

Relatório Old Town Road

Consultores Responsáveis:
Lucas de Sousa Cuadros

Requerente:
João Sábio

Brasília, 9 de novembro de 2025.



Sumário

	Página
1 Introdução	3
2 Referencial Teórico	4
2.1 Análise Descritiva Univariada	4
2.1.1 Frequência Relativa	4
2.1.2 Média	4
2.1.3 Mediana	4
2.1.4 Quartis	5
2.1.5 Variância	5
2.1.6 Desvio Padrão	6
2.1.7 Coeficiente de Variação	7
2.1.8 Boxplot	7
2.1.9 Gráfico de Dispersão	8
2.1.10 Tipos de Variáveis	9
2.1.11 Coeficiente de Correlação de Pearson	9
3 Análises	11
3.1 Análise da Receita Média nos últimos 10 anos	11
3.2 Análise do Comportamento das Variáveis Peso e Altura	11
3.3 Análise da relação entre idade e loja	13
3.4 Análise das Top 3 Lojas e Top 3 produtos em 1889	15
4 Conclusões	18

1 Introdução

Com o objetivo de oferecer uma visão ampla e estratégica sobre o desempenho das lojas e o perfil dos clientes na região analisada, este relatório apresenta quatro estudos estatísticos complementares desenvolvidos a partir do banco de dados disponibilizado. As análises buscam gerar insights que apoiem decisões comerciais e operacionais, com base em evidências quantitativas.

A primeira análise aborda a receita média das lojas entre os anos de 1880 e 1889, convertendo os valores para reais a partir da cotação de R\$ 5,31 por dólar. Esse estudo permite compreender a evolução do faturamento médio ao longo da década e identificar tendências de crescimento ou retração no período.

Em seguida, é investigada a relação entre peso e altura dos clientes, com o intuito de avaliar, se existe correlação estatisticamente relevante entre essas variáveis. Essa análise possibilita entender melhor padrões físicos entre os consumidores e eventuais implicações para segmentações de mercado.

A terceira etapa examina o perfil etário dos clientes na cidade de Âmbar Seco, destacando as diferenças de idade média e distribuição etária entre as lojas locais. Essa visão detalhada contribui para identificar o público predominante em cada estabelecimento e orientar estratégias de atendimento e comunicação.

Por fim, a quarta análise identifica os três produtos mais vendidos nas três lojas com maior receita no ano de 1889, evidenciando os itens de maior desempenho comercial e os estabelecimentos mais representativos em termos de faturamento.

Em conjunto, essas análises fornecem uma base sólida para compreender a dinâmica de mercado, o comportamento dos consumidores e o desempenho financeiro das lojas, apoiando a definição de ações estratégicas e oportunidades de melhoria.

2 Referencial Teórico

2.1 Análise Descritiva Univariada

2.1.1 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:

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

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

$$100 \times f_j$$

2.1.2 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:

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

2.1.3 Mediana

Sejam as n observações de um conjunto de dados $X = X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 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:

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

2.1.4 Quartis

Os quartis são separatrizes que dividem o conjunto de dados em quatro partes iguais. O primeiro quartil (ou inferior) delimita os 25% menores valores, o segundo representa a mediana, 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 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.

2.1.5 Variância

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

2.1.5.1 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
- μ = média populacional
- N = tamanho da população

2.1.5.2 Variância Amostral

Para uma amostra, a variância é dada por:

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

Com:

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

2.1.6 Desvio Padrão

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

2.1.6.1 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
- μ = média populacional
- N = tamanho da população

2.1.6.2 Desvio Padrão Amostral

Para uma amostra, o desvio padrão é dado por:

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

Com:

