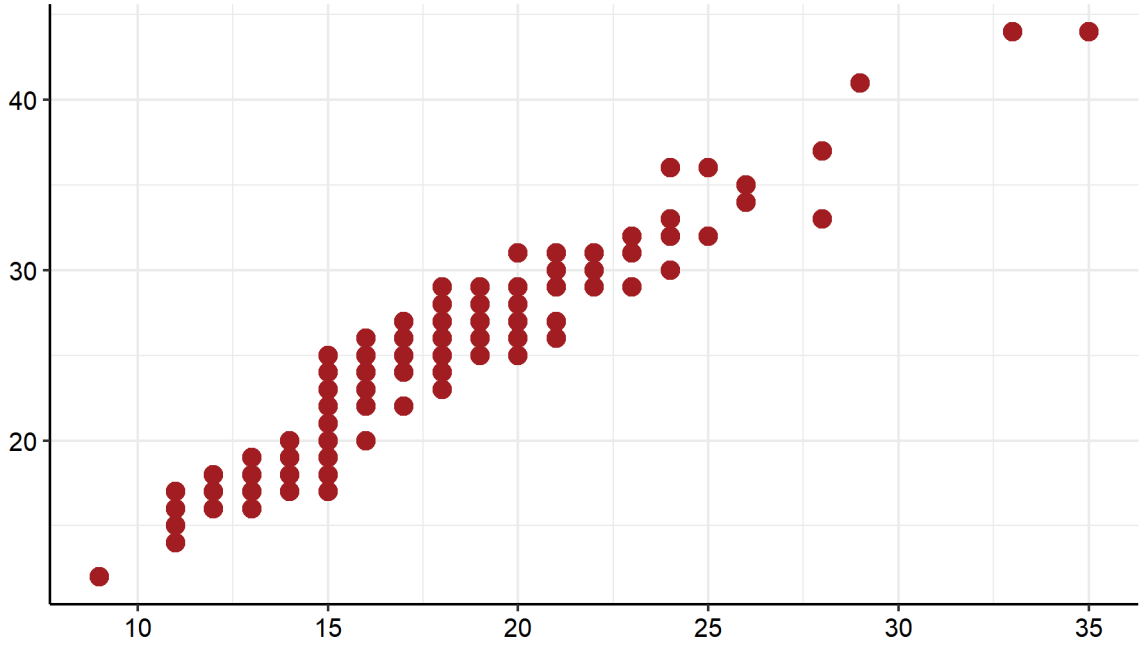


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada pelo losango branco e os pontos são *outliers*. Os *outliers* são valores discrepantes da série de dados, ou seja, valores que não demonstram a realidade de um conjunto de dados.

2.8 Gráfico de Dispersão

O gráfico de dispersão é uma representação gráfica utilizada para ilustrar o comportamento conjunto de duas variáveis quantitativas. A figura abaixo ilustra um exemplo de gráfico de dispersão, onde cada ponto representa uma observação do banco de dados.

Figura 2: Exemplo de Gráfico de Dispersão



2.9 Tipos de Variáveis

2.9.1 Qualitativas

As variáveis qualitativas são as variáveis não numéricas, que representam categorias ou características da população. Estas subdividem-se em:

- Nominais: quando não existe uma ordem entre as categorias da variável (exemplos: sexo, cor dos olhos, fumante ou não, etc)
- Ordinais: quando existe uma ordem entre as categorias da variável (exemplos: nível de escolaridade, mês, estágio de doença, etc)

2.9.2 Quantitativas

As variáveis quantitativas são as variáveis numéricas, que representam características numéricas da população, ou seja, quantidades. Estas subdividem-se em:

- Discretas: quando os possíveis valores são enumeráveis (exemplos: número de filhos, número de cigarros fumados, etc)

- Contínuas: quando os possíveis valores são resultado de medições (exemplos: massa, altura, tempo, etc)

2.10 Coeficiente de Correlação de Pearson

O coeficiente de correlação de Pearson é uma medida que verifica o grau de relação linear entre duas variáveis quantitativas. Este coeficiente varia entre os valores -1 e 1. O valor zero significa que não há relação linear entre as variáveis. Quando o valor do coeficiente r é negativo, diz-se existir uma relação de grandeza inversamente proporcional entre as variáveis. Analogamente, quando r é positivo, diz-se que as duas variáveis são diretamente proporcionais.

O coeficiente de correlação de Pearson é normalmente representado pela letra r e a sua fórmula de cálculo é:

$$r_{Pearson} = \frac{\sum_{i=1}^{n} \left[(x_i - \bar{x}) \, (y_i - \bar{y}) \right]}{\sqrt{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \times \sqrt{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2}}$$

Onde:

- $x_i =$ i-ésimo valor da variável X
- $y_i =$ i-ésimo valor da variável Y
- $\bar{x} =$ média dos valores da variável X
- $\bar{y} =$ média dos valores da variável Y

Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

2.11 Qui-Quadrado

A estatística Qui-Quadrado é uma medida de divergência entre a distribuição dos dados e uma distribuição esperada ou hipotética escolhida. Pode também ser usada para verificar independência ou determinar associação entre variáveis categóricas. É calculada pela seguinte fórmula:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Com:

- $O_i =$ frequência observada
- $E_i =$ frequência esperada

2.12 Teste de Hipóteses

O teste de hipóteses tem como objetivo fornecer uma metodologia para verificar se os dados das amostras possuem indicativos que comprovem, ou não, uma hipótese previamente formulada. Ele é composto por duas hipóteses:

$$\begin{cases} H_0 : \text{hipótese a ser testada (chamada de hipótese nula)} \\ H_1 : \text{hipótese alternativa que será aceita caso a hipótese nula seja rejeitada} \end{cases}$$

Essa decisão é tomada por meio da construção de uma região crítica, ou seja, região de rejeição do teste.

