

Modelo Projeto - Quarto

Consultores Responsáveis:

Laura Mello

Requerente:

João Sábio, Old Town

Road.Ltda

Brasília, 2 de novembro de 2025.

Sumário

	Página
1 Introdução	3
2 Referencial Teórico	4
2.1 Média	4
2.2 Mediana	4
2.3 Quartis	4
2.4 Variância	5
2.4.1 Variância Populacional	5
2.5 Desvio Padrão	5
2.5.1 Desvio Padrão Populacional	5
2.6 Boxplot	6
2.7 Gráfico de Dispersão	6
2.8 Tipos de Variáveis	7
2.8.1 Qualitativas	7
2.8.2 Quantitativas	7
3 Análises	8
3.1 A receita média das lojas registrada nos anos de 1880 até 1889	8
3.2 Variação Peso por Altura	8
3.3 Idade dos clientes de Âmbar Seco a depender da loja	9
3.4 O top 3 produtos mais vendidos nas top 3 lojas com maior receita em 1889	11
4 Conclusões	12

1 Introdução

O seguinte projeto tem como objetivo fornecer uma visão detalhada sobre o desempenho das lojas da região entre os anos de 1880 e 1889, por meio de análises estatísticas que incluem análise descritiva, análise de correlação e comparações entre múltiplas variáveis. Foi avaliado como as vendas das lojas evoluíram ao longo do tempo, o comportamento de vendas de produtos e características do perfil de clientes, com ênfase em identificar quais itens impulsionam o faturamento em cada loja.

O banco de dados utilizado foi fornecido pela Old Town Road Ltda. e contém registros completos das vendas realizadas na região nos últimos dez anos, incluindo informações detalhadas sobre produtos, quantidades vendidas, receita por item, além de características dos clientes, como idade, peso e altura. Com essa abrangência, a base permite analisar com precisão tanto padrões de consumo quanto o impacto financeiro de cada produto em diferentes lojas.

As análises foram realizadas utilizando o software do R.Studio, que permitiu a manipulação e visualização dos dados, realização de cálculos estatísticos e geração de tabelas e gráficos de forma clara e confiável.

2 Referencial Teórico

2.1 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

2.2 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:

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

2.3 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.

2.4 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.

2.4.1 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

2.5 Desvio Padrão

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

2.5.1 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:

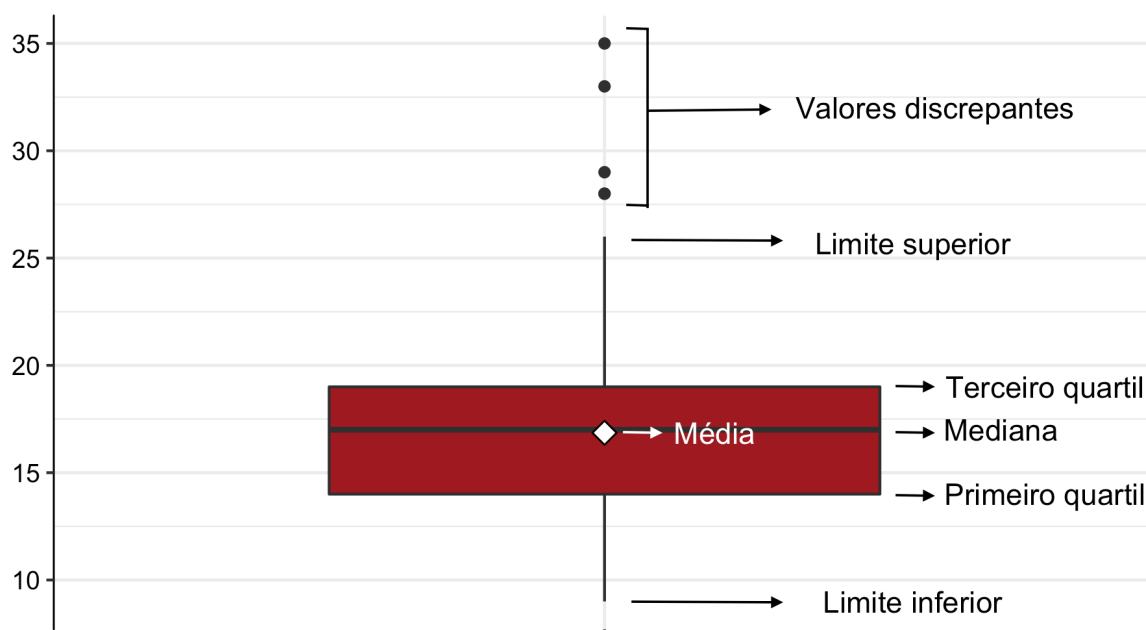
- X_i = i -ésima observação da população