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

2.1.7 Coeficiente de Variação

O coeficiente de variação fornece a dispersão dos dados em relação à média. Quanto menor for o seu valor, mais homogêneos serão os dados. O coeficiente de variação é considerado baixo (apontando um conjunto de dados homogêneo) quando for menor ou igual a 25%. Ele é dado pela fórmula:

$$C_V = \frac{S}{\bar{X}} \times 100$$

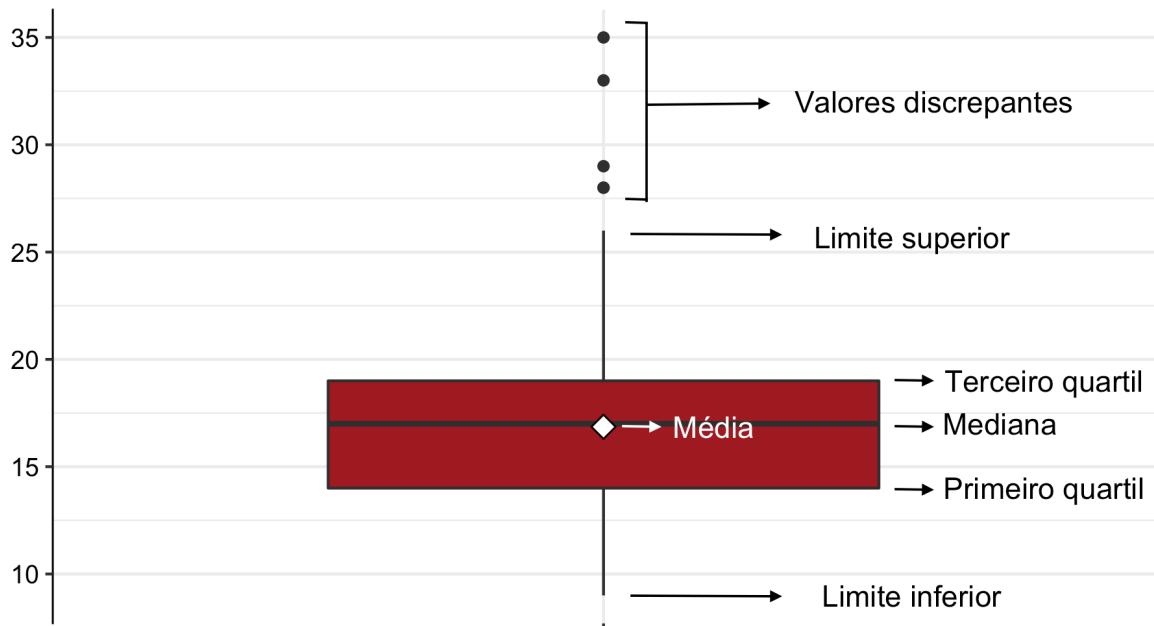
Com:

- S = desvio padrão amostral
- \bar{X} = média amostral

2.1.8 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

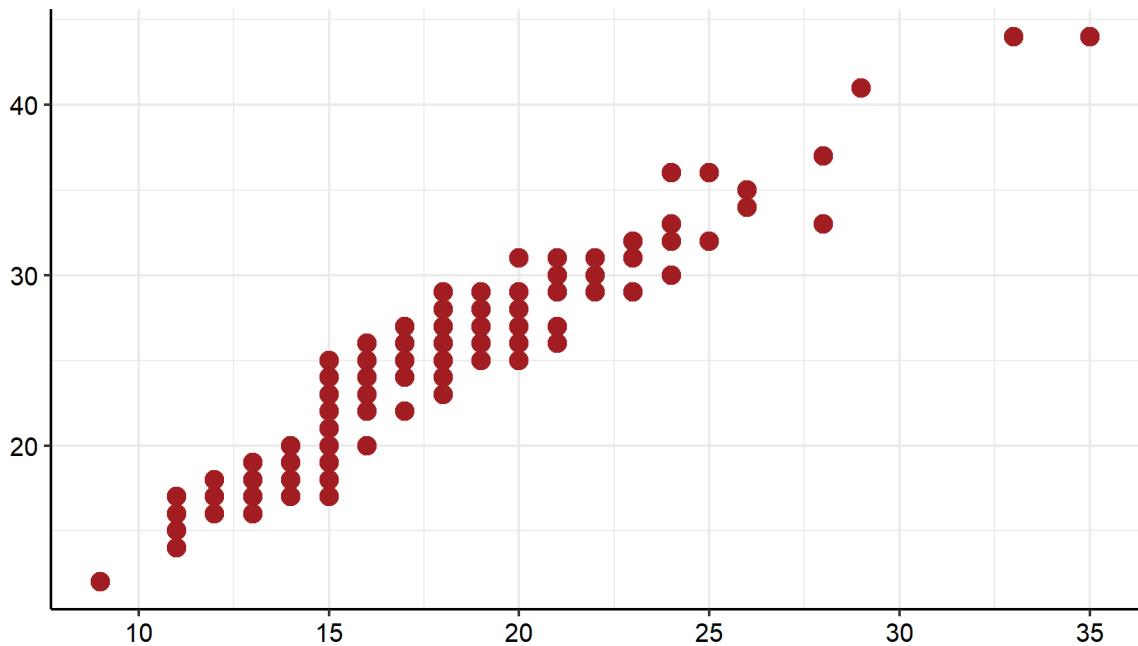


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada 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.

2.1.9 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



2.1.10 Tipos de Variáveis

2.1.10.1 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)

2.1.10.2 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)

2.1.11 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:

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

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

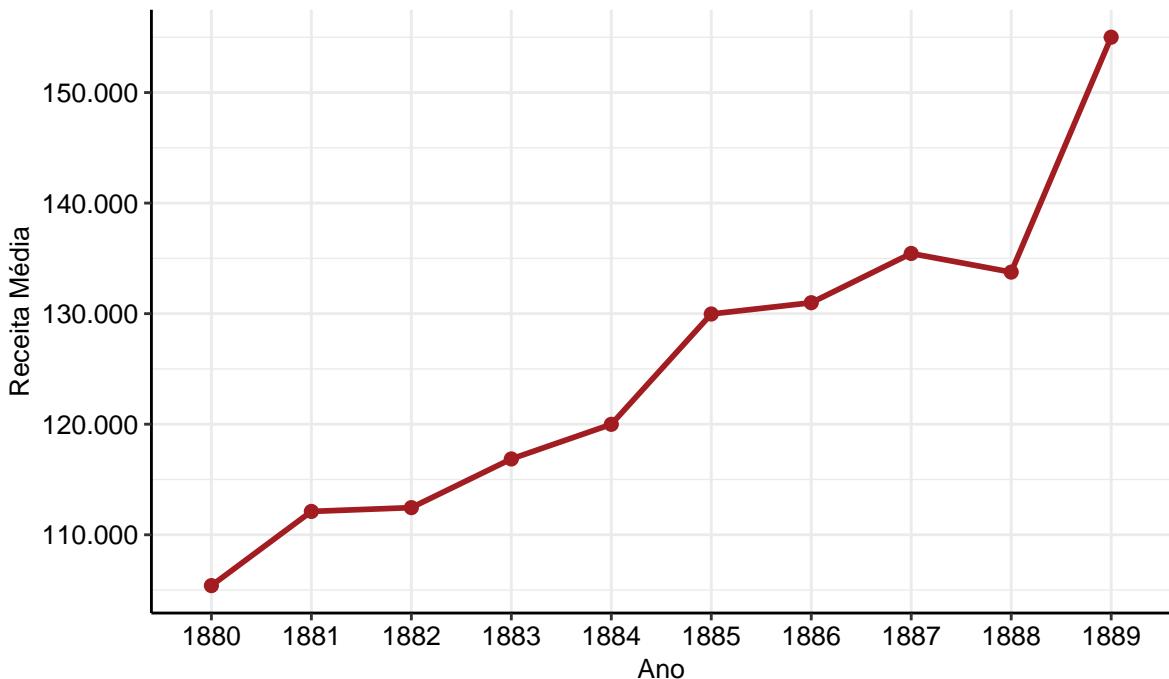
Vale ressaltar que o coeficiente de Pearson é paramétrico e, portanto, sensível quanto à normalidade (simetria) dos dados.