2.12.1 Tipos de teste: bilateral e unilateral

Para a formulação de um teste, deve-se definir as hipóteses de interesse. Em geral, a hipótese nula é composta por uma igualdade (por exemplo, $H_0 : \theta = \theta_0$). Já a hipótese alternativa depende do grau de conhecimento que se tem do problema em estudo. Assim, tem-se três formas de elaborar H_1 que classificam os testes em duas categorias:

- Teste Bilateral: Esse é o teste mais geral, em que a hipótese alternativa consiste em verificar se existe diferença entre os parâmetros de interesse, independentemente de um ser maior ou menor que o outro. Dessa forma, tem-se:

$$H_1 : \theta \neq \theta_0$$

- Teste Unilateral: dependendo das informações que o pesquisador possui a respeito do problema e os questionamentos que possui, a hipótese alternativa pode ser feita de forma a verificar se existe diferença entre os parâmetros em um dos sentidos. Ou seja,

$$H_1 : \theta < \theta_0 \text{ ou } H_1 : \theta > \theta_0$$

2.12.2 Tipos de Erros

Ao realizar um teste de hipóteses, existem dois erros associados: *Erro do Tipo I* e *Erro do Tipo II*.

- *Erro do Tipo I*: esse erro é caracterizado por rejeitar a hipótese nula (H_0) quando essa é verdadeira. A probabilidade associada a esse erro é denotada por α, também conhecido como nível de significância do teste.

- *Erro do Tipo II*: ao não rejeitar H_0 quando, na verdade, é falsa, está sendo cometido o *Erro do Tipo II*. A probabilidade de se cometer este erro é denotada por β.

2.12.3 Nível de significância (α)

Nível de significância do teste é o nome dado à probabilidade de se rejeitar a hipótese nula quando essa é verdadeira; essa rejeição é chamada de *erro do tipo I*. O valor de α é fixado antes da extração da amostra e, usualmente, assume 5%, 1% ou 0,1%.

Por exemplo, um nível de significância de $\alpha = 0,05$ (5%) significa que, se for tomada uma grande quantidade de amostras, em 5% delas a hipótese nula será rejeitada quando não havia evidências para essa rejeição, isto é, a probabilidade de se tomar a decisão correta é de 95%.

2.12.4 Estatística do Teste

Estatística do Teste é o estimador que será utilizado para testar se a hipótese nula (H_0) é verdadeira ou não. Ela é escolhida por meio das teorias estatísticas.

2.12.5 P-valor

P-valor, ou nível descritivo, é uma medida utilizada para sintetizar o resultado de um teste de hipóteses. Ele pode ser chamado também de *probabilidade de significância* do teste e indica a probabilidade de se obter um resultado da estatística de teste mais extremo do que o observado na presente amostra, considerando que a hipótese nula é verdadeira. Dessa forma, rejeita-se H_0 para $P-valor < \alpha$, porque a chance de uma nova amostra possuir valores tão extremos quanto o encontrado é baixa, ou seja, há evidências para a rejeição da hipótese nula.

2.13 Teste de Normalidade de Shapiro-Wilk

O Teste de Shapiro-Wilk é utilizado para verificar a aderência de uma variável quantitativa ao modelo da Distribuição Normal, sendo mais recomendado para amostras pequenas. A suposição de normalidade é importante para a determinação do teste a ser utilizado. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável segue outro modelo} \end{cases}$$

A amostra deve ser ordenada de forma crescente para que seja possível obter as estatísticas de ordem. A estatística do teste é dada por:

$$W = \frac{1}{D}\left[\sum_{i=1}^{k} a_i \left(X_{(n-i+1)} - X_{(i)}\right)\right]$$

Com:

- K aproximadamente $\frac{n}{2}$
- $X_{(i)} =$ estatística de ordem i
- $D = \sum_{i=1}^{n}(X_i - \bar{X})^2$, em que $\bar{X}$ é a média amostral
- $a_i =$ constantes que apresentam valores tabelados

2.14 Teste Qui-Quadrado

Os testes a seguir utilizam como base a estatística χ^2, apresentando mudanças nos graus de liberdade da sua distribuição de acordo com o teste que será utilizado. No geral,

$$\chi^2_v = \sum \frac{(o_i - e_i)^2}{e_i}$$

em que v expressa os graus de liberdade, o_i é a frequência observada e e_i é chamado de valor esperado e representa a frequência que seria observada se H_0 fosse verdadeira.

2.15 Teste de Kruskal-Wallis (Comparação de Médias)

O teste de Kruskal-Wallis é utilizado para comparar dois ou mais grupos independentes sem supor nenhuma distribuição. É um método baseado na comparação de postos, os quais são atribuídos a cada observação de uma variável quantitativa após serem ordenadas.

As hipóteses do teste de Kruskal-Wallis são formuladas da seguinte maneira:

$$\begin{cases} H_0 : \text{Não existe diferença entre os grupos} \\ H_1 : \text{Pelo menos um grupo difere dos demais} \end{cases}$$

A estatística do teste de Krukal-Waliis é definida da seguinte maneira:

$$H_{Kruskal-Wallis} = \frac{\left[\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i}\right] - 3(n+1)}{1 - \left[\frac{\sum_j (t_j^3 - t_j)}{n^3 - n}\right]} \approx \chi^2_{(k-1)}$$

Com:

- $k =$ número de grupos
- $R_i =$ soma dos postos do grupo i

- $n_i =$ número de elementos do grupo i
- $n =$ tamanho total da amostra
- $t_j =$ número de elementos no j-ésimo empate (se houver)

Se o p-valor for menor que o nível de significância α, rejeita-se a hipótese nula.

3 Análises

3.1 Faturamento Anual por Categoria

Para avaliar o faturamento no ano de 2022 foram utilizadas as variáveis categoria e preço. Categoria é uma variável qualitativa nominal que descreve os diferentes segmentos de roupas e calçados presente na loja, sendo eles moda feminina, moda masculina e moda infantil. Por sua vez, preço é quantitativa contínua e está relacionada ao valor de venda de cada produto, variando entre R$10,00 e R$100,00 reais.