- μ = média populacional
- N = tamanho da população

2.6 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

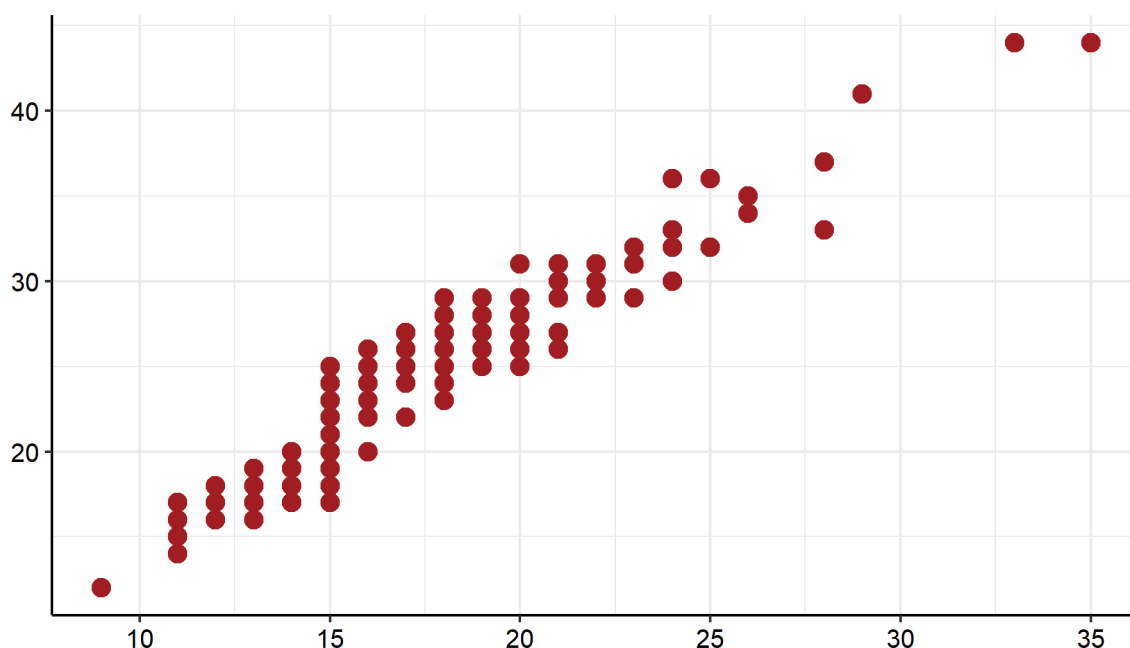


A porção inferior do retângulo diz respeito ao primeiro quartil, enquanto a superior indica o terceiro quartil. Já o traço no interior do retângulo representa a mediana do conjunto de dados, ou seja, o valor em que o conjunto de dados é dividido em dois subconjuntos de mesmo tamanho. A média é representada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

2.7 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



2.8 Tipos de Variáveis

2.8.1 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)

2.8.2 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)