3 Análises

3.1 Análise da Receita Média nos últimos 10 anos

Nesta análise, busca-se compreender se houve um aumento ou diminuição na quantidade de receita produzida, em reais média ao longo do tempo. Essas variáveis são quantitativa contínua e discreta, respectivamente. Essa análise partirá do cálculo da receita nos últimos 10 anos (1880 a 1889). Para analisarmos o comportamento das variáveis, faremos um gráfico de linha. E, para isso, calculamos a receita total, a partir da soma de todas as receitas em cada ano, e dividimos cada resultado encontrado pelo número de lojas (18). Assim, obtém-se a média de receita por ano.

Figura 3: Gráfico de linha da media por ano



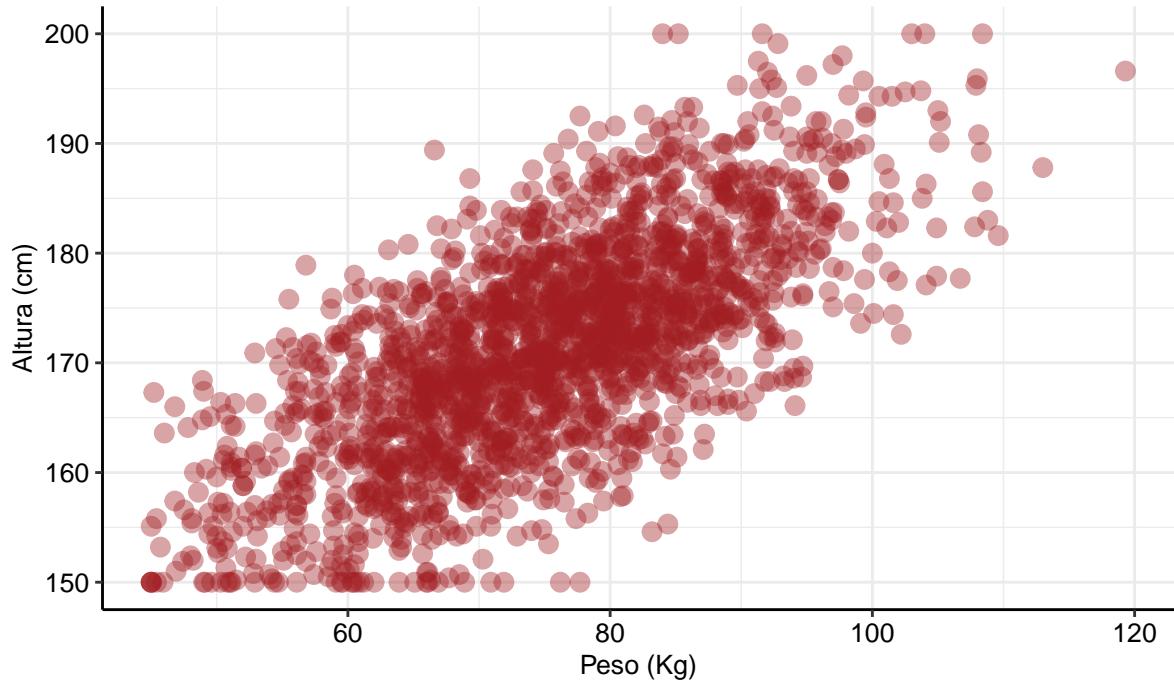
A partir da análise do **Figura 3** de linha, pode-se concluir que a receita média por loja vem crescendo significativamente com o passar dos anos. Tal análise pode ser feita com visualização da linha, que representa a receita média por ano, que é crescente. Entretanto, no ano de 1888, nota-se uma queda em tal número. Vale notar que o maior crescimento foi do ano de 1888 e 1889, ou seja, o método de venda que utilizaram para esse ano teve um sucesso muito grande.

3.2 Análise do Comportamento das Variáveis Peso e Altura

Esta análise tem como intuito entender a relação entre as variáveis peso(em kg) e altura (em cm), que são variáveis quantitativas contínuas. Para realizar essa análise

será necessário um gráfico para melhor entendimento visual da distribuição das alturas e idades. Além disso, será apresentado uma tabela para materializar os valores. Para realizá-los foi necessário converter todos os valores para os esperados e relacionar as variáveis.