Além disso, alguns registros de preço da venda e tipo de categoria não foram informados, essas linhas foram removidas para que a análise pudesse ser feita.

Figura 3: Gráfico de linhas do total vendido em 2022 ao longo dos meses por categoria

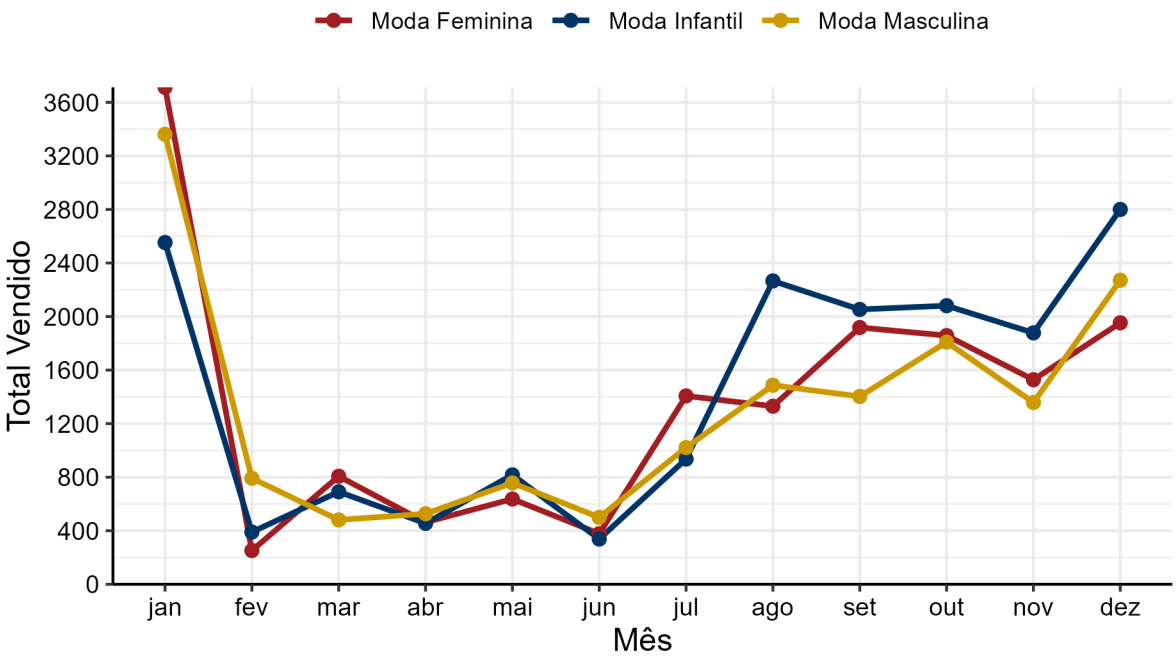


Tabela 1: Valor total vendido por categoria

Categoria	Total	Porcentagem
Moda Feminina	R$16.243	32,97%
Moda Infantil	R$17.258	35,03%
Moda Masculina	R$15.767	32,00%
Total	R$49.268	100%

Pela Figura 3 é possível ter uma ideia do total vendido ao longo dos meses e as tendências de cada categoria ao longo do ano. Em janeiro, todas as categorias re-

gistraram valores elevados, com destaque para as vendas de produtos masculinos e femininos que registraram seu máximo anual, enquanto em dezembro, produtos infantis lideraram nas vendas. Esses picos podem ter sido influenciados pelas festas de fim de ano e período de férias.

Em fevereiro há uma queda nas vendas em relação a janeiro e permanece baixa até junho. Esse é o intervalo que teve o menor registro de vendas em todas as categorias: feminina em fevereiro, masculina em março e a infantil em junho.

A partir de julho, houve um aumento em todas as categorias, com destaque nas vendas de produtos infantis em agosto, possivelmente relacionado ao início do ano letivo e/ou às férias escolares.

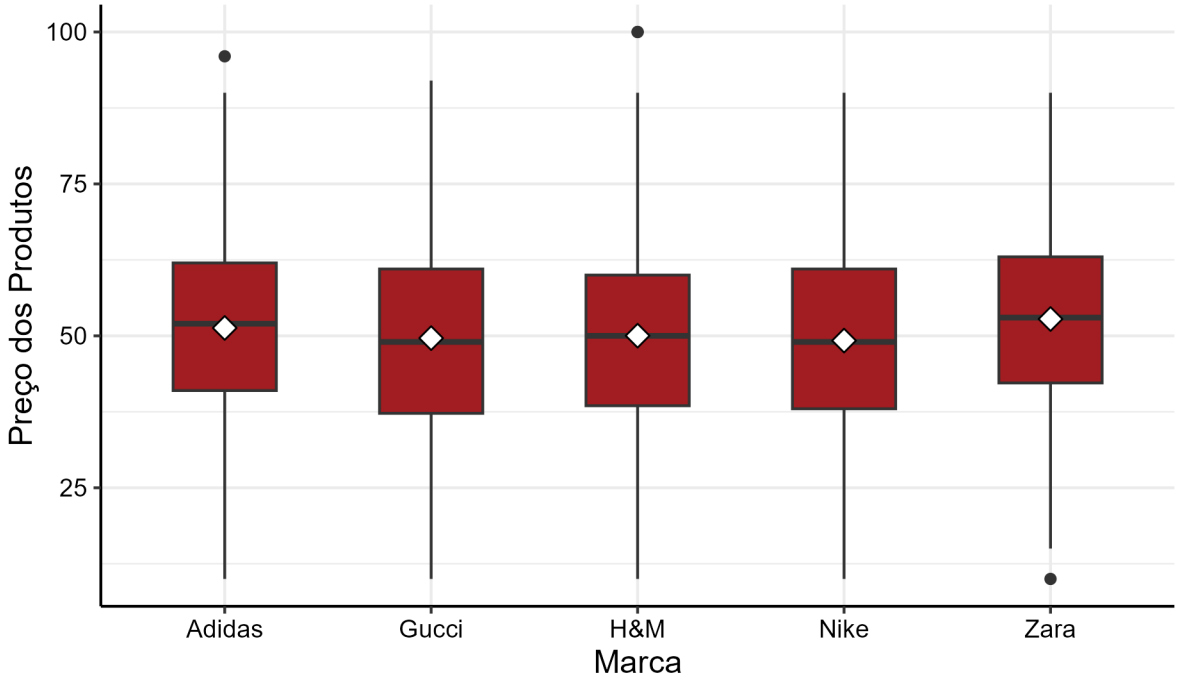
Ao examinar a Tabela 1 com o faturamento total ao longo do ano, nota-se que a categoria infantil obteve a maior proporção do faturamento em comparação as outras. No entanto, apesar das vendas de roupas e calçados infantis terem maior contribuição para um faturamento anual, a diferença proporcional entre as três não é tão expressiva. Isso evidencia um desempenho equilibrado entre todas as categorias da loja ao longo do ano.