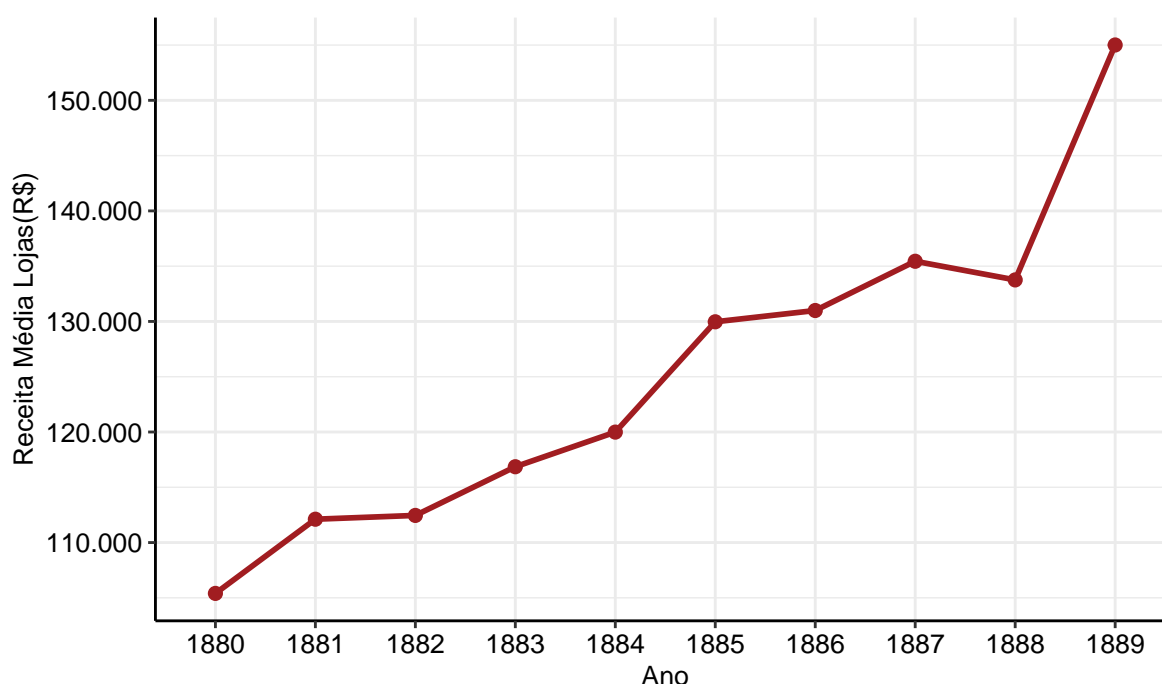
3 Análises

3.1 A receita média das lojas registrada nos anos de 1880 até 1889

Esta análise tem como objetivo compreender a variação da receita média das lojas da região no período entre 1880 e 1889. Para isso, foram consideradas duas variáveis principais: “ano”, que é quantitativa discreta e “ReceitaMédia”, que é quantitativa contínua, expressa em reais.

Os valores monetários foram convertidos para R\$ a fim de padronizar a análise financeira e facilitar a interpretação dos resultados.

Figura 3: Gráfico de linhas da receita média da lojas ao decorrer dos últimos dez anos