Figura 4: Gráfico de Dispersão Bivariado de peso por altura



A partir da análise do **Figura 4** é possível identificar um padrão, isso é, quanto maior a altura, maior tende a ser o peso, como esperado. Além disso, percebe-se uma maior concentração de pessoas no centro do gráfico. Tal fator possibilita a análise que maior parte da população tem peso entre 60Kg e 90Kg e altura na faixa de 160cm a 185cm.

Quadro 1: Medidas resumo das variáveis Altura (cm)

Estatística	Altura (cm)
Média	171,48
Desvio Padrão	9,87
Variância	97,38
Mínimo	150,00
1º Quartil	164,80
Mediana	171,75
3º Quartil	178,00
Máximo	200,00

Quadro 2: Medidas resumo das variáveis Peso (kg)

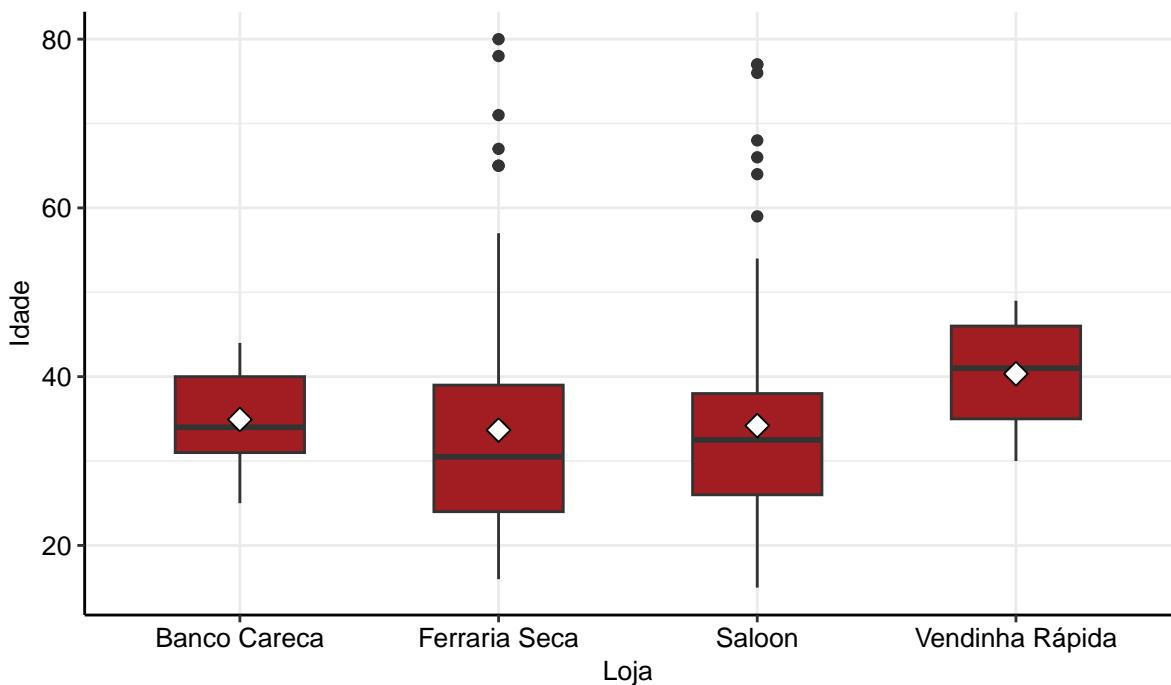
Estatística	Peso (kg)
Média	75,26
Desvio Padrão	11,93
Variância	142,25
Mínimo	45,04
1º Quartil	66,96
Mediana	75,37
3º Quartil	83,27
Máximo	119,41

Com os dados acima apresentados é possível vizualizar dados mais facilmente de modo que a interpretação seja facilitada. Por exemplo, a média de altura dos clientes é 171 cm, e o peso médio é 75,26 kg. A análise do gráfico, juntamente com a análise dos dados escritos no quadro dá indícios que existe uma correlação entre o peso e a Altura. Além disso, a população tem maior concentração em 171cm e 72,26 kg. Entretanto, há casos em que não seguem a concentração, como exemplo, o peso mínimo e máximo, 45,04 kg e 119,41 kg, e a altura máxima e mínima, 150 cm e 200 cm, respectivamente. Além disso, com o cálculo da correlação de pearson, tem-se que as variáveis tem uma correlação moderada a forte. Sendo mais específico, com um coeficiente igual a $\rho = 0,697$.