3.2 Variação do Preço por Marca

A análise da variação de preços por marca nesta loja será feita com base nas variáveis preço e marca. A váriável marca é qualitativa nominal e classifica as peças em Adidas, Gucci, H&M, Nike e Zara. Por outro lado, preço é quantitativa contínua, variando de R$10 até R$100.

Na base de dados, alguns registros de preço e marca não foram informados, e para a realização da análise, essas linhas foram desconsideradas.

Figura 4: Boxplot do preço dos produtos pela marca



[H] Quadro de medidas resumo da variação do preço por marca

Estatísticas	Adidas	Gu
Média	51,30	49
Desvio Padrão	16,41	15
Mínimo	10	1
1º Quartil	41,00	37
Mediana	52	4
3º Quartil	62	6
Máximo	96	9

A partir da análise da Figura 4 e do Quadro 3.2, temos uma explicação mais detalhada da variação de preços entre as diferentes marcas. Percebe-se que a Adidas e

a H&M possuem preços discrepantes, ultrapassando o limite superior do boxplot, enquanto a Zara apresenta valores abaixo do limite inferior. Essas disparidades indicam que essas marcas possuem valores que distoam do padrão esperado para cada uma delas.

Ao analisar a mediana e a amplitude interquartil, notamos uma proximidade dessas medidas entre todas as marcas, com a mediana variando entre [49; 53] e o desvio padrão entre [15,40; 17,02]. Esses dados sugerem uma notável uniformidade nos preços, com variações moderadas. Essa constância também é observada ao considerar a média, representada pelo losango branco.

Com o intuito de verificar essa suposta uniformidade apresentada pelo Quadro 3.2 será realizado o teste de normalidade de Shapiro-Wilk. As hipóteses testadas serão:

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável segue outro modelo} \end{cases}$$

Quadro 1: P-valor do teste de normalidade (Shapiro-Wilk) da variável preço

Variável	**P-valor**	**Decisão do teste**
Preço	0,187	Não rejeita H_0

Isso nos mostra que, a um nível de significancia de 5%, a variável preço segue de fato uma distribuição normal como a sugerida pelo quadro de medidas resumo, com uma concentração maior de valores em torno da média e um desvio padrão relativamente simetrico. Sabendo que a variável preço segue uma normal, o objetivo agora é verificar se as medianas entre todas as marcas são de fato semelhantes, conforme indicado pela Figura 4.

Para isso, será realizado o teste de Kruskal-Wallis dada as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Não existe diferença entre as marcas} \\ H_1 : \text{Pelo menos uma marca difere das demais} \end{cases}$$

Quadro 2: P-valor do teste de comparação da mediana de preço das marcas (Teste de Kruskall-Wallis)

Marca	P-valor	Decisão do teste
Adidas Gucci H&M Nike Zara	0,181	Não rejeita H_0

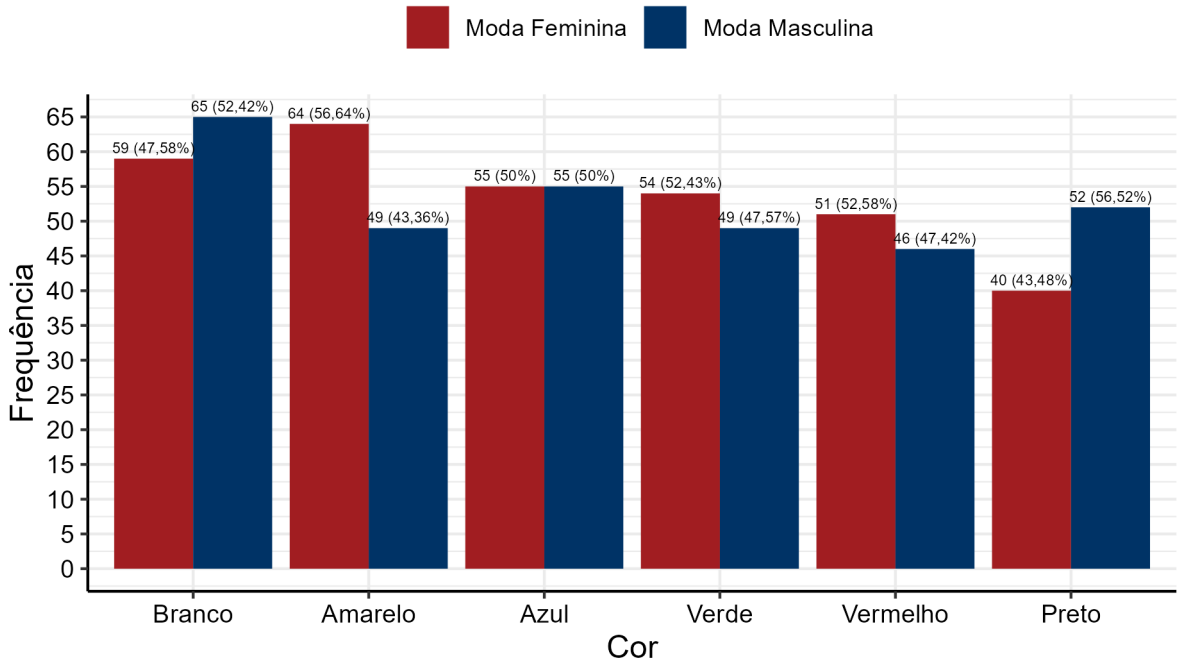
Com base no resultado do teste de Kruskal-Wallis, realizado com um nível de significância de 5%, onde foi obtido um p-valor de 0,181, não há evidências para rejeitar a hipótese nula. Portanto, não podemos dizer que há diferença significativa na mediana dos preços das marcas presentes na loja, concluindo que a distribuição dos preços é semelhante entre todas elas.