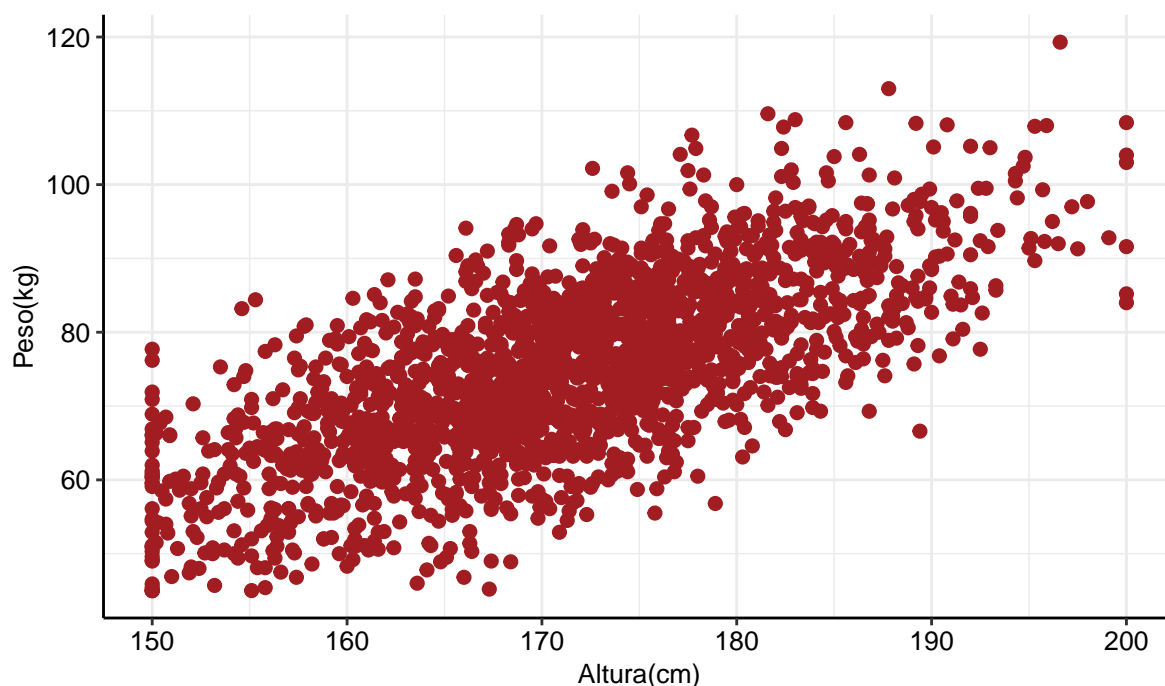


Como mostra a **Figura 3**, a receita média das lojas vem crescendo desde 1880, com exceção de 1888, ano em que houve uma pequena queda de R\$ 1.687,20 em relação ao período anterior. Por outro lado, o crescimento mais significativo ocorreu em 1889, com um aumento de R\$ 21.251,50 na receita média das lojas da região.

3.2 Variação Peso por Altura

Nesta análise busca-se compreender a relação entre a altura dos indivíduos, registrada originalmente em decímetros (dm) e convertida para centímetros (cm), e o peso, registrado em libras (lbs) e convertido para quilogramas (kg). Ambas as variáveis são quantitativas contínuas. O objetivo é verificar se há associação entre as medidas, avaliando se indivíduos mais altos tendem a apresentar maior peso.

Figura 4: Gráfico de dispersão do peso (kg) em função da altura (cm) dos indivíduos



Quadro 1: Medidas resumo do peso (kg) por faixas de altura (cm)

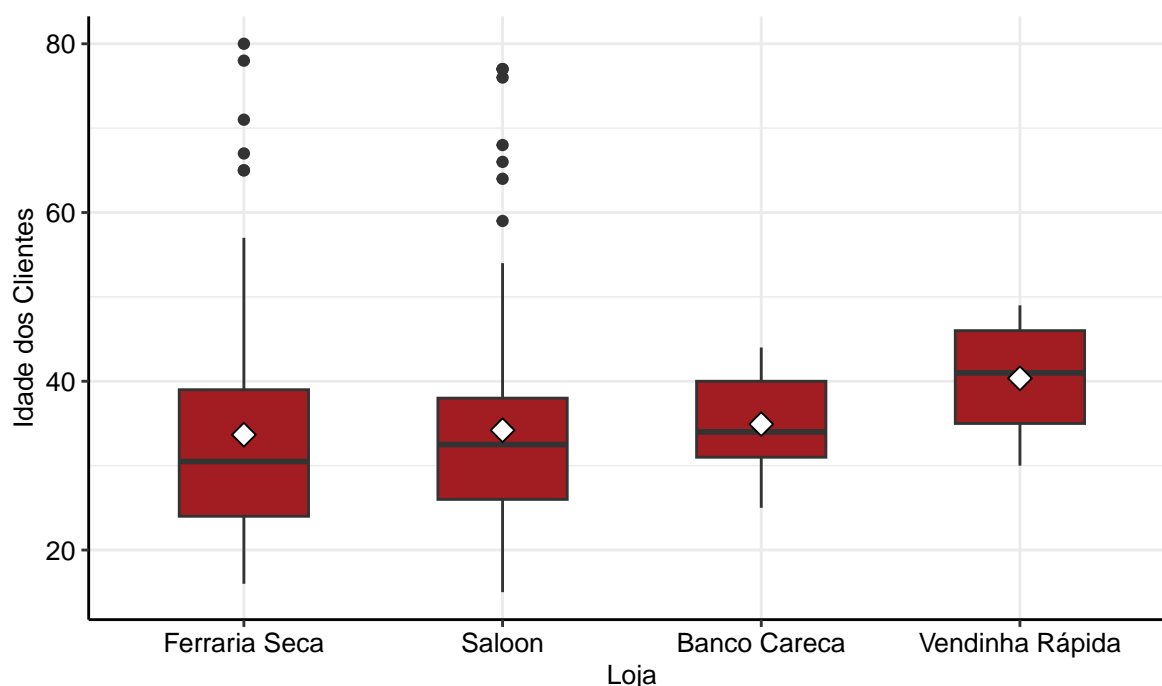
Faixa de altura	Média	Desvio Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
150–159	61,56	8,73	45,00	55,02	61,85	67,15	84,40
160–169	70,01	9,19	45,20	64,50	69,60	75,52	94,70
170–179	77,99	8,79	52,90	72,25	78,50	83,90	106,70
180–189	85,58	8,72	63,10	79,40	84,90	91,50	113,00
190–199	93,81	8,56	76,80	88,07	92,75	99,35	119,30