3.3 Análise da relação entre idade e loja

Esta análise busca vizualizar o comportamento das idades dos clientes de acordo com as lojas da cidade Âmbar Seco, que são variáveis quantitativa contínua e qualitativa nominal, respectivamente. Para vizualizar da melhor forma os dados, serão esboçados um Gráfico Boxplot e uma Tabela. Dessa forma, a visualização dos dados será facilitada.

Figura 5: Box-plot da idade por loja



A partir da análise do **Figura 5**, pode-se concluir que, dentre as 4 lojas e 415 clientes analisados, a Loja Vendinha Rápida tem um público, majoritariamente, mais velho. Além disso, nota-se que os clientes da Loja Ferraria Seca tendem a ter uma idade mais variada, enquanto isso, a loja banco careca tende a ter um público com a idade na faixa de 30 a 40 anos.

Quadro 3: Medidas resumo das variáveis idade(em anos) e loja

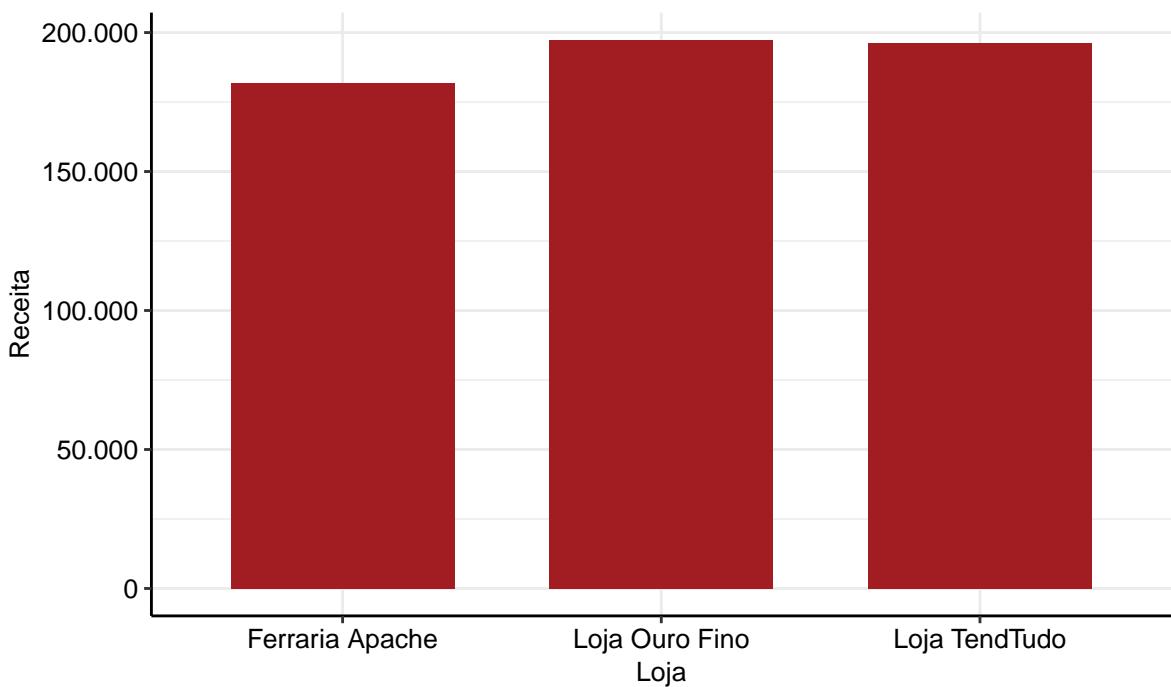
Estatística	Banco Careca	Vendinha Rápida	Saloon	Ferraria Seca
Média	34,9	40,3	34,2	33,7
Mediana	34	41	32,5	30,5
Desvio Padrão	5,57	6,03	12,7	13,3
Mínimo	25	30	15	16
Máximo	44	49	77	80
Variância	31,1	36,4	161	177