3.3 Relação entre Categoria e Cor

Nesta análise de relação, serão utilizadas as variáveis categoria e cor. Essas duas variáveis são classificadas como qualitativas nominais. Categoria contém informações sobre os diferentes segmentos de moda disponível na loja, como feminino, masculino e infantil. No entanto, para esta análise, concentraremos nossa atenção apenas nas categorias de moda feminina e masculina. A variável cor abrange informações sobre as cores dos produtos, que podem ser preta, azul, verde, vermelha, branca ou amarela.

Três produtos não tiveram sua cor informada e por isso essas linhas foram desconsideradas na análise.

Figura 5: Gráfico de colunas da cor da peça pela categoria de moda



Analisando o gráfico, podemos observar que o branco é a cor com maior frequência nas peças vendidas na categoria de moda masculina, enquanto o amarelo lidera as preferências de moda feminina. É interessante notar também, que a cor preta é a menos adquirida na categoria feminina, enquanto o vermelho é a menos popular entre os compradores de moda masculina.

Apesar dessas observações, é importante notar que as frequências de cores não apresentam discrepâncias consideráveis entre as categorias. As porcentagens das cores mais e menos populares estão relativamente próximas, indicando que, embora haja variação, não existe uma grande diferença nas preferências de cor entre os dois

grupos de consumidores.

O teste Qui-Quadrado será feita para avaliar se há associação entre as variáveis.

$$\begin{cases} H_0 : \text{Não há associação entre as variáveis} \\ H_1 : \text{As variáveis estão associadas} \end{cases}$$

Quadro 3: P-valor do teste de associação (Qui-Quadrado) entre categoria e cor

Variáveis	**P-valor**	**Decisão do teste**
Categoria Cor	0,511	Não rejeita H_0

Com base no resultado do teste Qui-Quadrado, que visa determinar a significância estatística da associação entre as variáveis, foi obtido um p-valor de 0,511. A um nível de significância de 5%, a conclusão é que não existem evidências suficientes para rejeitar a hipótese nula, ou seja, não é possível afirmar que há uma relação significativa entre a categoria de moda e as diferentes cores.

3.4 Relação entre Preço e Avaliação

A análise a seguir busca identificar uma possível relação entre o preço das roupas e as avaliações que elas receberam. Ambas as variáveis são consideradas quantitativas contínuas, sendo que preço varia entre 10 e 100 reais, e as avaliações variam no intervalo de 1,05 a 4,18.

Alguns valores de pagamento e avaliação não foram informados e, consequentemente, foram excluídos para a realização da análise.