Medidas resumo do peso (kg) por faixas de altura (cm)

3.3 Idade dos clientes de Âmbar Seco a depender da loja

Esta análise tem como objetivo investigar o perfil etário dos clientes das lojas localizadas na cidade de Âmbar Seco. Para isso, foram consideradas as variáveis Idade dos Clientes uma variável quantitativa contínua e Loja, variável qualitativa nominal que identifica cada estabelecimento.

Figura 5: Boxplot das idades dos clientes por loja em Âmbor Seco



Quadro 2: Medidas resumo da idade dos clientes por loja

NameStore	Média	Desvio Padrão	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo
Banco Careca	34,92	5,57	25,00	31,00	34,00	40,00	44,00
Ferraria Seca	33,67	13,31	16,00	24,00	30,50	39,00	80,00
Saloon	34,20	12,70	15,00	26,00	32,50	38,00	77,00
Vendinha Rápida	40,35	6,03	30,00	35,00	41,00	46,00	49,00

Medidas resumo das Idades dos Clientes por Loja

Conforme evidenciado pelo **Quadro 2** e na **Figura 5**, a distribuição das idades varia significativamente entre as lojas analisadas. A “Vendinha Rápida” apresenta a maior média (40,35 anos), seguida pelo “Banco Careca” (34,92), “Saloon” (34,2) e “Ferraria Seca” (33,67), indicando uma ordem crescente clara das médias. As idades na “Vendinha Rápida” e no “Banco Careca” são relativamente homogêneas, com desvios padrão de 6,03 e 5,57, concentrando-se entre 35 e 46 anos e 31 e 40 anos, respectivamente, enquanto “Saloon” e “Ferraria Seca” apresentam maior variabilidade (12,7 e 13,31), com quartis mais dispersos (1º quartil 26 e 24, 3º quartil 38 e 39) e extremos que vão de 15 a 77 anos e de 16 a 80 anos. Esses dados evidenciam que algumas lojas atendem faixas etárias mais homogêneas e outras apresentam clientes com idades bastante diversificadas.

3.4 O top 3 produtos mais vendidos nas top 3 lojas com maior receita em 1889

1889 Esta análise identifica os três produtos mais vendidos nas três lojas com maior receita em 1889, destacando aqueles que mais contribuíram para as vendas e para a receita total. Foram consideradas as variáveis Loja e Produto (qualitativas nominais), Quantidade Vendida (quantitativa discreta) e Receita (quantitativa contínua, em R\$), com os valores monetários padronizados em reais para facilitar a comparação. O objetivo é compreender o impacto desses produtos na receita das principais lojas e revelar o comportamento de consumo ao longo do ano de 1889.