Com isso, pode-se visualizar os dados de uma melhor forma. Através do quadro, pode-se concluir que as lojas são ordenadas, por idade, em ordem crescente da seguinte forma: Ferraria seca, Saloon, Banco Careca e Vendinha rápida. Entretanto, o cliente com maior idade frequenta a loja Ferraria Seca. Portanto, conclui-se que a loja que mais atrai o público de maior idade é a loja Vendinha rápida, e a loja que o público de menor idade frequenta é a Ferraria Seca.

3.4 Análise das Top 3 Lojas e Top 3 produtos em 1889

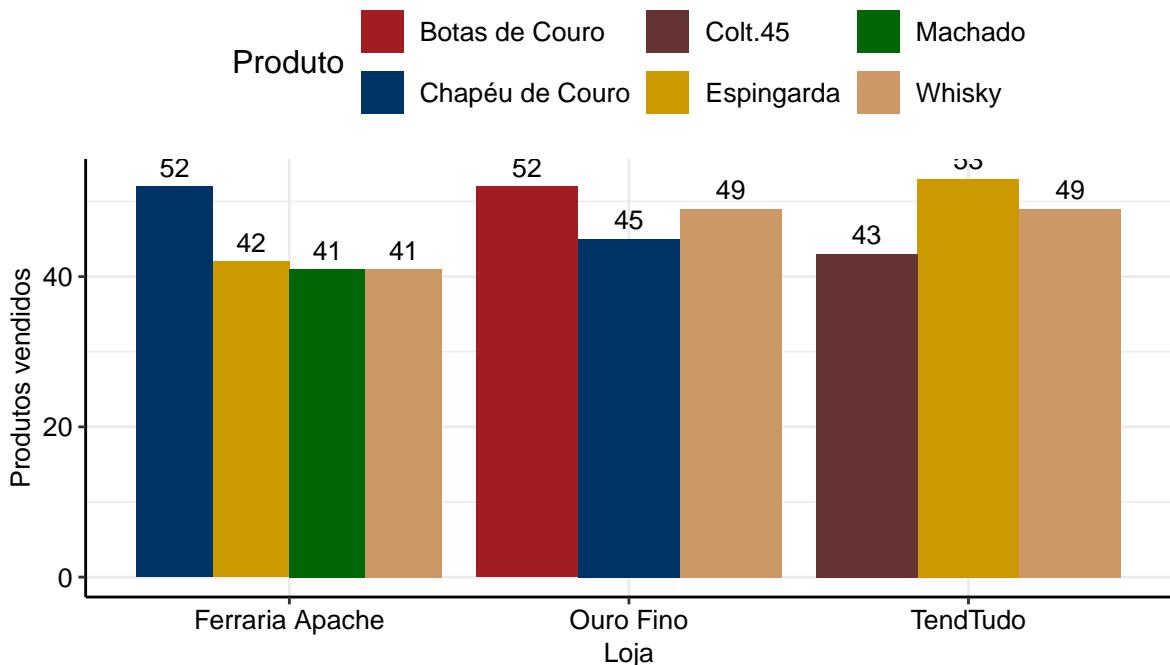
A Partir dessa análise, procura-se descobrir as 3 lojas com maior receita no ano de 1889, e dentre essas lojas, descobrir os três produtos mais vendidos em cada uma. Analisando a partir das duas variáveis (Loja e Produto), que são qualitativas nominais. Essa análise partirá de um gráfico contendo as tres lojas com maior receita e suas receitas, e em seguida, terá uma análise para cada loja contendo os 3 produtos mais vendidos e suas respectivas quantidades.

Figura 6: Três lojas com maior faturamento no ano de 1889



Com o **Figura 6**, pode se concluir que as 3 lojas com maior receita são, respectivamente, Loja Ouro Fino (197312,50 reais), Loja TendTudo (196340,30 reais) e Ferraria Apache (181689,10 reais). A partir dessa análise, pode-se realizar as análises individuais por loja. Essas tres lojas são responsáveis por 20,62% da receita total no ano de 1889.