Figura 6: Gráfico de dispersão do preço pela avaliação

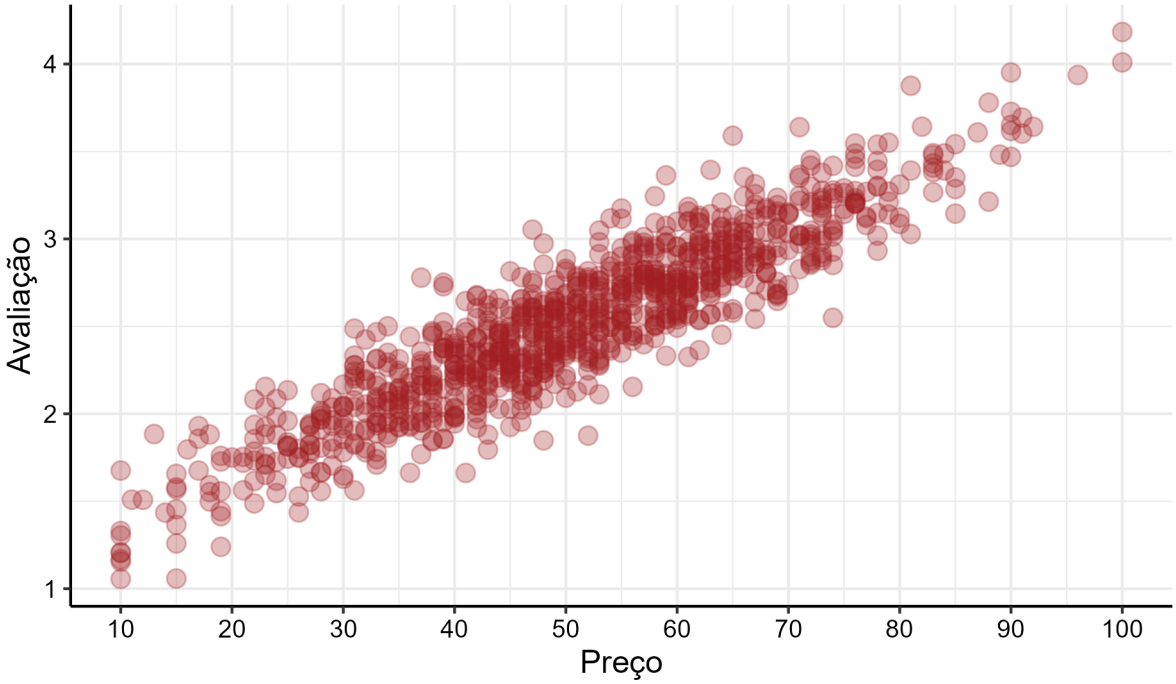


Figura 7: Boxplot da variável Preço

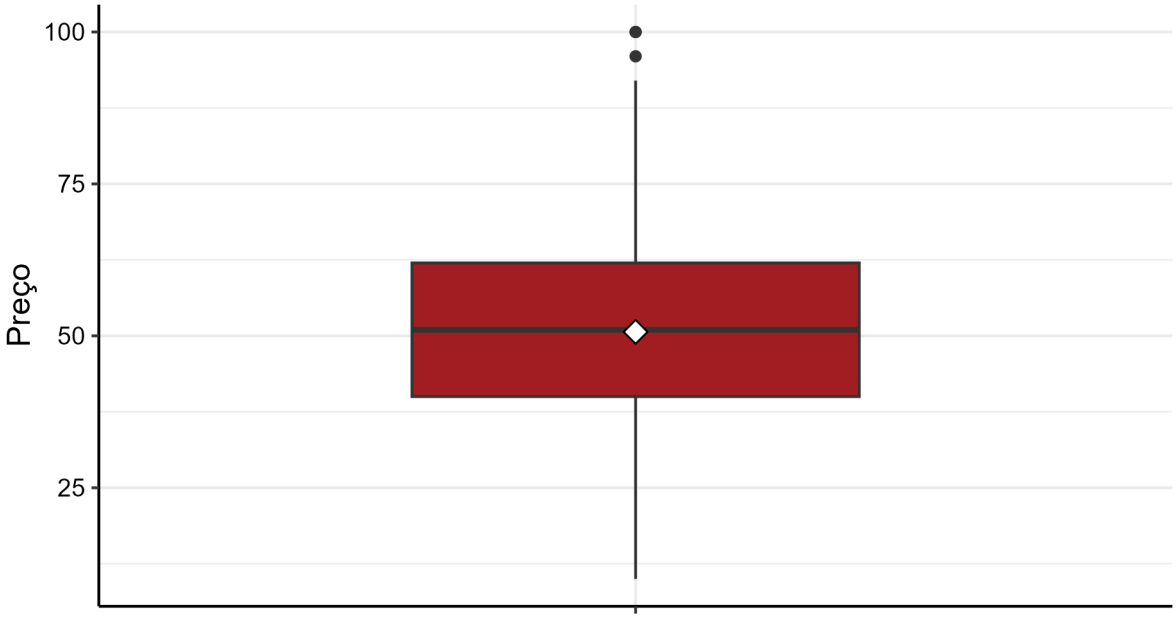
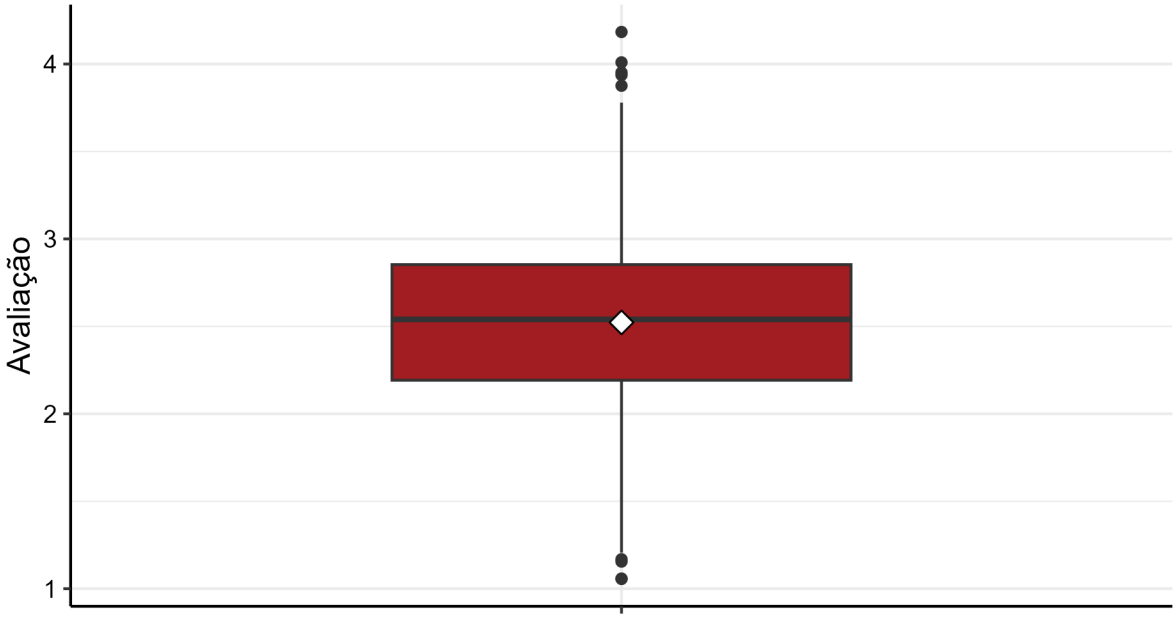


Figura 8: Boxplot da variável Avaliação



Quadro 4: Medidas resumo das avaliações e dos preços

Estatísticas	Avaliação	Preço
Média	2,52	50,64
Desvio Padrão	0,49	16,30
Mínimo	1,06	10
1º Quartil	2,19	40
Mediana	2,54	50,50
3º Quartil	2,84	62
Máximo	4,18	100

A análise da Figura 6 sugere uma relação entre o preço dos produtos e a avaliação atribuída pelos clientes. Ao calcular o coeficiente de correlação de Pearson, que avalia o grau de relação linear entre duas variáveis quantitativas, obtemos um coeficiente de 0,914. Esse valor indica que, de fato, existe uma forte correlação linear e positiva entre as duas variáveis, ou seja, à medida que o preço aumenta, a avaliação tende a ser melhor. Além disso, percebe-se que a maioria das vendas se concentram em preças com preços entre 40 e 70 reais, pois é a região com maior densidade de pontos, região essa que normalmente recebem avaliações entre 2 e 3.