Quadro 3: Top 3 produtos mais vendidos nas 3 lojas com maior receita em 1889

Loja	Produto	Quantidade Vendida	Receita (R\$)	% da Receita da Loja
Vendinha Rápida	Botas de Couro	55	12.572,75	4,52%
Vendinha Rápida	Cavalo	54	161.027,45	57,88%
Vendinha Rápida	Colt .45	54	43.449,71	15,62%
Loja do Forte	Cavalo	46	137.182,15	58,29%
Loja do Forte	Chapéu de Couro	42	6182,11	2,63%
Loja do Forte	Munição	40	748,36	0,32%
Loja Ouro Fino	Espingarda	63	61.981,72	27,00%
Loja Ouro Fino	Sela	52	28.122,82	12,25%
Loja Ouro Fino	Munição	47	880,98	0,38%

Top 3 produtos mais vendidos nas 3 lojas com maior receita em 1889

Conforme evidenciado pelo Quadro 3, os três produtos mais vendidos nas lojas com maior receita em 1889 apresentam padrões distintos de contribuição para o faturamento. Na “Vendinha Rápida”, embora “Botas de Couro” seja o produto mais vendido em quantidade (55 unidades), ele representa apenas 4,52% da receita da loja, enquanto o “Cavalo”, com quantidade semelhante, concentra 57,88% da receita, evidenciando que nem sempre o volume de vendas se traduz em impacto financeiro proporcional. Na “Loja do Forte”, o padrão se repete: o “Cavalo” domina a receita (58,29%), enquanto itens como “Chapéu de Couro” e “Munição” têm participação muito pequena, apesar de serem vendidos em quantidade considerável. Já na “Loja Ouro Fino”, a distribuição é mais equilibrada entre os dois primeiros produtos, com a “Espingarda” respondendo por 27% da receita e a “Sela” por 12,25%, mostrando que a receita da loja depende de mais de um produto para sustentar o faturamento. Esses dados permitem identificar quais produtos realmente impulsionam a receita em cada loja e perceber oportunidades estratégicas de foco em itens-chave para aumentar o faturamento.

4 Conclusões

O projeto desenvolvido para a Old Town Road Ltda., tem como objetivo compreender o funcionamento do mercado na região em que a empresa pretende investir. As quatro análises realizadas possibilitaram uma visão ampla sobre o comportamento dos clientes, o desempenho das lojas e os produtos mais relevantes para o faturamento.

A primeira análise abordou a evolução da receita média das lojas entre 1880 e 1889, revelando uma tendência geral de crescimento ao longo do período, com destaque para 1889, ano de maior expansão. Essa conclusão indica um cenário de mercado favorável e crescimento consistente, o que reforça o potencial econômico da região e apoia a decisão de João em investir.

Na segunda análise, que explorou a relação entre altura e peso dos clientes, foi observada uma tendência de associação entre as duas variáveis, indivíduos mais altos tendem a ter maior peso, embora exista uma grande variação. Esse resultado, ainda que descritivo, reforça a importância de conhecer o perfil físico médio dos clientes, o que pode auxiliar no planejamento de produtos específicos e em decisões sobre o público-alvo em potenciais investimentos.

A terceira análise, sobre a distribuição das idades dos clientes por loja, mostrou que há diferenças marcantes no perfil etário do público atendido. Algumas lojas concentram clientes em faixas etárias mais homogêneas, o que sugere um público-alvo bem definido, enquanto outras apresentam maior diversidade de idades, indicando um atendimento mais amplo. Esses resultados ajudam a entender quais lojas devem fazer campanhas mais segmentadas e quais podem se beneficiar de estratégias mais generalistas.

Por fim, a quarta análise identificou os três produtos mais vendidos nas três lojas com maior receita em 1889. Ficou evidente que, embora alguns produtos apresentem grande volume de vendas, nem todos contribuem de forma significativa para a receita total. Em contrapartida, determinados itens de menor quantidade vendida tiveram alto impacto financeiro. Esse conhecimento ajuda a compreender quais produtos realmente impulsionam os resultados e onde estão as melhores oportunidades de investimento.