Figura 7: Top 3 produtos Lojas



Com o **Figura 7**, nota-se que, na loja que têm maior receita, o produto mais vendido foi a bota de couro, com 52 unidades vendidas, gerando uma receita de 12115,56 reais, o que representa 6,14% da receita total da Loja nesse ano. Já o segundo produto mais vendido, foi o whisky, com 49 unidades vendidas, gerando uma receita de 2643,53 reais, o que representa 1,34% da receita total da Loja nesse ano. E, por fim, terceiro produto mais vendido foi o chapéu de couro, com 45 unidades vendidas, gerando uma receita de 6623,69 reais, o que representa 3,36% da receita total da Loja nesse ano.

Além disso, nota-se que, na loja que têm a segunda maior receita, o produto mais vendido foi a Espingarda, com 53 unidades vendidas, gerando uma receita de 52143,35 reais, o que representa 26,56% da receita total da Loja nesse ano. Já o segundo produto mais vendido, foi o whisky, com 49 unidades vendidas, gerando uma receita de 2643,53 reais, o que representa 1,35% da receita total da Loja nesse ano. E, por fim, terceiro produto mais vendido foi a Colt .45, com 43 unidades vendidas, gerando uma receita de 34598,84 reais, o que representa 17,63% da receita total da Loja nesse ano.

Por fim, nota-se que têm quatro produtos. Isso deu-se, pois o número de unidades vendidas dos produtos Machado e Whisky foram iguais (41). Enfim, na loja que têm a terceira maior receita, o produto mais vendido foi o Chapéu de Couro, com 52 unidades vendidas, gerando uma receita de 7654,04 reais, o que representa 4,21% da receita total da Loja nesse ano. Já o segundo produto mais vendido, foi a espingarda, com 42 unidades vendidas, gerando uma receita de 41321,14 reais, o que representa 22,74% da receita total da Loja nesse ano. E, por fim, os dois produtos com mesma quan-

tidade foram o Whisky, com 41 unidades vendidas, gerando uma receita de 2211,93 reais, o que representa 1,22% da receita totalda Loja nesse ano, e o Machado, com 41 unidades vendidas, gerando uma receita de 3864,35 reais, o que representa 2,13% da receita totalda Loja nesse ano.

4 Conclusões

Com base nas análises realizadas, foi possível obter uma compreensão mais ampla e fundamentada sobre o desempenho das lojas, o comportamento dos clientes e os produtos que mais se destacaram no período avaliado.

A avaliação da receita média entre 1880 e 1889 possibilitou identificar a trajetória financeira das lojas ao longo da década, apontando um crescimento significativo, no geral. Entretanto, houve uma queda na receita no ano de 1888. Tais análises são de extrema relevância e podem servir de referência para o planejamento estratégico e a gestão de desempenho futuro.

A análise da relação entre peso e altura dos clientes, conduzida com rigor estatístico, com o intuito de avaliar, se existe correlação estatisticamente relevante entre essas variáveis, apontou que há uma correlação de pearson igual a $\rho = 0,697$. Tal análise permitiu oferecer uma visão adicional sobre o perfil físico do público atendido e suas possíveis implicações.

Na sequência, o estudo sobre a idade dos clientes em Âmbar Seco trouxe informações valiosas sobre o perfil demográfico dos consumidores de cada loja, como o perfil etário de cada loja. Permitindo identificar diferenças de público e oportunidades de adaptação para certo público alvo.

Por fim, a identificação dos produtos mais vendidos nas lojas com maior receita em 1889 destacou quais itens tiveram papel mais relevante na composição das vendas, podendo acarretar em promoções e decisões comerciais mais assertivas.

De forma integrada, os resultados obtidos fornecem uma visão clara e orientada por dados, reforçando o compromisso da ESTAT Consultoria em transformar informações em insights estratégicos para a tomada de decisão. Esse conjunto de análises oferece análises concretas para aprimorar o desempenho das lojas, otimizar recursos e fortalecer a competitividade no mercado.