Ao analisar os resultados do Quadro 4, percebe-se que as médias são próximas as medianas em ambas as variáveis, e a média está em torno do valor central, entre o mínimo e o máximo das variáveis, o que sugere uma simetria na distribuição. Em outras palavras, essas informações indicam que ambas as variáveis podem seguir uma distribuição normal.

A fim de confirmar essa hipótese, será realizado um teste de Shapiro-Wilk para as duas variáveis.

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável segue outro modelo} \end{cases}$$

Quadro 5: P-valor do teste de normalidade (Shapiro-Wilk) da variável preço

Variável	P-valor	Decisão do teste
Preço	0,187	Não rejeita H_0

$$\begin{cases} H_0 : \text{A variável segue uma distribuição Normal} \\ H_1 : \text{A variável segue outro modelo} \end{cases}$$

Quadro 6: P-valor do teste de normalidade (Shapiro-Wilk) da variável avaliação

Variável	**P-valor**	**Decisão do teste**
Avaliação	0,915	Não rejeita H_0

Ao realizar o teste de Shapiro-Wilk, considerando um nível de significância de 5%, os p-valores obtidos foram de 0,187 e 0,915, para o preço e a avaliação, respectivamente. Não rejeitamos a hipótese de que a distribuição das duas variáveis são normais.

3.5 Motivo de Devolução por Marca

Nessa análise serão usadas as variáveis tipo de devolução e marca. Ambas são classificadas como qualitativas nominais. Tipo de devolução tem as seguintes categorias: arrependimento, não informado e produto com defeito. E as marcas são Nike, Adidas, Zara, Gucci e H&M.

Dos 990 regristros de compra que informaram a marca do produto, 347 foram devolvidos a loja.

Figura 9: Gráfico de colunas do motivo de devolução por marca

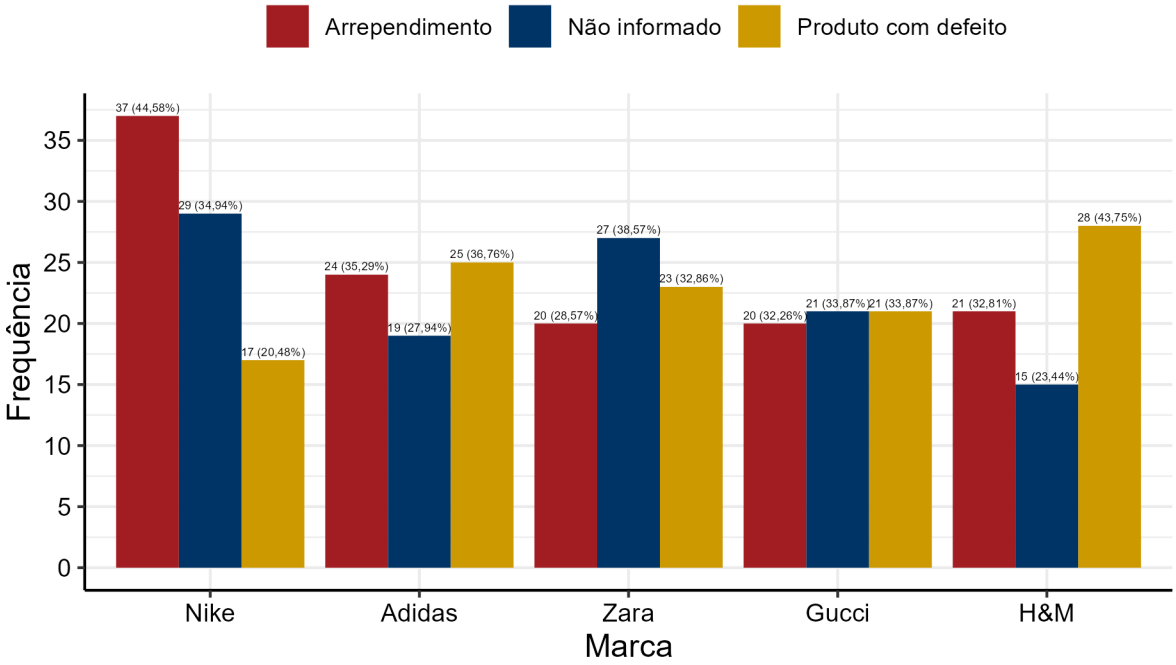


Tabela 2: Frequências da marca pelo motivo de devolução

Marca	Motivo Devolução			Total
	Arrependimento	Não Informado	Produto com Defeito	
Adidas	24	19	25	68
Gucci	20	21	21	62
H&M	21	15	28	64
Nike	37	29	17	83
Zara	20	27	23	70
Total	122	111	114	347

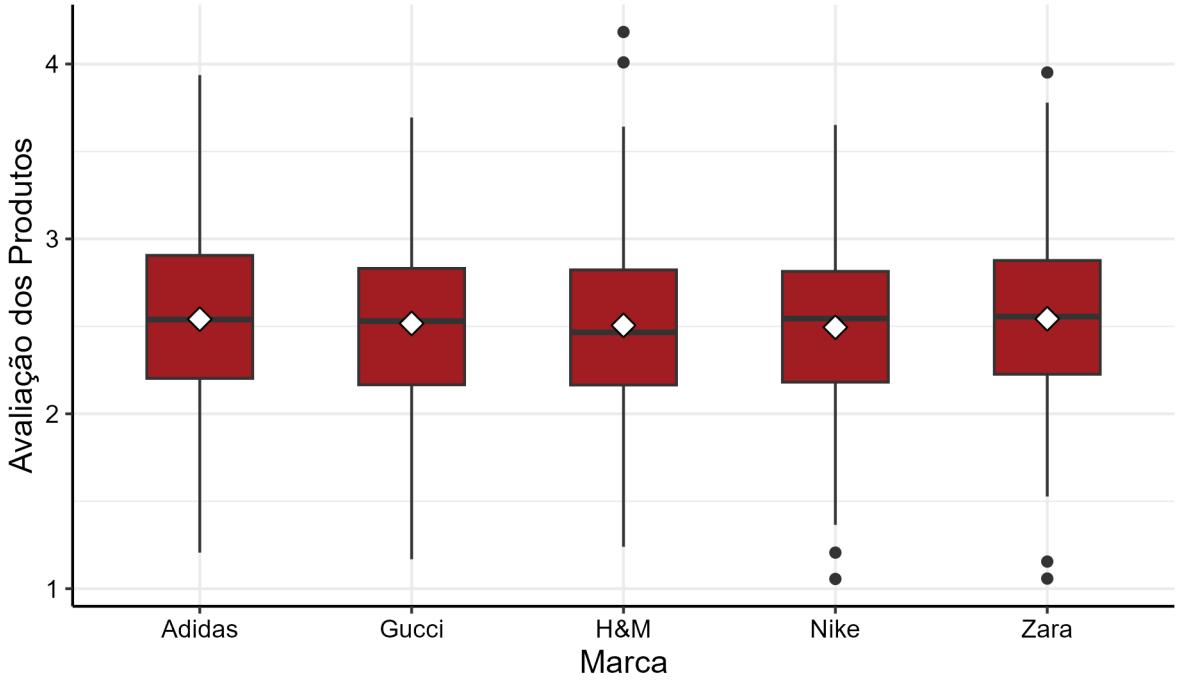
Ao analisar os resultados da Figura 9 e da Tabela 2, observa-se que a Nike destaca-se como líder em termos de frequência de devoluções por arrependimento e por motivos não informados. Além disso, representa 24% das devoluções totais.

Quando se trata de devoluções por defeito no produto, a H&M assume a liderança com 28 devoluções. No caso de produtos Adidas, tanto devoluções por arrependimento quanto por defeito apresentam frequências semelhantes, sendo que as devoluções não informadas são as menos registradas. Na Zara, observa-se o oposto da Adidas, as devoluções por motivos não informados lidera em números de registros em comparação aos demais motivos. E por fim, na marca Gucci, todos os motivos possuem frequências bem próximas, sem que nenhuma se sobressaia de forma considerável.

3.6 Avaliação Média por Marca

A análise a seguir pretende verificar a média de avaliação atribuída às marcas. A variável utilizada é quantitativa contínua e varia entre 1,06 e 4,18.

Figura 10: Boxplot das avaliações por marca



Quadro 7: Quadro de medidas resumo da variação da avaliação por marca

Estatísticas	Adidas	Gucci	H&M	Nike	Zara
Média	2,54	2,52	2,51	2,49	2,54
Desvio Padrão	0,50	0,48	0,49	0,50	0,48
Mínimo	1,21	1,17	1,24	1,06	1,06
1º Quartil	2,20	2,17	2,16	2,18	2,23
Mediana	2,54	2,53	2,47	2,54	2,56
3º Quartil	2,91	2,83	2,82	2,81	2,88
Máximo	3,94	3,69	4,18	3,65	3,95

Pela Figura 10 observa-se que tanto a Adidas quanto a Zara apresentaram uma média de 2,54, que constitui a melhor avaliação recebida entre as cinco marcas. Em contrapartida, a Nike registrou a menor média, 2,49, que, apesar de ser inferior, não difere substancialmente da melhor avaliação. Ao analisar o Quadro 7 nota-se que as médias de avaliação entre as cinco marcas são bem parecidas, com variações próximas.

Para verificar se as médias não diferem significativamente, será realizado um teste de Kruskal-Wallis.

$$\begin{cases} H_0 : \text{Não existe diferença na mediana entre as marcas} \\ H_1 : \text{Pelo menos uma mediana difere das demais} \end{cases}$$

Quadro 8: P-valor do teste de comparação da mediana das avaliações das marcas (Teste de Kruskall-Wallis)

Marca	P-valor	Decisão do teste
Adidas Gucci H&M Nike Zara	0,852	Não rejeita H_0

O teste confirma a suposição. Com um nível de significância de 5%, obteve-se um p-valor de 0,852, o que indica que não há evidências suficientes para rejeitar a hipótese de que não existe diferença entre as medianas. Isso indica que todas as marcas da loja tem o mesmo nível de satisfação em relação aos seus produtos, sem que uma se destaque em relação as demais.

4 Conclusão

Os resultados revelam uma proporção semelhante no faturamento anual entre todas as categorias de moda disponíveis na loja, com destaque especial para as vendas de fim de ano e janeiro, onde foi registrado o maior volume de vendas. Contudo, é necessário chamar a atenção para os meses entre fevereiro e junho, nos quais as vendas apresentaram uma queda expressiva. Além disso, não foi possível identificar uma relação significativa entre as categorias e a cor dos produtos.

No que diz respeito às marcas, não percebe-se uma diferença subtancial na variação dos preços de venda entre elas. A distribuição dos preços em todas as marcas é semelhante, sendo que nenhuma se destaca como a mais cara ou barata em comparação com as outras, mesmo que algumas marcas possuam produtos com valores consideravelmente acima ou abaixo do esperado. O mesmo padrão é observado em relação à avaliação média recebida pelos produtos em relação às marcas, indicando um nível de satisfação semelhante entre todas elas.

Na análise da relação entre preço e avaliação, observa-se que produtos mais caros tendem a receber avaliações melhores. No entanto, é importante notar que a maioria das vendas corresponde a produtos com valores medianos. Isso destaca a importância de equilibrar a avaliação e o preço para otimizar as vendas.

Contudo, é fundamental estar atento aos casos de devolução e seus motivos, especialmente aqueles relacionados a produtos defeituosos, a fim de evitar avaliações negativas e redução no faturamente por conta de itens devolvidos. Além disso, dar uma atenção aos casos devolvidos por motivo de arrependimento é essencial, dado que essa situação ocorre com maior frequência. Considerando que a lei não obriga o cliente a informar o motivo nessas situações, compreender o comportamento e dar suporte ao cliente pode reduzir esses casos.